

RESEARCH

Open Access



MonuNet: a high performance deep learning network for Kolkata heritage image classification

A. Sasithradevi¹, Sabari nathan², B. Chanthini³, T. Subbulakshmi^{4*} and P. Prakash⁵

Abstract

Kolkata, renowned as the City of Joy, boasts a rich tapestry of cultural heritage spanning centuries. Despite the significance of its architectural marvels, accessing comprehensive visual documentation of Kolkata's heritage sites remains a challenge. In online searches, limited imagery often fails to provide a detailed understanding of these historical landmarks. To address this gap, this paper introduces MonuNet, a high-performance deep-learning network specifically designed for the classification of heritage images from Kolkata. The development of MonuNet addresses the critical need for efficient and accurate identification of Kolkata's architectural marvels, which are significant tangible cultural heritages. The dataset used to train MonuNet is organized by heritage sites, each category within the dataset represents distinct sites. It includes images from 13 prominent heritage sites in Kolkata. For each of these sites, there are 50 images, making it a structured collection where each category (heritage site) is equally represented. The proposed network utilizes a unique architecture incorporating a Dense channel attention module and a Parallel-spatial channel attention module to capture intricate architectural details and spatial relationships within the images. Experimental evaluations demonstrate the superior performance of MonuNet in classifying Kolkata heritage images with an accuracy of 89%, Precision of 87.77%, and Recall of 86.61%. The successful deployment of MonuNet holds significant implications for cultural preservation, tourism enhancement, and urban planning in Kolkata, aligning with the United Nations Sustainable Development Goals (SDGs) for sustainable city development. By providing a robust tool for the automatic identification and classification of heritage images, MonuNet promises to enrich online repositories with detailed visual documentation, thereby enhancing accessibility to Kolkata's cultural heritage for researchers, tourists, and urban planners alike.

Keywords Kolkata, Heritage image, Deep learning, Dense channel attention module, Parallel-spatial channel attention module

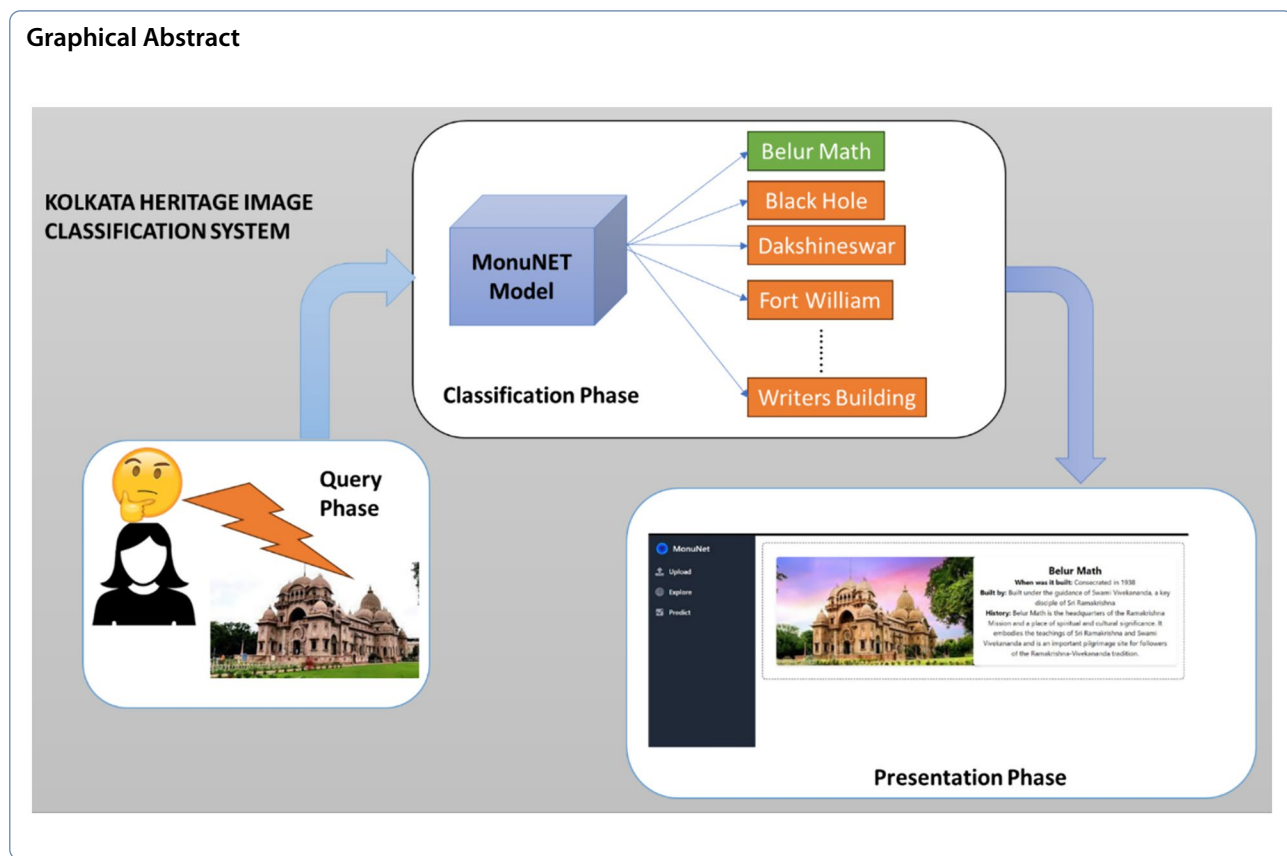
*Correspondence:

T. Subbulakshmi
subbulakshmi.t@vit.ac.in

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



Introduction

Kolkata, the City of Joy, stands as a living testament to India’s rich and diverse cultural heritage [1]. Nestled along the banks of the Hooghly River, this city has witnessed the flow of empires, the birth of literary giants, and the evolution of unique architectural marvels. From the colonial-era the city edifices temples and memorials and also the Kolkata’s cultural heritage includes intricate tapestry that spans centuries [2, 3]. The choice of Kolkata as the focal point for this work stems from the city’s unparalleled wealth of cultural heritage, particularly its monuments. These monuments are tangible cultural heritages, each with its unique story; represent the legacy of civilizations that have left an indelible mark on the city’s landscape [4]. While Kolkata’s cultural heritage is a source of immense pride, it also demands responsible preservation and recognition to ensure that these treasures endure for generations to come. The development of Information and communication technology (ICT) has paved a significant way to preserve large amount of data in the form of text and images [5–7].

Currently, many research works has reported efficient image classification for various applications, but very few research works have been devoted for classification of Tangible cultural heritage like monuments and

buildings. Machine learning (ML) and Deep Learning (DL) techniques are employed in the classification of the monument images [8–11]. The importance of image classification lies in its ability to facilitate efficient recognition and understanding of the vast array of monuments scattered across the Globe. Using ICT tools for creating a digital library for these monuments along with artificial intelligence (AI) for the identification of these historical sites also provide a bridge to the past, offering insights into their historical and architectural significance. The societal benefits of this research are manifold. Effective monument recognition can enhance tourism experiences, making Kolkata’s cultural heritage more accessible to both residents and visitors. Moreover, it aids in the preservation efforts, as it enables the assessment of structural and conservation needs. It also helps in urban planning where urban planners can use this model to identify the locations of monuments and incorporate this information into city development plans, ensuring that new structures and infrastructure projects respect the city’s historical heritage. By actively participating in the safeguarding of cultural treasures, this work contributes to the sustainable development of the city and its economy, aligning with the United Nations Sustainable Development Goals (SDGs).

