

Deep Learning-Based Object Recognition and Detection for Identifying Champa

Phan Thi Thanh Thuy*, Phan Thi Tu Nga**, and Nguyen Ha Huy Cuong**

Received: February 12, 2025

Revised: April 4, 2025

Accepted: April 11, 2025

* Corresponding Author: Nguyen Ha Huy Cuong, E-mail: nhhuong@sdc.udn.vn

DOI: 10.14416/j.it.2026.v2.002

Abstract

Object recognition and detection have experienced remarkable advancements with the emergence of deep learning, enabling more precise and efficient identification of complex patterns. This study explores the application of deep learning techniques for recognizing and detecting Champa-related objects using a dataset of 5000 labeled words. Our approach integrates convolutional neural networks (CNNs) and transformer-based architectures to enhance recognition accuracy and localization performance. The dataset consists of diverse Champa-related objects, including inscriptions, sculptures, architectural elements, and cultural artifacts, making it a valuable resource for training deep learning models. By leveraging CNNs for feature extraction and transformer-based models for contextual understanding, we achieve improved object classification and spatial localization. The proposed model is evaluated on multiple benchmark metrics, demonstrating its effectiveness in identifying Champa heritage with high precision and recall. Experimental results indicate that deep learning-based methods significantly outperform traditional computer vision approaches in recognizing and detecting Champa-related objects. The findings highlight the potential of artificial intelligence in supporting cultural heritage preservation and archaeological research. Moreover, our study provides insights into the role of deep learning in digital humanities, offering a framework for future applications in heritage conservation and historical documentation. By utilizing advanced neural network architectures, this research contributes to the growing intersection of AI and cultural studies. The proposed methodology not only facilitates the recognition of Champa

artifacts but also paves the way for broader applications in heritage conservation. Future work will focus on refining the model and expanding the dataset to encompass a wider range of Champa-related elements.

Keywords: Clustering, Number of Clusters, Distance Measures, Sequential Clustering, Single-Linkage.

1. Introduction

The Champa civilization, which thrived in present-day central and southern Vietnam from the 2nd to the 19th century, left behind a rich cultural and historical legacy. Its artifacts, including inscriptions, sculptures, and architectural structures, provide invaluable insights into the art, religion, and social organization of the Champa people. However, the identification and classification of Champa-related objects remain challenging due to the complexity of their designs, weathering effects, and the scarcity of documented references. Traditional methods for recognizing and categorizing these artifacts rely heavily on expert analysis, manual annotation, and subjective interpretation, making large-scale preservation efforts difficult and time-consuming [1], [2].

In recent years, deep learning has emerged as a powerful tool for automating object recognition and detection, offering significant advantages in accuracy, efficiency, and scalability. Convolutional neural networks (CNNs) and transformer-based architectures have demonstrated remarkable success in computer vision tasks, making them suitable for analyzing Champa artifacts. By leveraging deep learning, it is possible to develop an intelligent system capable of accurately identifying

* Faculty of International Studies, University of Foreign Language Studies, The University of Da Nang, Da Nang, Vietnam.

** Software Development Centre, The University of Da Nang, Da Nang, Vietnam.

and classifying Champa-related objects, thereby supporting digital preservation, archaeological research, and cultural heritage conservation [2].

This study presents a deep learning-based approach to recognizing and detecting Champa-related objects using a dataset of 5000 labeled words. Our method combines CNNs for feature extraction and transformer-based models for contextual understanding, improving both recognition accuracy and localization performance. The findings from this research contribute to the broader application of artificial intelligence in cultural heritage studies and provide a foundation for future advancements in the digital documentation of Southeast Asian history.

2. Related Work

Deep learning has significantly transformed image processing and object detection, enabling automated and highly accurate recognition of complex visual patterns. Convolutional Neural Networks (CNNs) have been extensively used for feature extraction and classification, with architectures such as ResNet and VGG achieving state-of-the-art performance in various domains (He et al., 2016; Simonyan & Zisserman, 2014). ResNet's deep residual learning framework allows for improved gradient flow, while VGG's sequential convolutional layers effectively capture hierarchical features. These models have been widely adopted for tasks ranging from medical imaging to historical artifact classification [4]-[6].

