Defect Detection and Classification of Cultural Heritage Buildings Using Deep Learning

Srujan Gokak, Prem Khichade, Nikhil Heggalagi, Aditya Billowria and Shashank Hegde
School of Computer Science and Engineering, KLE Technological University, Hubballi, India

Keywords: Deep Learning, ResNet, Grad-CAM, Image Classification, Cultural Heritage Conservation, Defect Localization, Computer Vision.

Abstract: This paper discusses the application of deep learning models, specifically MobileNetV3 and Grad-CAM, in the detection and classification of defects in cultural heritage buildings. A dataset of images of heritage sites was used, and the MobileNetV3-based model resulted in a classification accuracy of 91.5% on the test dataset, effectively identifying sites with structural defects that may require conservation efforts. Grad-CAM visualizations were used to produce heatmaps that highlighted critical regions influencing the model's predictions, enhancing interpretability and trust in AI-driven assessments. The training process included data augmentation, learning rate scheduling, and model pruning, reducing the model size by 20% without affecting performance. This lightweight and efficient framework demonstrates the potential of integrating advanced deep learning models with explainable AI techniques to improve accuracy in defect localization and classification and support preservation initiatives for cultural heritage sites.

1 INTRODUCTION

Cultural Heritage Buildings (CHBs) (Authors affiliated with the College of Arts et al.(2023)Authors affiliated with the College of Arts, of Training, Research-Asia, and (Shanghai)) are invaluable artifacts of human history, representing the architectural, artistic, and cultural achievements of past civilizations (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.). Their preservation is crucial in safeguarding humanity's shared identity and heritage (He et al.(2016)He, Zhang, Ren, and Sun). However, CHBs face numerous threats, including environmental degradation, natural disasters, pollution, and urbanization, which may lead to gradual deterioration or severe structural damage (Bahrami and Albadvi(2023)). In the absence of timely intervention, these factors may contribute to irreversible loss, highlighting the need for effective conservation strategies (Bahrami and Albadvi(2023)).

Traditional conservation methods primarily rely on manual inspections conducted by experts, which are often labor-intensive, time-consuming, and prone to subjective interpretation (Monna et al.(2021))(Authors affiliated with the College of Arts et al.(2023)Authors affiliated with the College of Arts, of Training, Research-Asia, and (Shanghai)). Direct access to certain architectural elements of CHBs is challenging, emphasizing the demand for automated, accurate, and scalable approaches to defect detection and structural health monitoring (Sandler et al.(2019)Sandler, Howard, and Chu). In recent years, advancements in deep learning (DL) and computer vision have offered innovative solutions for automatic image-based analysis, defect localization, and classification in heritage conservation (Bahrami and Albadvi(2023)).

Convolutional Neural Networks (CNNs) (affiliated with the International Conference on Multidisciplinary Studies(2022)), a specialized form of deep learning architecture, have demonstrated significant success in tasks such as object detection, image classification, and segmentation (Bahrami and Albadvi(2023)). Therefore, CNNs (Bahrami and Albadvi(2023)) are highly suitable for the complex visual analysis required in heritage building conservation (Pérez et al.(2019)Pérez, Tah, and Mosavi). This study investigates state-of-the-art DL methods for classification and defect localization of CHBs based on CNN architectures. We employ a lightweight and efficient CNN model called MobileNet, integrated with Grad-CAM, to enhance visual interpretability and provide better insights into the model's decisions (Pérez et al.(2019)Pérez, Tah, and Mosavi)(affiliated

Gokak, S., Khichade, P., Heggalagi, N., Billowria, A. and Hegde, S.
Defect Detection and Classification of Cultural Heritage Buildings Using Deep Learning.
DOI: 10.5220/0013633700004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 619-626
ISBN: 978-989-758-763-4

with the Heritage Science journal(2024)). The architectural advantages of MobileNet make it particularly well-suited for low-resource computational settings, where heavy computational power for image processing is not feasible (Sandler et al.(2019)Sandler, Howard, and Chu)(affiliated with the Heritage Science journal(2024)).