The proposed work centers on the critical task of classifying and recognizing these monuments using image analysis, specifically leveraging deep learning techniques. The method used for our work involves organizing the dataset specifically by heritage sites, where each site is a distinct category and is represented with an equal number of images. This differs from traditional methods of dataset creation which often involve collecting a large, diverse set of images without such strict categorization or equal representation per category. Traditional datasets may prioritize a broad collection of images to cover more general scenarios without focusing on specific, evenly represented categories. In contrast, our structured approach by specific heritage sites ensures that the model can learn detailed and nuanced features unique to each location, which is critical for tasks requiring high specificity like heritage site recognition. The availability of comprehensive datasets on Kolkata's monuments posed a major challenge in this research work. A dataset was created by collecting the different view of 13 heritage monuments with each category having 50 images. The inclusion of 50 images per heritage site in the MonuNet dataset balances comprehensiveness and manageability, allowing for varied representation of features and perspectives critical for training deep learning models. This uniformity ensures each site is represented equally, avoiding biases from disproportionate representation. Such balance is essential in machine learning, as it enhances model robustness and consistency across different data categories, ultimately improving accuracy and generalizability. This method aligns with best practices in the field, emphasizing the importance of a sufficient and balanced number of samples for each category. Despite the vast cultural wealth that the city harbors, there was a conspicuous absence of publicly accessible, well-annotated datasets for image classification and recognition tasks. This dearth of data, crucial for training and validating the deep learning models, prompted to create a dataset. The

collected dataset not only serves as the backbone to this research but also addresses a critical gap in the resources available for the preservation and recognition of Kolkata's cultural heritage. The Fig. 1 shows the various images from the Kolkata heritage monument dataset.

The organization of this article in a nutshell as follows: Section 2 reviews the existing works related to heritage image classification. Section 3 provides the dataset curation process. Section 4 gives the detailed explanation about our proposed workflow for Kolkata heritage image classification system. Section 5 discusses the experimental evaluation of our MonuNet on curated dataset. Section 6 concludes our work.

Related works

This section discusses the various works reported on the classification of cultural heritage monument images not only limited to tangible cultural heritage [12, 13]. Early works and advancements: Developing an efficient image classification technique has its application in manifolds. The initial application of CNN models for classifying heritage monuments and the major landmarks of Pisa has reported. A dataset of 1227 images with 12 classes was used and various feature extraction techniques were employed before applying CNN model for classification of images [14]. The use of a pretrained MobileNet V2 model optimized with Bayesian optimization for classifying UNESCO certified heritage sites in India, facilitated by a crowdsourced dataset has been explored. The dataset was collected by crowdsourcing platform for Indian digital heritage space monuments dataset, integrated into a web app [11]. CNN based Indian heritage monument identification was proposed for its application in various domains like education, tourism, cultural preservation and so on. The model was deployed on cloud platforms, and a constant process of feedback collection and model refinement was also included for automating the monument identification process with information

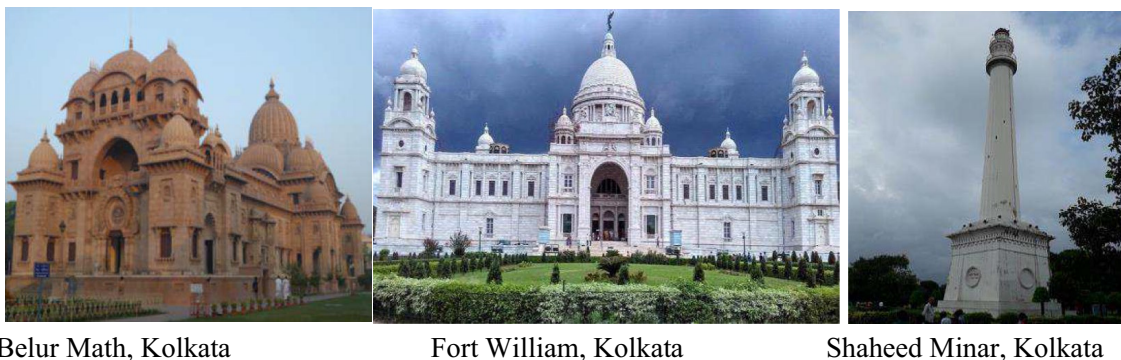


Fig. 1 Images from dataset used to train MonuNet

up-gradation [15]. Another work introduces the CAH10 dataset, comprising 3080 images representing 10 classes of Chinese architectural heritage [16]. This work aims to enhance image retrieval in cultural heritage, particularly in the context of local architectural heritage in China and also introduces an innovative deep hashing method to tackle data scarcity. Point clouds for architectural classification details the innovative approach of transforming heritage images into 3D point clouds for improved classification accuracy using deep learning methods [17, 18]. In this approach firstly, the images are converted to 3D point clouds. Then, feature extraction is done and the dataset is trained on a deep learning classifier. The results demonstrate that the deep learning framework surpasses existing state-of-the-art methods in the semantic segmentation of Point Clouds, achieving enhanced accuracy in architectural element classification by effectively capturing local geometric features.

Additional studies have explored the classification of architectural features, such as bridges [19] and street façades [20], emphasizing their structural characteristics [21]. Other research has utilized machine learning tools like WEKA to apply models to architectural datasets [22]. These models classify images using three features: fuzzy histograms, edge histograms, and DCT coefficients. The dataset employed consists of 150 images of heritage sites, with features extracted using four tree algorithms: J48, Hoeffding Tree, Random Tree, and Random Forest [22]. Moreover, further developments include mobile applications designed to facilitate easy access to and identification of various monuments. Deployed using the Python Flask framework, these applications enable users to photograph monuments and access a wealth of information such as nearby tourist attractions, hotel recommendations, user ratings, price ranges, reviews, monument descriptions, official websites, and 360-degree views [10].

Works addressing classification challenges: Heritage image classification has played a huge role in the cultural image recognition of many countries [23]. Various image processing techniques were employed to analyze the ancient monuments and heritage sites of Thailand. The goal was to glean information and craft narratives about these sites, with the aim of preserving knowledge and sparking interest for future generations. Researchers utilized convolutional neural networks (CNN) to classify and extract pertinent details from images of cultural heritage [24], a project supported by King Mongkut's University of Technology North Bangkok. Additionally, a proposal was made to classify digital documentation images of cultural heritage sites using CNN [25]. The dataset used in this study, known as the Architectural Heritage Elements Dataset, is available in various versions of different sizes. The primary aim of this research

is to leverage the growing volume of digital documentation within the cultural heritage field, which includes images captured by non-professionals using smartphones. However, these images often suffer from a lack of clarity, missing source captions, and inadequate categorization, which complicates the classification process.

Another study highlighted the significance of image classification for cultural preservation and the challenges posed by imbalanced categories in image datasets. The research showed that clustering and focusing on local features, such as dominant image gradients, can achieve high classification rates even with incomplete data [10]. Additionally, deep learning techniques have been applied to monitor the structural health of heritage buildings. This particular study concentrated on utilizing a variety of machine learning algorithms and pre-trained models to predict factors like compressive strength and damage scenarios that influence the structural integrity of heritage buildings. A range of standard ML algorithms and pre-trained models were employed to assess the structural integrity of these buildings [26].

It is evident from the literature that there is a huge space for ML and DL techniques in the field of heritage image classification. Also, the earlier works on the perspective of conservation and preservation of Kolkata heritage images was not available. With the notion to preserve and document those images, we observed about the characteristics of Google Lens. One of the significant features that hinders us to get the heritage related information from it is its limited knowledge about local Indian sites. Also, it may not provide in-depth context about our local heritage. Hence, we take an initiative step to perform digital preservation of Kolkata heritage images. We propose Kolkata heritage image search (KHIS) system to accomplish the task and the novel contributions are:

- Curation of Kolkata heritage image dataset which comprises the significant monuments with different architectural style in Kolkata
- Creation of "MonuNet"—a new deep learning model to learn the complex characteristics in the architecture style of Kolkata
- Evaluation of MonuNet on curated dataset to categorise the query image into one of the 13 categories,
- Presentation of details about the Kolkata monument through web-app to the user.