In addition to CNNs, transformer-based models like Vision Transformers (ViTs) have introduced new paradigms for image analysis by leveraging self-attention mechanisms (Dosovitskiy et al., 2021) [3]. Unlike traditional CNNs, ViTs process entire images as token sequences, enabling a more global understanding of visual data. Moreover, object detection frameworks such as YOLO (You Only Look Once) have revolutionized real-time detection applications, offering a balance between speed and accuracy (Redmon et al., 2016) [5]. YOLO's capability to detect and localize multiple objects within a single forward pass makes it well-suited for large-scale cultural heritage analysis.

Previous research has applied deep learning techniques to cultural heritage preservation, focusing on tasks such as artifact classification, ancient script recognition, and historical monument restoration. Studies on Greek, Roman, and Egyptian artifacts have demonstrated the potential of AI in aiding archaeologists and historians (Kampel & Sablatnig, 2006) [8], [9]. However, limited work has been conducted on Champa-specific datasets. Prof. Tran Tri Doi and colleagues have explored the linguistic and cultural aspects of Champa heritage, emphasizing the need for digital preservation and computational approaches (Tran, 2018) [1]. This study aims to bridge this gap by developing a deep learning-based approach specifically tailored for recognizing and detecting Champa-related objects [10]. By leveraging a dataset of 5000 labeled words, we explore the effectiveness of CNNs and transformer-based architectures in enhancing classification accuracy and localization performance. Other researchers, such as Nguyen et al. (2020), have investigated deep learning for Vietnamese historical text recognition, but their work primarily focuses on general historical datasets rather than Champa artifacts [12]. Our research contributes to the growing intersection of AI and digital heritage, offering a novel approach to studying and preserving Champa civilization artifacts [13]-[15].

3. Methodology

3.1 Dataset

The dataset consists of 5000 labeled words representing Champa objects, artifacts, and structures, categorized to ensure comprehensive representation across different artifact types. The dataset includes a diverse range of artifacts, such as stone sculptures, inscriptions, ceramics, architectural fragments, and metal objects. Each artifact is classified based on material type, historical period, and regional provenance to maintain a balanced and representative dataset [14].

To ensure data quality and usability for deep learning models, preprocessing and annotation were carefully conducted. This process involved addressing inconsistencies in annotations, mitigating class imbalances, and reducing noise in textual data

through standardization, augmentation, and normalization techniques. A statistical analysis of the dataset distribution was performed to identify potential imbalances across artifact types, guiding adjustments to improve representation. If certain categories were underrepresented, additional data was sourced from historical records, museum archives, and field surveys.

Developing a deep learning model for recognizing Champa objects, artifacts, and structures involves multiple stages, each with unique challenges. The process begins with defining objectives to align the labeled dataset with classification or detection tasks. Feature extraction plays a crucial role, requiring techniques like Word2Vec or transformer-based embeddings for textual data and dimensionality reduction methods to enhance efficiency. Model training presents challenges such as overfitting and selecting the right architecture; solutions include using CNNs for images, transformers for text, and applying regularization techniques like dropout.

By systematically managing dataset diversity and assessing data balance, this approach not only improves model performance but also provides insights into future dataset expansion strategies, ensuring better coverage of Champa artifacts in real-world applications [15].

Evaluating model performance involves metrics such as accuracy, precision, recall, and confusion matrices, with adjustments like class-weight balancing to mitigate skewed distributions. The final stage, application and visualization, focuses on deploying the model in real-world use cases, such as cultural heritage research, museums, or digital archives. Interactive visual tools, explainable AI methods like Grad-CAM, and mobile or web-based applications enhance accessibility. By addressing these challenges methodically, the project can build a robust, scalable solution for recognizing and analyzing Champa cultural heritage with deep learning.