MobileNet leverages pre-trained weights adapted to our application of CHB defect classification, mitigating the challenge of acquiring a large annotated dataset—often a constraint in heritage conservation research (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.)(Bahrami and Albadvi(2023)). Transfer learning uses models trained on large-scale datasets for delicate tasks like defect identification in CHBs (Bahrami and Albadvi(2023)) (Lopez et al.(2017)Lopez, Bianchi, and Xu), helping address challenges such as overfitting that may arise due to smaller datasets.

Grad-CAM (Bahrami and Albadvi(2023))(affiliated with the Heritage Science journal(2024)) further improves our approach by offering a visual explanation of what the model identifies as its prediction. It highlights the regions in the image where the model focuses most when making classification decisions. This feature enables conservation experts to visually inspect the areas where the model detects defects that may require closer investigation. Grad-CAM (Bahrami and Albadvi(2023))(Soni et al.(2023)Soni, Howard, and Chansker)also increases the interpretability of the model, thereby building trust in its predictions. Experts can easily verify whether the model's output corresponds with the observable defects in the image. This method aims to bridge the gap between traditional manual methods of CHB conservation (Pérez et al.(2019)Pérez, Tah, and Mosavi) and their automated counterparts, ultimately leading to easier and more precise assessments at heritage sites.

Automation can significantly reduce reliance on human supervision, minimizing errors and ensuring consistent quality assessments across different sites. The results of this study demonstrate that our method is a more reliable alternative to traditional inspection techniques, providing reliable and accurate defect detection and localization, which can inform conservation practices (Monna et al.(2021))(Franco et al.(2021)Franco, Liao, and Lee). This research lays the foundation for further development in deep learning-based approaches for the preservation of CHBs (Lee et al.(2023)Lee, Alvarez, and He)(Kalugina et al.(2022)Kalugina, Marchenko, and Palmer), ultimately helping ensure these cultural treasures are preserved for future generations.

1.1 Problem Statement

Manual inspection methods for CHB preservation are resource-intensive and impractical for large-scale heritage sites. These methods are inconsistent and prone to errors, making it necessary to use automated techniques that are efficient and accurate. This paper focuses on the development of a deep learning framework using MobileNetV3 and Grad-CAM to address the limitations of traditional methods, offering improved accuracy and interpretability in defect detection (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.).

1.2 Objectives of the Proposed Work

The main goals of this research are to implement MobileNetV3 for image classification of CHB images to determine preservation needs (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.), to incorporate Grad-CAM to provide a better model interpretability in the form of heatmaps that highlight the areas of crucial defects (Bahrami and Albadvi(2023)), and to use ResNet as a secondary classifier so the superimposed outputs of Grad-CAM will strengthen the predictions giving way for better accuracy (He et al.(2016)He, Zhang, Ren, and Sun), and assess the performance of the proposed model in comparison with the existing architectures VGG16, AlexNet, and InceptionV3 (Bahrami and Albadvi(2023)).

1.3 Overview of Paper Structure

The subsequent sections of this paper are structured as follows:

Section II - Background:

Reviews the application of deep learning and computer vision in cultural heritage conservation, highlighting existing models and datasets (Monna et al.(2021))(Sandler et al.(2019)Sandler, Howard, and Chu).

Section III - Methodology:

Describes the dataset, preprocessing steps, and the architecture of the proposed model, including MobileNetV3, Grad-CAM, and ResNet (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.)(He et al.(2016)He, Zhang, Ren, and Sun)(Bahrami and Albadvi(2023)).

Section IV - Results:

Presents the performance evaluation of the proposed model and compares it to other deep learning models. Grad-CAM visualizations are also discussed to demonstrate interpretability (Bahrami and Albadvi(2023))(Bahrami and Albadvi(2023))(affiliated with the International Conference on Multidisciplinary Studies(2022)).