Kolkata monument data curation process

As there is always a critical requirement to preserve the Kolkata heritage monument, we are taking the initial step to achieve it. To the best of our knowledge, there is no public dataset available for our work. Hence, it is mandatory to curate the required dataset. We started to collect

the dataset from open sources such as flicker, google, instagram and shutterstock. Our keywords for image search were “Kolkota Monument”, “Kolkata monuments”, “Kolkata heritage sites”, “Kolkata historical landmarks”, “Kolkata architectural heritage”, “Kolkata famous buildings”, “Kolkata iconic structures”, “Kolkata historical monuments”, “Kolkata tourist attractions”, “Kolkata architectural landmarks”, “Kolkata cultural heritage”. We carefully selected thirteen of the most significant heritage places located in Kolkata, as designated by the West Bengal Heritage Commission (WBHC) [3], an authoritative body responsible for the preservation and promotion of heritage sites within the state. The WBHC is recognized for its rigorous criteria in identifying sites of historical and cultural importance, ensuring that our selection is grounded in a well-respected framework. We also searched using specific keywords like “Dakshineswar Kali Temple Kolkata”, “Marble Palace Kolkatta”, “Shaheed Minar”, “Victoria Memorial Kolkatta”, “Raj Bhavan” and so on. The list of monuments chosen for our work and the details are listed in Table 1.

Belur Math is the headquarters of Ramakrishna Math and Mission. The Black Hole Monument commemorates the Kolkata incident. Dakshineswar Kali Temple, dedicated to Goddess Kali, is a revered site. Fort William, originally built by the British, now hosts the Indian Army’s Eastern Command. Howrah Bridge, a cantilever bridge over the Hooghly River, is a Kolkata landmark. The James Prinsep Memorial, in neo-Gothic style, celebrates James Prinsep, a scholar who deciphered Brahmi script. Marble Palace, noted for its marble architecture and art collections, dates back to the 19th century. Metcalfe Hall features a Neoclassical design with Corinthian columns and houses a library. Raj Bhavan is the residence

of West Bengal’s Governor. Shaheed Minar and St. Paul’s Cathedral, with its Indo-Gothic architecture, are key historical sites. The Victoria Memorial and Writers’ Building are iconic British-era structures. These sites are each represented by 50 images in the MonuNet dataset, aimed at preserving Kolkata’s architectural heritage.

Proposed framework for Kolkata heritage image classification

The classification using the MonuNet model is crucial because it enables the system to categorize each query image into one of the thirteen monument classes. This categorization helps in organizing the data efficiently, enhancing the search functionality, and improving user experience by quickly matching query images to their corresponding monument categories. Classification is fundamental to making the heritage image search system intuitive and responsive, facilitating educational, research, and preservation efforts related to Kolkata’s cultural heritage. The entire pipeline of our Kolkata heritage image classification includes three phases: (i) query phase and (ii) classification phase and (iii) presentation phase. During query phase, anonymous user rises an image query with a motivation to get the heritage information lying behind the query image. In the classification phase, We extract the high level features using the expertise driven “MonuNet” model and classifies the query into one of the thirteen monument classes. In the context of MonuNet, ‘high-level features’ refer to the distinguishing attributes of images that are learned by the model to effectively differentiate between the thirteen monument classes. For example, these features could include architectural elements like the shape and style of arches, the presence of specific motifs or sculptures, and

Table 1 Details of specific Kolkota monuments

S.No	Monument	Architectural style
1	Belur Math	Blend of Hindu, Islamic, Christian, and Buddhist
2	Black Hole Monument	Historical commemorative monument, no specific architectural style
3	Dakshineswar kali temple	Bengali Navaratna architectural style
4	Fort William	British colonial military architecture
5	Howrah Bridge	Cantilever truss bridge
6	James Prinsep Memorial	Neo-Gothic architectural style
7	Marble Palace	Neoclassical architecture with marble facade
8	Metcalfe Hall	Neoclassical architectural style
9	Raj Bhavan	Blend of neoclassical and British colonia
10	Shaheed Minar	Indo-Saracenic architectural style
11	St.Paul’s Cathedral	Indo-Gothic architectural style
12	Victoria Memorial	Indo-Saracenic Revival architectural style
13	Writers Building	Blend of British and Mughal architectural styles

unique color patterns of the buildings. Such features help MonuNet recognize and classify an image of a monument based on its architectural signature and stylistic elements, enabling accurate identification and classification within the system. The presentation phase provides the insightful and valuable information to the user according to the classification result. MonuNet acts as the brain and expert for classifying the query.

Evolution of MonuNet

Recognizing the challenge in classifying heritage site images in Kolkata, we proposed developing a dedicated deep learning model. The MonuNet work was initiated with a focus on developing a model that could accurately classify images from 13 major heritage sites. Our team began collecting images, aiming to gather 50 high-quality images for each site to ensure comprehensive training data. After that, the basic architecture of MonuNet was designed, incorporating convolutional neural networks and early versions of attention mechanisms. MonuNet was tested, revealing initial flaws in classification

accuracy and model efficiency. To improve performance, advanced features like parallel spatial channel attention modules were integrated to enhance feature extraction and focus on relevant details in images. Extensive testing and optimization phases were conducted to refine the model’s accuracy and efficiency.

MonuNet architecture details

The fundamental (DenseNet121[]) blocks of MonuNet consists of three different units as shown in Fig. 1. Unit 1 includes three consecutive operations as 2D convolution (kernel size 7X7 with a stride of 2), batch normalization and activation. The convolution process extracts the spatial features from the curated Kolkota heritage images. Batch normalization aids MonuNet in the convergence process through normalizing and stabilizing the activations in each mini batches. The activation operation introduces non linearity into the MonuNet through ReLU function. Based on the information presented in Fig. 2, it can be inferred that Unit 1 and Unit 2 share common operations but differ from each other