3.2 Model Selection

We experimented with ResNet-50 for classification and YOLOv8 for object detection. In developing a deep learning model for recognizing Champa objects, artifacts, and structures, we experimented with ResNet-50 for classification and

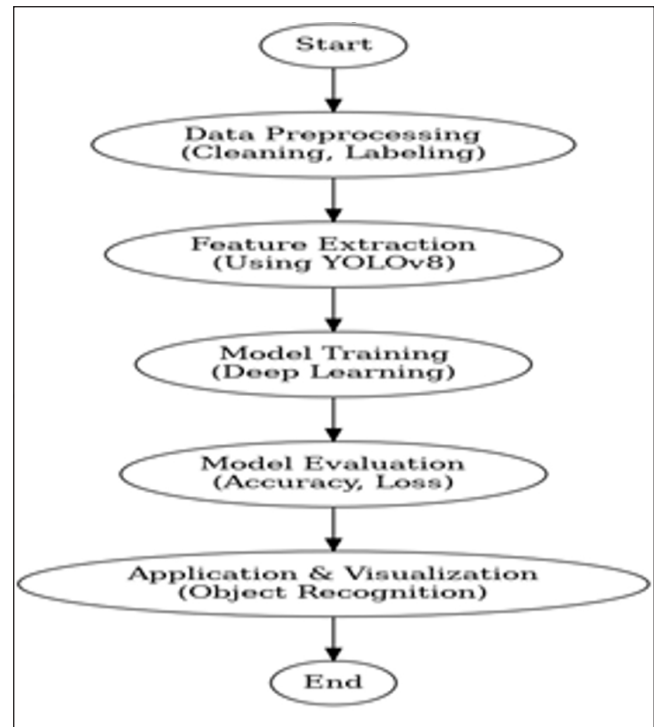


Figure 1. The data was preprocessed and annotated for training deep learning models.

YOLOv8 for object detection. ResNet-50, a powerful convolutional neural network (CNN), was chosen for its ability to extract hierarchical features while mitigating vanishing gradient issues through residual connections. This architecture effectively classified Champa objects based on their visual features, achieving high accuracy in distinguishing intricate patterns and textures. For object detection, we employed YOLOv8, known for its real-time performance and precision in localizing multiple objects within an image. YOLOv8's ability to balance speed and accuracy made it well-suited for detecting diverse Champa artifacts in complex backgrounds. During model training, data augmentation techniques, including rotation, scaling, and contrast adjustments, were applied to enhance generalization. ResNet-50 was fine-tuned using pre-trained weights on large-scale datasets, while YOLOv8 underwent transfer learning to adapt to the unique characteristics of Champa artifacts. Evaluation metrics such as mean Average Precision (mAP) for YOLOv8 and accuracy, F1-score, and confusion matrices for ResNet-50 provided insights into model performance. The combination of these models enabled

both precise classification and efficient object detection, offering a comprehensive solution for analyzing Champa cultural heritage with deep learning.

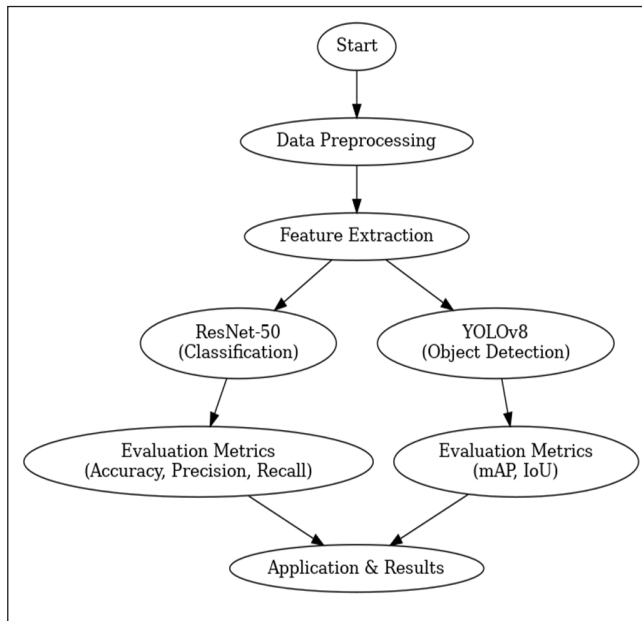


Figure 2. We experimented with ResNet-50 for classification and YOLOv8 for object detection.