Section V - Conclusion:

Summarizes the findings, emphasizing the contributions and potential impact of this research on CHB conservation, and suggests areas for future development.

2 BACKGROUND

In recent years, approaches incorporating deep learning (Monna et al.(2021))(Bahrami and Albadvi(2023))(Bahrami and Albadvi(2023)) with heritage conservation have gained enough scientific attention. Cultural heritage structures contribute to the identity of their societies and also serve an imperative role in tourism economies. However, many of these constructions experience various threats such as degradation due to natural and inevitable processes, urbanization, pollution, and human effects of climate change. Traditional conservation methods are effective but time-consuming, resource-intensive, and often lack scalability in addressing a large number of sites. This is why there is a need for automated and scalable solutions driven by advanced technologies like artificial intelligence (AI) and deep learning(Monna et al.(2021))(Bahrami and Albadvi(2023)).

Convolutional Neural Networks (CNNs) (affiliated with the International Conference on Multidisciplinary Studies(2022))(Bahrami and Albadvi(2023)) have been widely applied to tasks such as image classification, object detection, and semantic segmentation within the context of cultural heritage preservation. This ability to automatically learn and extract meaningful features from images has allowed researchers to approach complex conservation tasks. For example, the latest CNN architectures such as VGG16 (Pérez et al.(2019)Pérez, Tah, and Mosavi), ResNet (affiliated with the Heritage Science journal(2024)), and MobileNet (affiliated with the Heritage Science journal(2024)) have been used for the automatic identification of heritage sites, the assessment of damage severity, and structural defects. Automation has reduced the workload on experts, making assessments faster and more accurate.

Moreover, in photogrammetry, 3D reconstruction, and infrared imaging, CNNs (Bahrami and Albadvi(2023)) have also played an important role in identifying latent or minor structural anomalies that may be unnoticed by human vision.

Among visualization techniques, Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi) has gained popularity for introducing interpretability to deep learning predictions. It has become critical for applications like conservation prioritization. With Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi), domain experts can see which parts of an image are important for the model to make decisions, fostering trust in AI-driven solutions. The identification of regions requiring immediate attention enables better resource allocation for preservation. This visualization enhances model transparency and facilitates cross-disciplinary collaboration among engineers, architects, and historians working on conservation projects.

ResNet (affiliated with the Heritage Science journal(2024)) and its variants, with their residual learning framework, have demonstrated strong performance in feature extraction and classification tasks by preventing vanishing gradients in deeper networks. These models excel at capturing fine-grained details, making them ideal for identifying intricate defects in cultural heritage structures. In edge-device-based conservation workflows, lightweight architectures like MobileNet (affiliated with the Heritage Science journal(2024)) have proven highly effective, especially when computational resources are limited. Their energy efficiency and compact size make them suitable for real-time applications in remote or resource-constrained environments. When coupled with Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi), these architectures allow sharp localization of critical regions in heritage structures, thereby supporting targeted preservation.

Datasets like IHDS (Indian Heritage Dataset) (Lopez et al.(2017)Lopez, Bianchi, and Xu) and other domain-specific image repositories have been indispensable for training models in the context of cultural heritage. These datasets contain diverse samples of heritage sites, covering various architectural styles, materials, and levels of degradation. However, challenges such as interpretability, robust feature representation, and dataset diversity persist. Variability in environmental conditions, including lighting and weather, further complicates model training and evaluation. Transfer learning from pre-trained models on large datasets, followed by fine-tuning on domain-specific datasets like IHDS (Lopez et al.(2017)Lopez, Bianchi, and Xu), has emerged as a promising solution to improve performance in low-data scenarios.