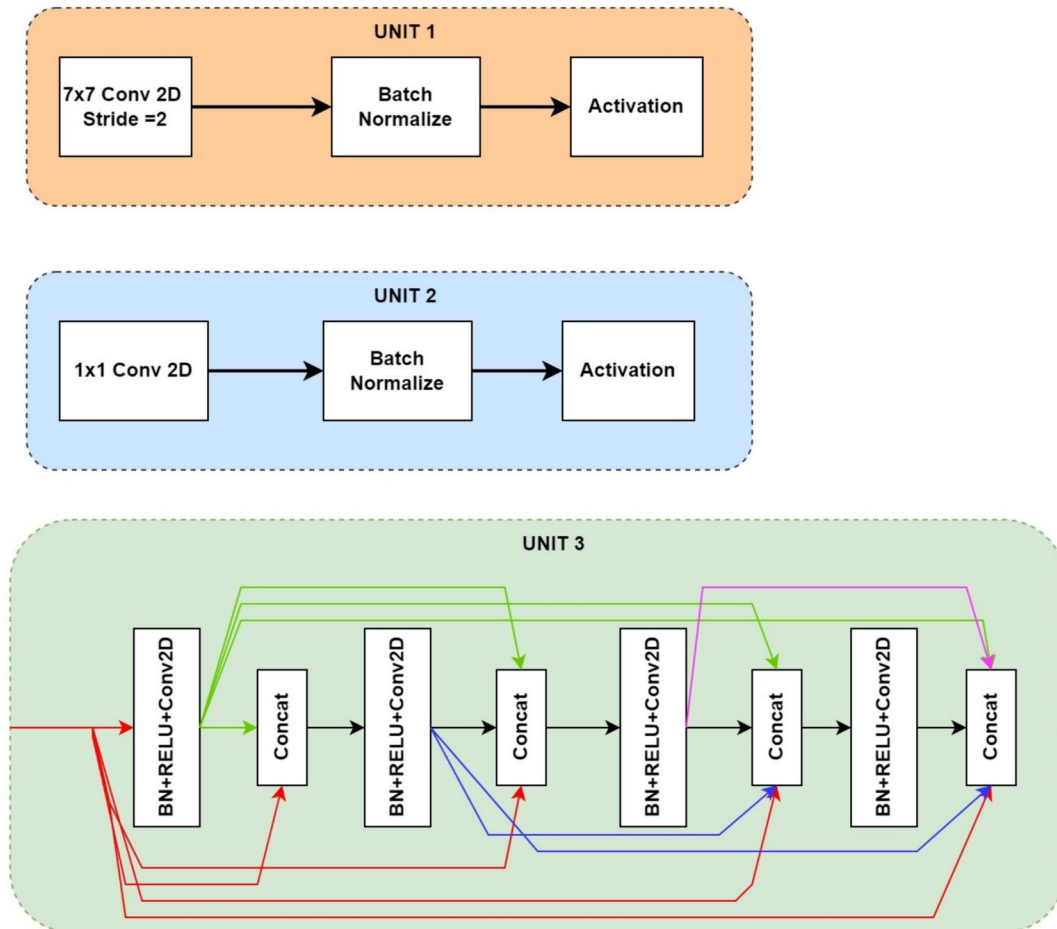


Fig. 2 Fundamental blocks of MonuNet

solely in terms of the convolution process. The kernel size for the 2D convolution process in unit 2 is “1X1”. Unit 3 extracts and learns the complicated features from the heritage images. It consists of four repetition layers comprising a combination of batch normalization, relu, 2D Convolution followed by concatenation. The normalization function ensures that the MonuNet attains effective learning stage with reduced internal covariate shift. Relu activation supports the derivation of complex inter-relationships between the images. The hierarchical features are learned using the multi-scale features obtained using 2D convolution process. The concatenation of these layers enhances the power of MonuNet by creating the abstract and meaningful representations of input images.

These three fundamental units form the architecture of MonuNet as visualized from Fig. 3. The units are arranged to make up the five different blocks. Each block is combined with the next block by max and average pooling. Dense channel attention module (DCAM) operates on the intermediate representation or the feature maps obtained from block 5. The channel based attention map is computed and the significance of each channel is learned. The attention map from each channel is combined with the attention map of consecutive channels to attain dense connectivity. The DCAM operation is mathematically represented as,

$$C_A = \text{softmax}(\text{ReLU}(W_2.\text{avgpool}(W_1.Y))) \tag{1}$$

$$Y' = Y.C_A \tag{2}$$

where ‘Y’ denotes the intermediate feature map, avg-pool represents the average pooling operation, W_1 & W_2 are the updatable weights, *ReLU* denotes the activation function and *softmax* function normalizes the activation scores across all attention channels. The enhanced feature representation obtained from DCAM is further tuned by parallel spatial channel attention module.

A parallel spatial channel attention module is an advanced feature in deep learning architectures designed to enhance model performance by focusing on the most informative features of input data. This module operates by applying attention mechanisms separately but simultaneously across spatial and channel dimensions of input data. The spatial attention focuses on identifying important regions within the data, enhancing relevant spatial features, while the channel attention selectively emphasizes informative channels, effectively capturing interdependencies between different feature maps. By processing these two attention mechanisms in parallel, the module can more efficiently and accurately refine feature representation, leading to improvements in tasks like image recognition, segmentation, and classification. This dual focus allows for a more nuanced understanding and processing of complex input data, making it particularly

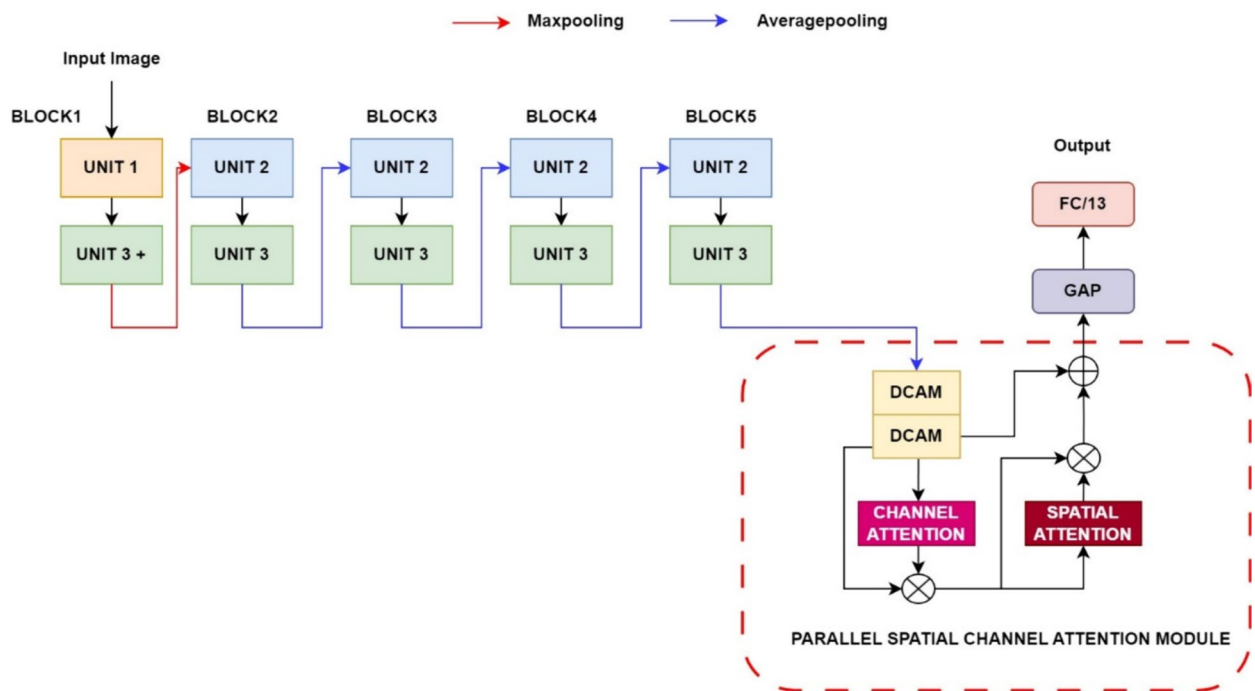


Fig. 3 Architecture of MonuNet

useful in deep learning models that handle high-dimensional datasets.

Experimental results and discussion

This section discusses about the experimental setup, evaluation metrics and performance evaluation of MonuNet in categorizing the query image into 13 classes.

Experimental setup

We utilized Nvidia 1080 GTX GPU for evaluation of MonuNet model. The evaluation process involved our carefully curated dataset, augmented to populate the images. The augmentation process through five transformations such as height shift, width shift, rotation, horizontal and vertical flips. A common baseline is maintained for image resolution as 224×224 . Weight update during training phase is attained by Adam optimizer with the initial learning rate is set as 0.001. If the validation loss holds even after 5 epochs, it is adjusted to 10%. Categorical loss entropy and batch size of 2 is used. Training occurs till 250 epochs, if overfitting is not inferred. Early stopping strategy is used to alleviate the overfitting issue.