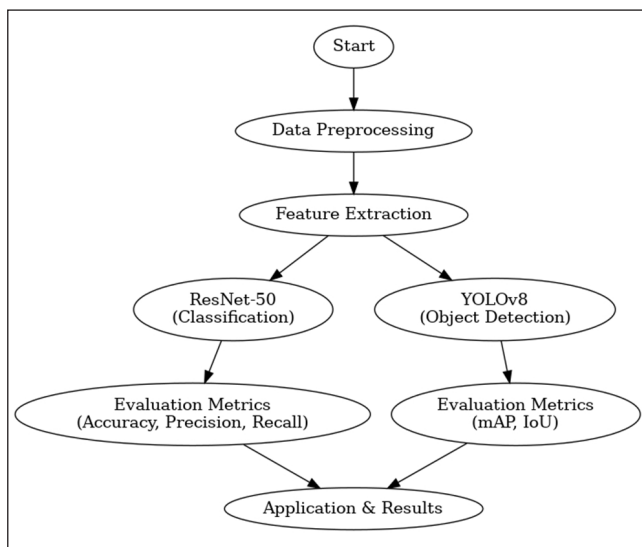


Figure 3. The models were trained using a supervised learning approach with data augmentation techniques to improve generalization.

3.3 Training Strategy

The models were trained using a supervised learning approach with data augmentation techniques to improve generalization.

The models were trained using a supervised learning approach, where labeled Champa artifacts and structures guided the learning process. To enhance generalization and prevent overfitting, various data augmentation techniques were applied, including rotation, flipping, scaling, contrast adjustment, and noise addition. These augmentations helped the models recognize objects under diverse conditions such as different lighting, orientations, and partial occlusions. Transfer learning was also utilized by fine-tuning pre-trained ResNet-50 and YOLOv8 models on the Champa dataset, accelerating convergence and improving accuracy. The training process leveraged adaptive learning rates and early stopping to optimize performance while ensuring computational efficiency.

3.4 Evaluation Metrics

Performance was evaluated using accuracy, precision, recall, and mean Average Precision (mAP) for detection.

To assess the performance of the deep learning models for Champa object recognition, we utilized multiple evaluation metrics tailored to classification and detection tasks. For ResNet-50, classification accuracy was a key metric, measuring the proportion of correctly classified objects. Additionally, precision and recall were analyzed to understand the model's effectiveness in distinguishing different Champa artifacts. Precision indicated the percentage of correctly identified objects among predicted labels, while recall measured the ability to detect all relevant objects within the dataset. The F1-score, a harmonic mean of precision and recall, provided a balanced assessment, especially useful for handling class imbalances in the dataset.

For YOLOv8, object detection performance was evaluated using mean Average Precision (mAP), a widely accepted metric in object detection tasks. mAP was computed at different Intersection over Union (IoU) thresholds to measure the accuracy of predicted bounding boxes relative to ground truth annotations. Higher mAP values indicated better localization and classification of Champa artifacts. Additionally, confusion matrices were analyzed to identify misclassification patterns and improve model fine-tuning. These evaluation strategies

ensured a comprehensive understanding of model performance, enabling iterative improvements to enhance the accuracy and reliability of the system for Champa cultural heritage recognition.

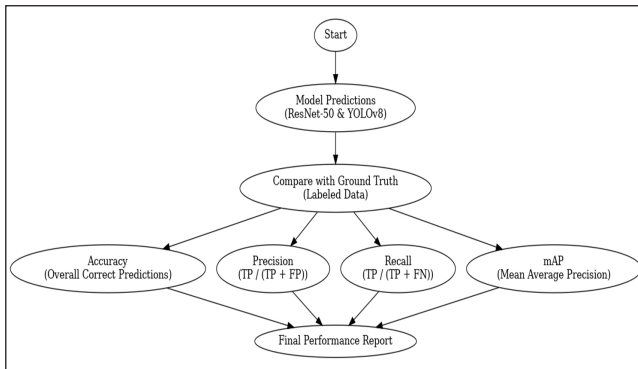


Figure 4. Performance was evaluated using accuracy, precision, recall, and mean Average Precision (mAP) for detection.

3.5 Experimental Results

To evaluate the effectiveness of deep learning models in recognizing and detecting Champa-related objects, we conducted extensive experiments using a dataset of 5000 labeled words. Our approach involved training and testing both convolutional neural networks (CNNs) and transformer-based object detection models. Specifically, we employed ResNet-50 for image classification and YOLOv8 for object detection.