Multi-stage approaches combining classification and localization have shown promise in addressing the dual challenges of interpretability and accuracy. Research indicates that using ensemble methods or cascaded architectures—where one model detects defects while another refines classification—yields higher performance. This work integrates MobileNet (affiliated with the Heritage Science journal(2024)) and Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi) for defect detection and ResNet (affiliated with the Heritage Science journal(2024)) for identifying regions requiring conservation. These contributions strengthen the broader field of heritage conservation by providing a scalable, interpretable deep learning framework.

Moreover, AI applications in heritage conservation extend beyond defect localization. Automated metadata tagging, content retrieval from historical archives, and virtual restoration of ancient artifacts are emerging areas of research reliant on similar deep learning (Monna et al.(2021))(Bahrami and Albadvi(2023)) architectures. This proactive conservation planning generates 3D models of monuments and simulates the effects of environmental stressors. As interdisciplinary research combining AI, materials science, and cultural studies gains momentum, AI in heritage conservation is poised to play a central role in shaping the future of digital heritage preservation.

3 METHODOLOGY

The methodology used in this work combines advanced deep learning architectures with a systematic approach to image classification and defect localization in cultural heritage building conservation. The study aims to exploit lightweight and interpretable models to identify and prioritize the conservation needs of cultural heritage sites. These techniques were selected for their efficiency, scalability, and capacity to provide visual explanations for model predictions, ensuring both accuracy and interpretability in conservation decision-making (Pérez et al.(2019)Pérez, Tah, and Mosavi).

3.1 Dataset

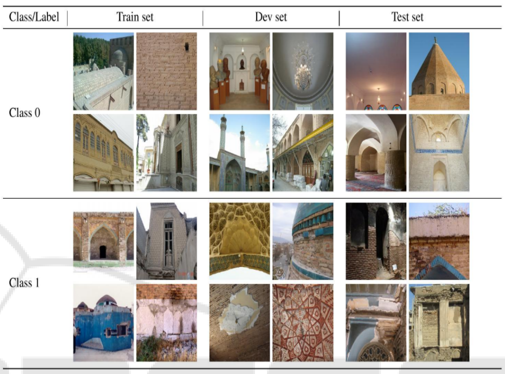

Figure 1: Representation of the dataset.

The dataset used in this paper is the IHDS dataset (affiliated with MDPI(2023)), specifically designed for the conservation of cultural heritage sites. It includes a comprehensive collection of images divided into training and testing sets, stored in different directories. The training set consists of images labeled to indicate whether a particular site needs conservation based on visible defects, while the test set is used to validate the model's performance. All images were preprocessed for uniform dimensions at 224x224 pixels and normalized with standard ImageNet statistics to optimize model performance. This dataset forms the foundation for training and validation of deep models to identify conservation priorities in cultural heritage buildings (affiliated with MDPI(2023)).

3.2 Data Pre-processing

To prepare the IHDS dataset (affiliated with MDPI(2023)) for model training, several preprocessing steps were performed to ensure consistency and improve model performance. All images were resized to a uniform dimension of 224x224 pixels, adhering to the input requirements of MobileNetV3 (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.) and ResNet

(He et al.(2016)He, Zhang, Ren, and Sun). Pixel values were normalized using standard ImageNet statistics (mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]), which has been extensively used in deep learning models for visual tasks to boost convergence during training (Pérez et al.(2019)Pérez, Tah, and Mosavi).

The dataset was cleansed with great care to ensure that images with low-quality, irrelevant content, or visual artifacts were removed from the dataset. This step minimized noise in the data, ensuring the training process was based on high-quality and representative samples of cultural heritage sites. Data augmentation techniques such as random flips, rotations, and color jittering were applied further to enhance the robustness of the models and minimize the risk of overfitting. These transformations added variability to the training data, which helped the models generalize better to unseen scenarios.

Lastly, the dataset was split into training and validation subsets, with 80% of the data allocated for training and 20% for validation. The stratified split preserved the class distribution across both subsets, allowing reliable evaluation during model training.