Performance assessment metrics

- Accuracy provides the correctly identified categories by MonuNet by referring the ground truth in dataset labels.
- Confusion matrix helps to provide insights into the MonuNet model's performance in tabular format. The rows refer the classes in ground truth and column refers to the identified categories by MonuNet.
- Precision denotes the ability of MonuNet in identifying the positive classes and can be represented mathematically as,

$$Precision = \frac{actualpositive \cap predictedpositive}{actualpositive \cap predictedpositive + actualnegative \cap predictedpositive} \quad (3)$$

- Recall depicts the Monunet's true positive rate and it is written as,

$$Recall = \frac{actualpositive \cap predictedpositive}{actualpositive \cap predictedpositive + actualpositive \cap predictednegative} \quad (4)$$

- F1score denotes the overall performance of Monunet and it is written as,

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

Performance evaluation of MonuNet on curated dataset

Figure 4 portrays the confusion matrix for the classification performance of MonuNet on curated dataset. Each row in the matrix denotes the samples in an actual class while each column indicates the samples in a predicted class. The diagonal elements denote the number of instances for which the predicted class is equal to the actual class, thus indicating correct classifications by the MonuNet model. Our MonuNet model shows good performance for various classes like James Prinsep Memorial, Dakshineswar Kali Temple, Shaheed Minar and Metcalfe Hall with the best score of 1. Few monuments have been interpreted as others due to complexity in architectural style. It can be clearly noted from the non-zero entries in the off-diagonal elements. For Fort William, 38% of its instances were misclassified as Raj Bhavan, and similarly, 14% of Raj Bhavan instances were misclassified as Fort William. Also, our model finds difficulty in categorizing St. Paul's Cathedral and Victoria Memorial. Despite few misclassifications, MonuNet seems to do a good classification process with most of the classes having an accuracy greater than 80%.

Table 2 enlists the performance evaluation metrics for MonuNet with James Prinsep Memorial achieving perfect precision, recall, and F1-score of 1. A high precision with slightly low recall can be inferred for Belur Math and Howrah bridge. Also, a potential space for model improvement can be noted for the Fort William and Shaheed Minar classes. Figure 5 shows the receiver operating characteristic for Monunet and it depicts the strong classifying capability with AUC as 1 for Dakshineswar Kali Temple and James Prinsep Memorial. Even though the macro average AUC is 0.93, AUC for Fort William is quite poor as 0.80.

Webapp development

Following the selection of MonuNet as the best-performing model, we seamlessly integrated it into a user-friendly web application, featuring various functionalities:

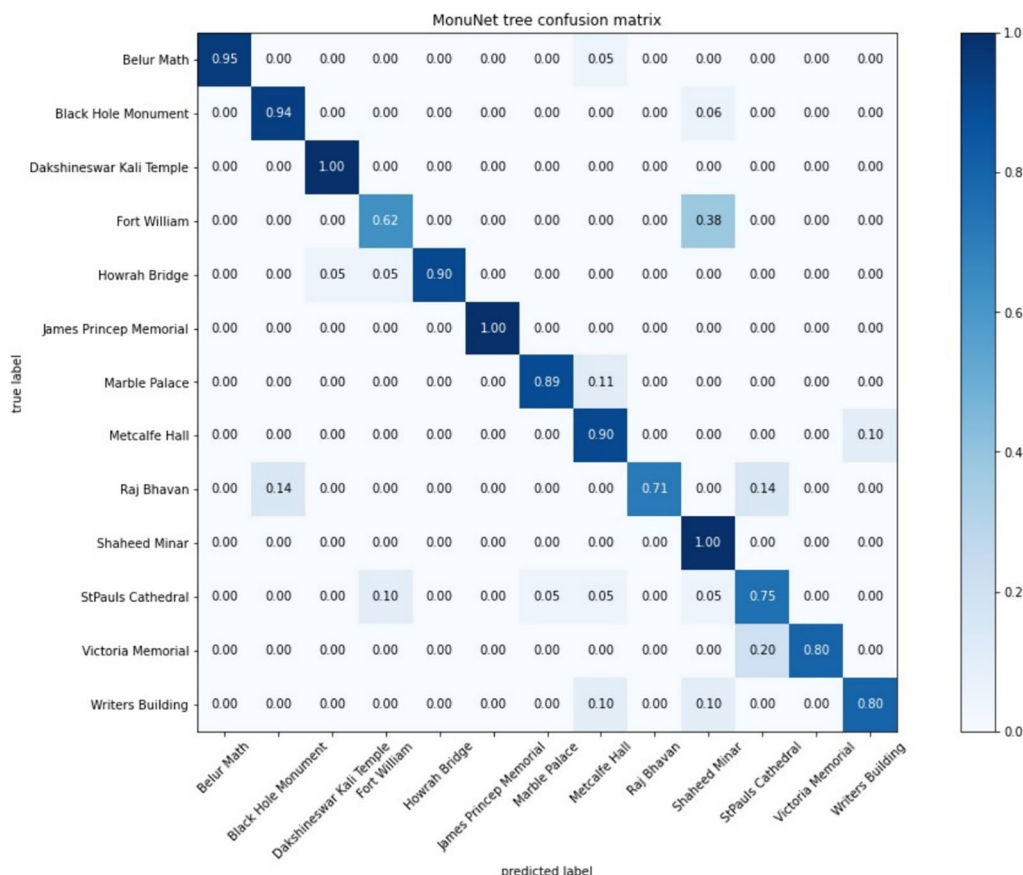


Fig. 4 Confusion matrix-MonuNet

Table 2 Performance metrics for MonuNet

Monuments	Precision	Recall	F1-score
Belur Math	1	0.95	0.97
Black Hole Monument	0.94	0.94	0.94
Dakshineswar Kali Temple	0.95	1	0.98
Fort William	0.62	0.62	0.62
Howrah Bridge	1	0.9	0.95
James Prinsep Memorial	1	1	1
Marble Palace	0.89	0.89	0.89
Metcalfe Hall	0.69	0.9	0.78
Raj Bhavan	1	0.71	0.83
Shaheed Minar	0.6	1	0.75
StPauls Cathedral	0.83	0.75	0.79
Victoria Memorial	1	0.8	0.89
Writers Building	0.89	0.8	0.84

- Client–Server architecture: the web application follows a client–server architecture, employing React for the client-side UI, Firebase for backend services (authentication and data storage), Node.js for server-

side logic, and Fast API to create API endpoints. Tailwind CSS is utilized for styling React components with a utility-first approach.

- Webapp sections: the webapp comprises three sections: upload, explore, and predict. In the Upload section, users can input images of Kolkata monuments, triggering the trained model to predict the monument’s identity as seen in Fig. 6. The result, the name of the monument, is then displayed. An additional action button allows users to access detailed information about the predicted monument.

From Fig. 6, We can see how the model perfectly identified the image of Belur Math which was uploaded by the user. We also see extra information on Belur math like when it was built, by whom it was built by and a short history of the monument. Uploaded images are stored in a database, accessible through the Explore section as seen in Fig. 7. Users can peruse additional monuments and gather more information, providing a valuable reference for tourists based on the searches of others.