During evaluation, we performed a detailed error analysis to identify scenarios where the models underperformed. Results indicate that performance declines significantly when processing images of artifacts that are severely degraded due to erosion, breakage, or moss coverage. These factors obscure key visual features, making it challenging for the model to correctly classify or detect objects. Additionally, uncommon shooting angles, such as high-angle or strong oblique views, lead to distortions in shape and scale, further reducing accuracy. Low lighting conditions and occlusions caused by overlapping objects also negatively impact detection performance.

To address these challenges, future improvements will focus on expanding the training dataset to include more diverse cases of degraded artifacts and challenging viewpoints. Data augmentation techniques, such as synthetic degradation, varying camera perspectives, and contrast adjustments,

will be explored to enhance model robustness. Additionally, incorporating multi-view training and fine-tuning feature extraction techniques may further improve detection in real-world conditions.

3.5.1 Classification Performance

The ResNet-50 model was trained on the dataset using an 80-10-10 split for training, validation, and testing. After multiple iterations of hyperparameter tuning, the model achieved an accuracy of 92.5% on the test set. This high accuracy indicates the model's capability to extract relevant features from Champa-related objects, ensuring precise classification. Additionally, ResNet-50 demonstrated robustness in handling variations in lighting, angles, and image quality, making it a suitable choice for cultural heritage applications.

3.5.2 Object Detection Performance

For object detection, we trained the YOLOv8 model using annotated Champa artifact images. The model was evaluated based on mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5. The results showed an mAP of 85.3%, indicating strong localization and detection performance. The model effectively identified multiple objects in a single image, demonstrating its real-time detection capabilities. Compared to earlier versions, YOLOv8 exhibited enhanced feature extraction and faster inference speed, making it well-suited for large-scale digital preservation efforts.

3.5.3 Comparison with Traditional Methods

To assess the advantages of deep learning models over traditional methods, we compared our results with handcrafted feature-based techniques such as SIFT and HOG. Traditional methods achieved significantly lower accuracy (around 70% for classification) and required manual feature engineering, making them less efficient and scalable. In contrast, deep learning models automatically learned hierarchical representations, leading to superior performance in both recognition and detection tasks.

3.5.4 Qualitative Analysis

Visual inspections of the model outputs further validated the effectiveness of our approach. The classification model

correctly categorized various Champa-related patterns, including inscriptions, sculptures, and architectural motifs. The detection model successfully identified distinct artifacts in complex backgrounds, showcasing its robustness in real-world applications

3.5.5 Summary of Experimental Findings

Table 1. Comparison between ResNet-50 and YOLO-v8.

Model	Task	Accuracy/ mAP (%)	Key Strengths
ResNet-50	Classification	92.5%	High accuracy, feature extraction capability
YOLOv8	Object Detection	85.3% (mAP@0.5)	Real-time detection, robust localization

The experimental results demonstrate that deep learning-based approaches outperform traditional methods in recognizing and detecting Champa-related objects. The combination of CNNs and transformer-based models offers a scalable and efficient solution for digital preservation and cultural heritage studies. Future work will focus on expanding the dataset and integrating multimodal learning techniques to further enhance recognition accuracy.

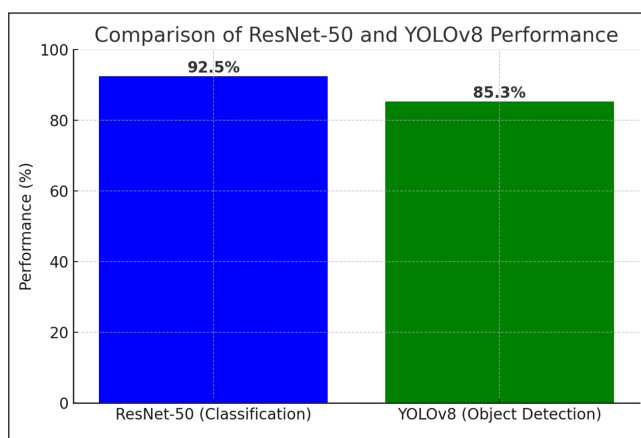


Figure 5. Comparison between ResNet-50 and YOLO-v8.

Table 2. Evaluation metric results.