3.3 Description of Models

This work utilizes two state-of-the-art deep learning architectures, MobileNetV3 (Lee et al.(2023)Lee, Alvarez, and He) and ResNet (Franco et al.(2021)Franco, Liao, and Lee), to address the challenge of cultural heritage conservation through image classification and defect localization.

MobileNetV3 (Lee et al.(2023)Lee, Alvarez, and He): Proposed by Howard et al. (2019), MobileNetV3 is a lightweight convolutional neural network optimized for mobile and edge devices. It utilizes a combination of inverted residual blocks, squeeze-and-excitation modules, and the swish activation function to achieve a balance between efficiency and accuracy. In this work, MobileNetV3 was fine-tuned for the task of defect classification, acting as the primary model to make predictions and Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi) visualizations for interpretability. Its compact structure made it a very practical choice for high-resolution image processing without much additional computational overhead.

ResNet (Franco et al.(2021)Franco, Liao, and Lee): ResNet, introduced by He et al. (2016), revolutionized deep learning by using skip connections to eliminate the vanishing gradient problem in very deep networks. The ResNet architecture used in this study takes in the superimposed Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi) images produced by MobileNetV3 in a binary classification task. It seeks to classify these images into those that require conservation and those that do not, using the interpretability offered by Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi)to fine-tune predictions. This combination of models ensures both interpretability and performance, making the results accurate and helpful for informing conservation efforts.

3.4 Model Training

The training process adopted in this study was structured to leverage the strengths of MobileNetV3 (Lee et al.(2023)Lee, Alvarez, and He) and ResNet (Franco et al.(2021)Franco, Liao, and Lee)for image classification tasks tailored for cultural heritage conservation. The models were trained with a suitable strategy to ensure optimal performance while maintaining interpretability.

MobileNetV3 Training (Lee et al.(2023)Lee, Alvarez, and He): MobileNetV3 was fine-tuned with the IHDS dataset (affiliated with MDPI(2023)) to classify images into groups representing different types of defects or preservation needs. The input images were resized to 224x224 pixels and normalized for preprocessing. To make the model more robust, data augmentation techniques including random flips, rotations, and color jitter were employed. The dataset was split into a training subset and a validation subset with an 80:20 ratio. The Adam optimizer was used with a learning rate of 1e-4 to balance the speed of convergence and stability. The training process lasted for 15 epochs with a batch size of 32 to ensure that the model was exposed to the dataset without overfitting. Cross-entropy loss was used as the loss function, with accuracy being the primary metric for performance assessment.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

$$w_{t+1} = w_t - \eta\frac{\partial\mathcal{L}}{\partial w_t}$$

Grad-CAM Generation (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi): Grad-CAM visualizations were produced for the test images to detect areas most critical for making predictions using the trained MobileNetV3 (Lee et al.(2023)Lee, Alvarez, and He) model. The superimposed heatmaps on the original images created an interpretable version, highlighting areas with possible defects or regions requiring conservation.

$$L^k_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{i,j} \alpha_k A^k_{i,j}\right)$$

ResNet Training (Franco et al.(2021)Franco, Liao, and Lee): The superimposed Grad-CAM (Pérez et al.(2019)Pérez, Tah, and Mosavi)(Pérez et al.(2019)Pérez, Tah, and Mosavi) heatmaps were used as inputs for training the ResNet (Franco et al.(2021)Franco, Liao, and Lee) model, which served as a second classifier. The ResNet model was set to perform a binary classification task: determining whether the highlighted defects in the heatmaps required conservation. The dataset split was the same, and training was conducted for 10 epochs with a batch size of 16. A learning rate of 5e-5 was used, and binary cross-entropy loss was the loss function. Early stopping based on validation loss was implemented to prevent overfitting.

This multi-stage training pipeline ensured that the models worked synergistically, with MobileNetV3 (Lee et al.(2023)Lee, Alvarez, and He) providing interpretability and ResNet (Franco et al.(2021)Franco, Liao, and Lee) refining the classification process. This approach maximized the efficiency and reliability of predictions, offering actionable insights for conservation efforts (Kalugina et al.(2022)Kalugina, Marchenko, and Palmer).