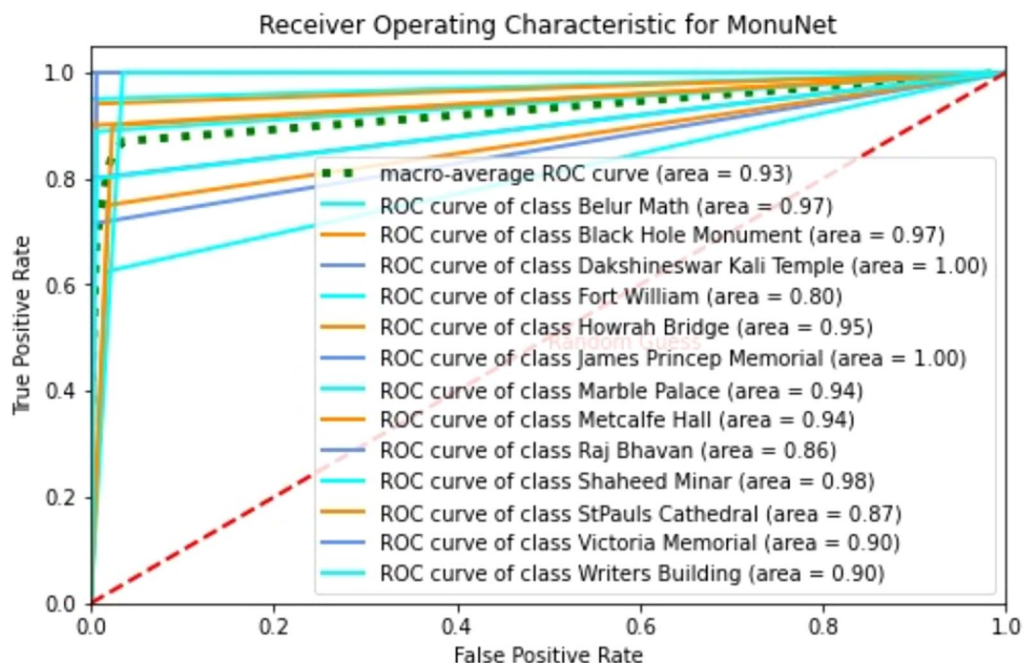


Fig. 5 ROC for MonuNet

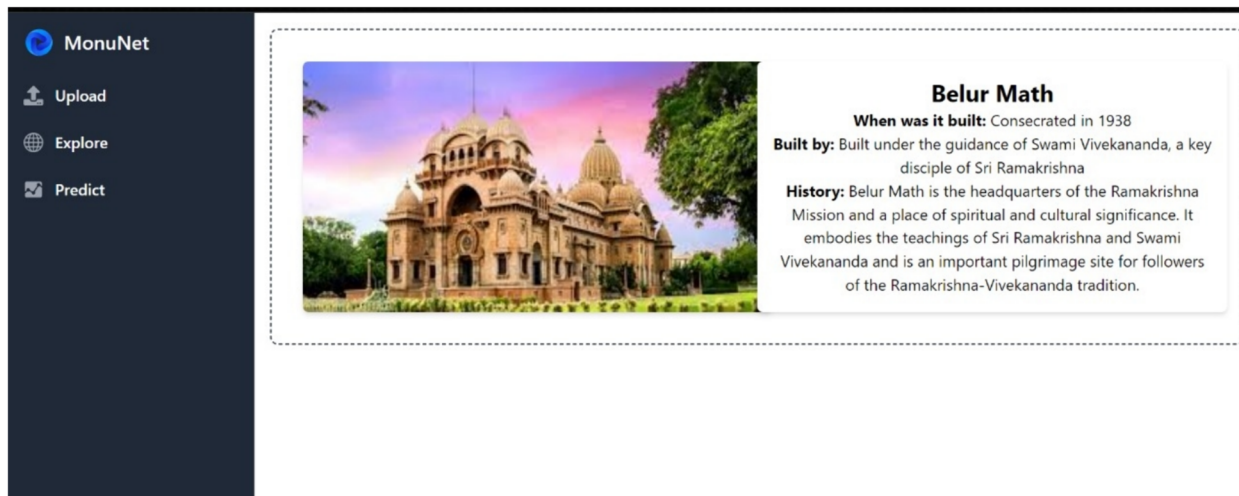


Fig. 6 Implementation of the trained model on a webapp

This web application not only showcases the practical utility of our work but also contributes to the preservation and promotion of Kolkata’s rich cultural heritage. The seamless integration of technology and cultural preservation underscores the potential impact of our project on both technological and historical fronts. The technology implemented in MonuNet can also complement the data capabilities of Historical Geographic Information Systems (HGIS) by enhancing the accessibility and specificity of information related to cultural heritage

sites. While HGIS provides a broad spatial and temporal mapping of historical data, MonuNet specializes in the detailed classification and recognition of heritage site imagery. This synergy can allow users to not only see where heritage sites are located and understand their historical context through HGIS maps but also to visually identify and classify these sites through images analyzed by MonuNet.

One major advantage of integrating MonuNet with HGIS is the enriched user experience in educational and

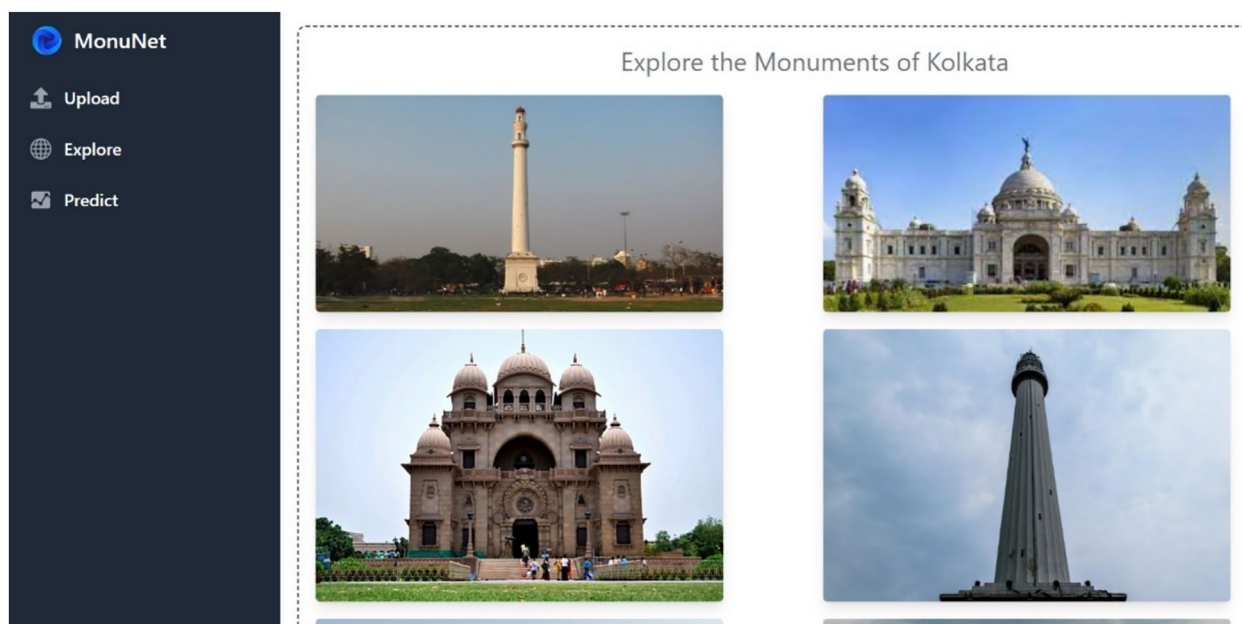


Fig. 7 Uploaded images under the explore section of WebApp

Table 3 MonuNet Vs DenseNet

Metric	MonuNet	DenseNet
Accuracy (%)	89	87.7
FLOPs	2.3 GFLOPs	3.8 GFLOPs
Parameters	1.2 million	3.5 million
Inference Time (ms)	45	60

research settings. Users can cross-reference visual data with geographical information, making it easier to connect visual architectural features with their historical and spatial contexts. Furthermore, this integration will enhance digital archives by linking precise image classification with detailed historical maps, thereby improving data retrieval and the accuracy of information provided to users interested in heritage studies.

Comparison with existing DenseNet versions

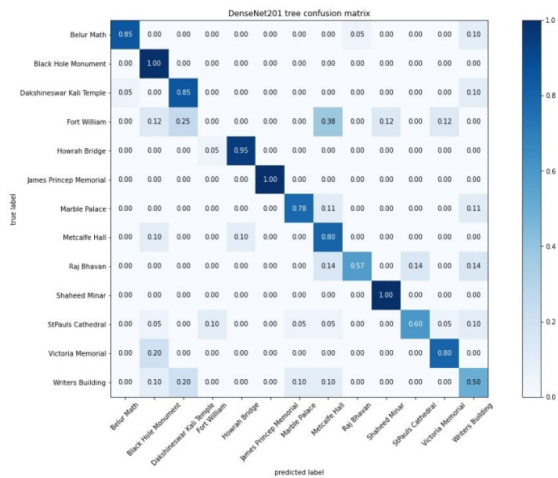
We followed common baselines for the comparison purpose. Both models were trained and tested using an 80–20 split of the dataset, with identical training procedures to ensure a fair comparison. The experiments were conducted on a machine with an NVIDIA GTX 1080 Ti GPU, 32 GB RAM, and an Intel i7 processor. MonuNet outperformed DenseNet in terms of accuracy, computational complexity, and inference time while using fewer parameters (ref. Table 3). This demonstrates MonuNet’s efficiency and effectiveness in classifying heritage site images with a lower computational burden.