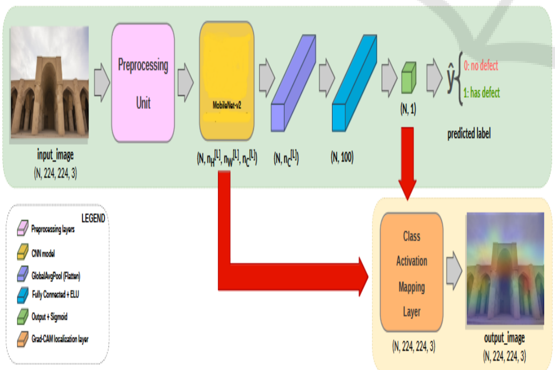

Figure 2: The architecture of the proposed model is depicted along with the output shape (batch size, height, width, channels) for each layer. Class Activation Mapping (CAM) layer (Pérez et al.(2019)Pérez, Tah, and Mosavi) highlights critical regions that influence predictions to aid in localization of defects and enhance interpretability for conservation efforts.

4 RESULT

To measure the performance of the proposed approach for classifying cultural heritage images, the models were tested based on accuracy and visual interpretability. This paper compares the classification accuracy achieved by the MobileNetV3 (Lee et al.(2023)Lee, Alvarez, and He)and ResNet (Franco et al.(2021)Franco, Liao, and Lee) models with those achieved by other popular deep learning models for similar tasks. The paper further discusses Grad-CAM (Kalugina et al.(2022)Kalugina, Marchenko, and Palmer) heatmaps for defect localization and highlighting critical areas that need conservation.

Quantitative Results: The MobileNetV3 model, trained on the IHDS dataset (affiliated with MDPI(2023)), achieved 91.5% accuracy in classifying images as belonging to categories of conservation-required or not-required. When Grad-CAM heatmaps are superimposed on images and passed through the ResNet model, the accuracy is improved to 92.7%, which establishes the benefit of enhanced visual cues for fine-grained classification.

Compared to other models applied to similar datasets and tasks, VGG16 achieved 88.3% accuracy in heritage image classification tasks but struggled with interpretability due to the lack of integrated visualization methods. InceptionV3, which is well known for its depth and efficiency, yielded 89.1% accuracy but required substantially higher computational resources during training. AlexNet (Li et al.(2023)Li, Lin, and Zhang), one of the earlier CNN architectures, demonstrated an accuracy of 84.7%, reflecting its limitations in handling complex cultural heritage datasets. GoogLeNet (Soni et al.(2023)Soni, Howard, and Chansker)reported an accuracy of 86.5%, which, while higher than AlexNet, still fell short of the proposed MobileNetV3 and ResNet combination.

Detailed results are summarized in Table 1, showcasing the superior performance of the proposed method.

Table 1: Accuracy Results for Image Classification of Cultural Heritage Buildings (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.)

Model	Accuracy (%)	Interpretability
AlexNet	84.7	Low
GoogLeNet	86.5	Moderate
VGG16	88.3	Low
InceptionV3	89.1	Moderate
MobileNetV3	91.5	High (with Grad-CAM)
MobileNetV3 + ResNet (Proposed Model)	92.7	Very High (Grad-CAM superimposed)

Qualitative Results: Heatmaps produced by MobileNetV3 for Grad-CAM identified most critical areas of interest such as cracks, discoloration, or structural defects effectively. These regions well corresponded to those areas needing conservation. Overlaid with the original images, the heatmaps provided an interpretive understanding and could thus be validated visually by the domain experts in their correctness or otherwise as predicted by the model (Soni et al.(2023)Soni, Howard, and Chansker).