Table 4 Performance comparison with related DenseNet models

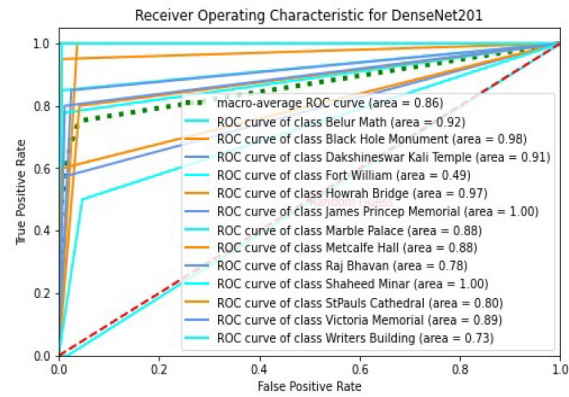
Methods	Accuracy (%)	Precision (%)	Recall (%)
MonuNet	89	87.77	86.61
DenseNet201 [27]	79	73.46	74.61
DenseNet169 [28]	85	83.23	82.62
DenseNet121 [29]	85	83.15	81.62

Looking into the performance of various DenseNet models (see Table 4), it is clear that MonuNet outshines other DenseNet counterparts with the superior accuracy of 89%. The highest precision of 87.77% depicts its effectiveness in reducing the false positives. Also, MonuNet provides a recall of 86.62% and it shows the vital nature to enhance the true positives. But, DenseNet201 shows a little lag in the performance with scores-accuracy-79% and F1score-73.23%. It reveals that the balance between precision and recall is not well maintained as such MonuNet. DenseNet169 and DenseNet121 shares similar accuracy score of 85%. The slight variations in the depth of network, especially from DenseNet169 to DenseNet201 do not provide any advantage in rising the accuracy. Therefore, in practical applications where computational resources are a limiting factor, the marginal performance trade-off seen in DenseNet121 may justify its deployment over the more computationally demanding MonuNet.

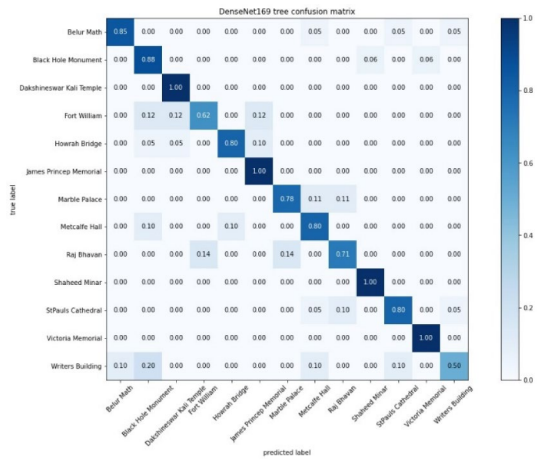
When comparing the performance (See Fig. 8) of these models, DenseNet169 and DenseNet121 stand out as



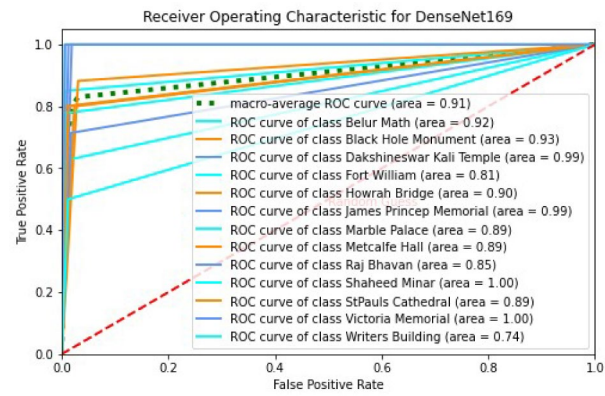
(a) Confusion matrix - DenseNet201



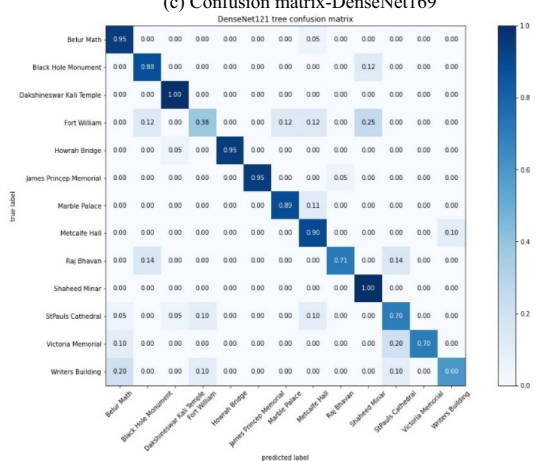
(b) RoC-DenseNet201



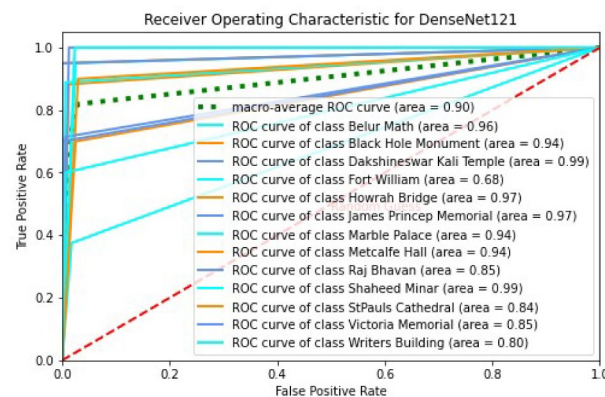
(c) Confusion matrix-DenseNet169



(d) RoC-DenseNet169



(e) Confusion matrix-DenseNet121



(f) RoC-DenseNet121

Fig. 8 Plots for RoC and confusion matrix

more accurate and reliable classifiers than DenseNet201, based on the AUC values and confusion matrices. The precision in classification and the ability to discern between classes, as observed in the confusion matrices, along with the ROC curves, point towards DenseNet169 and DenseNet121 as favorable choices for Kolkata heritage image classification tasks where accurate class differentiation is crucial. DenseNet201, while still a competent model, may require further tuning or class-specific threshold adjustments to improve its performance in line with its counterparts. Comparing MonuNet against DenseNet201, DenseNet169, and DenseNet121 reveals distinct performance trends. MonuNet's ROC curve boasts a macro-average AUC of 0.93, reflecting its robust discriminative power, which surpasses DenseNet201's 0.86 and is on par with DenseNet169 and DenseNet121's 0.91 and 0.90 respectively. The confusion matrix for MonuNet indicates high precision in class-specific predictions, with several classes achieving near-perfect recognition. Notably, MonuNet demonstrates fewer misclassifications for challenging classes compared to DenseNet201, suggesting more reliable class separability. Overall, MonuNet's superior ROC AUC and precise classification, as evidenced in its confusion matrix, underscore its effectiveness for tasks requiring high-fidelity image classification.