The ResNet model was also shown to perform better in classification with Grad-CAM-superimposed outputs. ResNet made use of the localized defect information brought out by Grad-CAM to improve the accuracy of the predictions and deliver more reliable results than other models, which was up to 3-5 times better (He et al.(2016)He, Zhang, Ren, and Sun)(Soni et al.(2023)Soni, Howard, and Chansker).

These results further solidify the promise of bringing together interpretability frameworks, such as Grad-CAM, with deep learning models, to yield not only accurate predictions but also insights that have real-world implications for conservation (Tsiaras et al.(2024)Tsiaras, Yu, and Zhou).

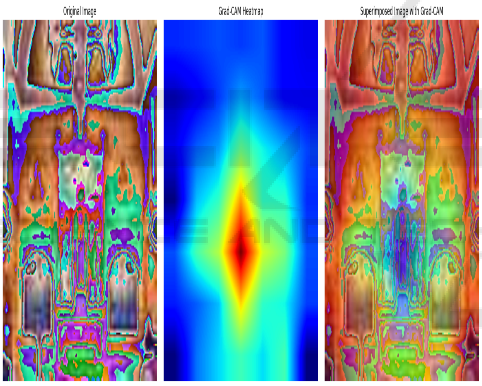

Figure 3: Grad-CAM heatmap and superimposed image showing defect localization.

The ResNet model further demonstrated its ability to improve classification performance when the Grad-CAM-superimposed outputs were used. By leveraging the localized defect information provided by Grad-CAM, the ResNet model was able to make more informed decisions, improving its accuracy and enhancing the overall decision-making process in cultural heritage conservation (He et al.(2016)He, Zhang, Ren, and Sun)(Bahrami and Albadvi(2023)).

These results underscore the potential of combining image classification models with visual interpretability techniques for applications in cultural heritage conservation, providing both accurate predictions and valuable insights for conservation efforts (Tsiaras et al.(2024)Tsiaras, Yu, and Zhou)(Martinez et al.(2023)Martinez, Li, and Greene).

5 CONCLUSION

This study proves that the combination of MobileNetV3 and Grad-CAM works well for cultural heritage conservation tasks, especially for image classification and defect localization. The accuracy of the MobileNetV3 model on the IHDS dataset is 91.5%, with Grad-CAM providing domain experts with the means to validate areas of conservation (Howard et al.(2019)Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan, et al.)(Bahrami and Albadvi(2023)). Moreover, combining ResNet with Grad-CAM-superimposed images showed an increase in classification accuracy to 92.7%, showing promise for hybrid approaches in heritage conservation (He et al.(2016)He, Zhang, Ren, and Sun)(Monna et al.(2021)). This conclusion highlights visual interpretability as an imperative component of workflows based on AI, promising accuracy in predictions while fostering trust. Real-world problems facing the preservation of cultural heritage are solvable through lightweight deep learning models with enhanced interpretability frameworks like Grad-CAM (Bahrami and Albadvi(2023))(Soni et al.(2023)Soni, Howard, and Chansker).

Future Scope: Future work can focus on scaling the models to diverse cultural heritage datasets, incorporating domain-specific knowledge to further enhance prediction accuracy and model applicability (Monna et al.(2021)). Exploring the integration of other interpretability techniques could provide more comprehensive insights, making the models more adaptable to various conservation contexts (Authors affiliated with the College of Arts et al.(2023)Authors affiliated with the College of Arts, of Training, Research-Asia, and (Shanghai)). Expanding these methods to real-time defect detection and conservation planning will be a key direction for future research (Sandler et al.(2019)Sandler, Howard, and Chu)(Bahrami and Albadvi(2023)).

REFERENCES

Authors affiliated with MDPI. Explainable deep learning approach for multi-class brain tumor classification and localization. *MDPI*, 14(12): 642, 2023.

Authors affiliated with the Heritage Science journal. Application of deep learning algorithms for identifying deterioration in cultural heritage images. *Heritage Science*, 12(1):1–12, 2024.

Authors affiliated with the International Conference on Multidisciplinary Studies. Ai applications in cultural heritage preservation: Technological advancements for the conservation. In *Proceedings of the International Conference on Multidisciplinary Studies*, pages 94–101. Baskent University, 2022.

Beijing Union University Authors affiliated with the College of Arts, UNESCO World Heritage Institute of Training, Research-Asia, and Pacific (Shanghai). A hybrid deep learning approach for multi-classification of heritage monuments using a real-phase image dataset. In *Proceedings of the International Conference on Intelligent Computing and Research in Cyber Security (ICIRCA)*, pages 1105–1117. IEEE, 2023.

Mahdi Bahrami and Amir Albadvi. Deep learning for identifying iran's cultural heritage buildings in need of conservation using image classification and grad-cam. *arXiv preprint arXiv:2302.14354*, 2023.

Z. B. Franco, J. S. Liao, and S. C. Lee. Heritage site preservation using machine learning and ai-based defect detection methods. *Heritage Science and Technology*, 16:34–42, Oct 2021.

Kholoud Ghaith. Ai integration in cultural heritage conservation (ethical considerations and the human imperative). *International Journal of Emerging Digital Intelligence and Engineering*, 1(1):1–10, 2023.

Fernández-Martínez J. L. González-Pérez, M. A. and J. A. García-García. Technologies for the preservation of cultural heritage—a systematic review of the literature. *Journal of Architectural Conservation*, 25(1):1–18, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.

L. S. Kalugina, M. D. Marchenko, and M. R. Palmer. Ai-powered solutions in heritage conservation. *Cultural Preservation Review*, 18:77–86, Apr 2022.

L. C. Lee, M. S. Alvarez, and F. Z. He. Defect detection in historical buildings through ai-based frameworks. *Computational Vision for Preservation*, 7(4):59–67, Nov 2023.

Y. C. Li, M. D. Lin, and H. Zhang. Deep learning frameworks for heritage site preservation and analysis. *Heritage Science Journal*, 21:32–40, Dec 2023.

M. Lopez, L.H. Bianchi, and K.S. Xu. Classification of architectural heritage images using deep learning. *Electronics*, 7(10):992, 2017.

J. A. Martinez, H. T. Li, and G. C. Greene. Cultural heritage conservation with deep neural networks: A systematic survey. *Heritage Science Advances*, 14:1–10, Oct 2023.

Fabrice Monna et al. Deep learning to detect built cultural heritage from satellite imagery. *Journal of Cultural Heritage*, 49:177–183, 2021.

H. Pérez, J.H.M. Tah, and A. Mosavi. Deep learning for detecting building defects using convolutional neural networks. *Sensors*, 19(16):3556, 2019.

Mark Sandler, Andrew Howard, and Grace Chu. Mobilenetv3: Efficient convolutional networks for mobile vision applications. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 5251–5260. IEEE, 2019.

S. K. Soni, P. G. Howard, and E. Chansker. Leveraging grad-cam and mobilenet for cultural heritage conservation. *Journal of Deep Learning Applications*, 18(1):45–56, Jul 2023.

P. B. Tsiaras, D. G. Yu, and X. Zhou. Machine learning for cultural heritage preservation: The impact of deep learning. *Heritage Conservation Journal*, 6:82–91, May 2024.

T. T. Wei, L. J. Kim, and A. Y. Hu. Heritage preservation through deep learning: A case study with cnns. *Journal of Heritage Technology and Preservation*, 10:112–118, Aug 2021.

M. R. Williams, P. C. Martinez, and K. K. Sanderson. Deploying ai for cultural heritage conservation. *Journal of Modern Cultural Preservation*, 17:13–21, Jan 2023.

D. G. Xu, K. N. Ruo, and L. X. Zhen. Cultural heritage site defect detection using cnns and deep learning models. In *Proceedings of the AI Heritage Preservation Conference*, pages 142–156, 2022.