Conclusion and future work

We introduced MonuNet, a specialized deep learning network tailored for Kolkata's heritage image classification, trained on a curated dataset of 13 key monuments. Through rigorous evaluations, MonuNet achieved over 80% accuracy across most classes, outperforming existing DenseNet models. Our Pipeline promises to digitally preserve and recognize Kolkata's cultural heritage effectively. Looking forward, we aim to diversify and expand the dataset, refine the model with advanced attention mechanisms, and incorporate user feedback for iterative improvements. Deploying our framework as an interactive web application can bolster tourism, education, and cultural preservation efforts. Collaborative partnerships with local organizations and academic institutions will ensure the dataset's continuous enrichment and relevance. Future endeavors also include optimizing the web application's user interface and integrating augmented reality features for an immersive heritage exploration experience. In conclusion, MonuNet mark a significant advancement in leveraging deep learning for Kolkata's cultural heritage preservation, with potential applications benefiting researchers, heritage enthusiasts, and the broader community.

Author contributions

Conceptualization—Sasithradevi A, experimentation—Sabari Nathan, implementation—Chanthini. B, results evaluation—Subbulakshmi. T, paper drafting—P.Prakash. All authors reviewed the manuscript.

Funding

Open access funding provided by Vellore Institute of Technology.

Availability of data and materials

Data will be made available based on reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Centre for Advanced Data Science, Vellore Institute of Technology, Chennai, TamilNadu, India. ²Couger Inc, Shibuya-Ku, Tokyo, Japan. ³School of Electronics Engineering, Vellore Institute of Technology, Chennai, TamilNadu, India. ⁴School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, TamilNadu, India. ⁵Department of Electronics Engineering, MIT, Anna University, Chennai, TamilNadu, India.

Received: 10 April 2024 Accepted: 22 June 2024

Published online: 16 July 2024

References

1. City of joy. https://iiche.org.in/chemcon2023/Explore_Bengal.pdf.
2. Biswas A. Continuity in tradition—a narrative on the cultural heritage of para and adda in Kolkata. *Urban Des Plan*. 2021. <https://doi.org/10.1680/jurp.21.00016>.
3. Government of Bengal. West Bengal Heritage Commission. <https://wbhc.in/home/landing>.
4. Chowdhury S. Kolkata's Heritage status: the question of survival. 2019. <https://doi.org/10.13140/RG.2.2.21451.44320>.
5. Yunari N, Yuniarno EM, Purnomo MH. Indonesian batik image classification using statistical texture feature extraction Gray Level Co-occurrence Matrix (GLCM) and Learning Vector Quantization (LVQ). *J Telecommun Electron Comput Eng*. 2018;10:67–71.
6. Carriero VA, Gangemi A, Mancinelli ML, Marinucci L, Nuzzolese AG, Presutti V, et al. ArCo: the Italian cultural heritage knowledge graph. Cham: Springer International Publishing; 2019. p. 36–52.
7. Giulio, R, Maietti, F, Plaia, E, Medici, M, Ferrari, F, & Turillazzi, B. Integrated data capturing requirements for 3d semantic modelling of cultural heritage: the inception protocol, *ISPRS - Int Arch Photogramm Remote Sens Spat Inf Sci*. <https://doi.org/10.5194/isprs-archives-XLII-2-W3-251-2017>. 2017.
8. Kavitha S, Mohanavalli S, Bharathi B, Rahul CH, Shailesh S, Preethi K. Classification of Indian monument architecture styles using bi-level hybrid learning techniques. Singapore: Springer Nature Singapore; 2022. p. 471–88.
9. Llamas J, Lerones PM, Medina R, Zalama E, Gómez-García-Bermejo J. Classification of architectural heritage images using deep learning techniques. *Appl Sci*. 2017;7:992.
10. Cosovic M, Jankovic R. CNN classification of the cultural heritage images. In: 2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH). IEEE; 2020. p. 1–6.
11. Kulkarni U, Meena SM, Gurlahosur SV, Mudengudi U. Classification of cultural heritage sites using transfer learning. In: 2019 IEEE Fifth International Conference on Multimedia Big Data. IEEE. 2019. p. 391–7.
12. Fan T, Wang H, Deng S. Intangible cultural heritage image classification with multimodal attention and hierarchical fusion. *Expert Syst Appl*. 2023;231:120555.
13. Dou J, Qin J, Jin Z, Li Z. Knowledge graph based on domain ontology and natural language processing technology for Chinese intangible cultural heritage. *J Vis Lang Comput*. 2018;48:19–28.

14. Janković R. Machine learning models for cultural heritage image classification: comparison based on attribute selection. *Information*. 2019;11:12.
15. Sasithradevi A, Sabarinathan SS, Roomi SMM, Prakash P. KolamNetV2: efficient attention-based deep learning network for tamil heritage art-kolam classification. *Herit Sci*. 2024;12:60. <https://doi.org/10.1186/s40494-024-01167-8>.
16. Prasomphan S. Toward fine-grained image retrieval with adaptive deep learning for cultural heritage image. *Comput Syst Sci Eng*. 2023;44:1295–307.
17. Ma K, Wang B, Li Y, Zhang J. Image retrieval for local architectural heritage recommendation based on deep hashing. *Buildings*. 2022;12:809.
18. Grilli E, Özdemir E, Remondino F. Application of machine and deep learning strategies for the classification of heritage point clouds. *Int Arch Photogramm Remote Sens Spat Inf Sci*. 2019;XLII-4/W18:447–54.
19. Pierdicca R, Paolanti M, Matrone F, Martini M, Morbidoni C, Malinverni ES, et al. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sens*. 2020;12:1005.
20. Cardelliccchio A, Ruggieri S, Nettis A, Renò V, Uva G. Physical interpretation of machine learning-based recognition of defects for the risk management of existing bridge heritage. *Eng Fail Anal*. 2023;149:107237.
21. Law S, Seresinhe CI, Shen Y, Gutierrez-Roig M. Street-Frontage-Net: urban image classification using deep convolutional neural networks. *Int J Geogr Inf Sci*. 2020;34:681–707. <https://doi.org/10.1080/13658816.2018.1555832>.
22. Triantis D, Pasiou ED, Stavrakas I, Kourkoulis SK. New perspectives in structural health monitoring of restored elements of cultural heritage monuments. *Procedia Struct Integr*. 2024;55:185–92.
23. Mishra M. Machine learning techniques for structural health monitoring of heritage buildings: a state-of-the-art review and case studies. *J Cult Herit*. 2021;47:227–45.
24. Zou H, Ge J, Liu R, He L. Feature recognition of regional architecture forms based on machine learning: a case study of architecture heritage in Hubei Province, China. *Sustainability*. 2023;15:3504.
25. Murugesan S, Ramshankar N, Hiba Mariam HKP, Kk A. Heritage identification of monuments using deep learning techniques. *J Data Acquis Process*. 2023;38:1927–35.
26. Abed MH, Al-Asfoor M, Hussain ZM. Architectural heritage images classification using deep learning with CNN [Paper presentation]. Proceedings of the 2nd International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, Bari, Italy. January, 2020 <http://ceur-ws.org/Vol-2602/>.
27. Salim F, Saeed F, Basurra S, Qasem SN, Al-Hadhrami T. DenseNet-201 and xception pre-trained deep learning models for fruit recognition. *Electronics*. 2023;12:3132.
28. Dalvi PP, Edla DR, Purushothama BR. Diagnosis of coronavirus disease from chest X-ray images using DenseNet-169 architecture. *SN Comput Sci*. 2023;4:214.
29. Cinar N, Ozcan A, Kaya M. (2022). A hybrid DenseNet121 Coronavirus disease from chest x-ray images using densenet-169 architecture. *Sn Comput. Sci*. 4: 214 (2023). <https://doi.org/10.1007/s42979-022-01627-7>
30. UNet model for brain tumor segmentation from MR Images. *Biomedical Signal Processing and Control*. 76: 103647.
30. Mobile Application. <https://attractions.io/use-case/mobile-apps-for-heritage-and-cultural-attractions>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.