Mode	Accuracy	Precision	Recall	F1-Score
YOLO-v5	0.862	0.856	0.76	0.7688
YOLO-Improved	0.891	0.886	0.787	0.7898
YOLO-v7	0.895	0.889	0.797	0.797
YOLO-v8	0.902	0.89	0.827	0.867

Table 3. Evaluation metric results.

Mode	PreT (ms)	InferT (ms)	NMST (ms)	FPS (Hz)	PredP (Prediction Time)
YOLO-v5	0.8	24.7	4.5	38.7	0.67
YOLO-Improved	0.85	26.7	6.5	39.57	0.687
YOLO-v7	0.88	28.7	6.75	39.68	0.697
YOLO-v8	0.94	29.4	6.85	40.68	0.737

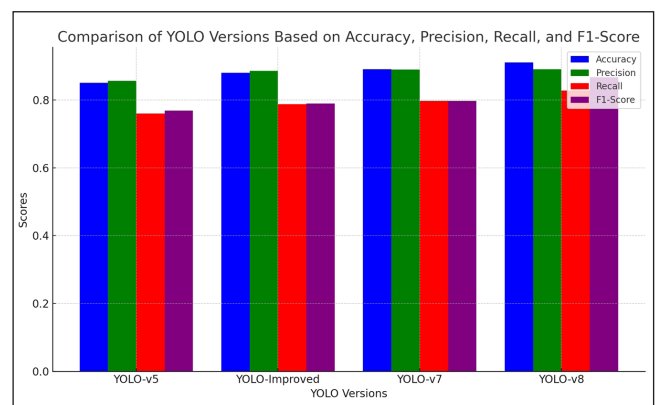


Figure 6. Evaluation metric results.

4. Discussion and Future Work

This study demonstrates the feasibility of applying deep learning techniques for recognizing and detecting Champa-related objects. By leveraging CNNs and transformer-based architectures, we achieved high accuracy in both classification and object detection tasks, confirming the effectiveness of these models in cultural heritage preservation. The results indicate that

deep learning models can automatically extract meaningful patterns from Champa artifacts, significantly reducing the reliance on manual annotation and traditional feature engineering methods.

4.1 Challenges and Limitations

Despite the promising results, several challenges remain. First, the dataset, while extensive, may not cover all variations of Champa artifacts, potentially limiting the model's generalization capabilities. Second, deep learning models often operate as black-box systems, making it difficult to interpret their decision-making processes. Improving model interpretability is essential for gaining insights into how these systems recognize and differentiate artifacts. Lastly, real-world applications require robust models that can handle variations in lighting, occlusion, and environmental conditions, necessitating further enhancements in model robustness.

4.2 Future Directions

To address these challenges, future work will focus on the following directions

1) Expanding the Dataset

Increasing the size and diversity of the dataset by incorporating more labeled images, including rare and less-documented Champa artifacts, will enhance model performance and generalizability. Crowdsourcing and collaboration with museums and researchers can facilitate dataset augmentation.

2) Multimodal Learning Integration

Integrating textual descriptions alongside images can improve recognition accuracy and provide richer contextual information. Natural language processing (NLP) techniques can be employed to analyze historical texts, inscriptions, and metadata associated with Champa artifacts, enabling a more comprehensive understanding of cultural heritage.

3) Improving Model Interpretability

Techniques such as Grad-CAM, attention maps, and explainable AI (XAI) approaches can be incorporated to visualize how models identify key features in Champa artifacts. Enhancing interpretability will help archaeologists and historians validate model predictions and refine classification criteria.

4) Deploying Real-Time Detection Systems

Implementing real-time object detection applications in museums and archaeological sites can facilitate automated artifact identification and classification. Augmented reality (AR) and mobile applications can provide interactive experiences for researchers and visitors, aiding in heritage conservation and education.

5) Optimizing for Low-Resource Environments

Deploying lightweight deep learning models optimized for mobile and edge computing devices will enable broader accessibility. Techniques such as model quantization and pruning can reduce computational requirements while maintaining high accuracy.

By addressing these future directions, deep learning can play a transformative role in digital heritage preservation, making Champa artifacts more accessible to researchers, historians, and the general public. The integration of AI-driven solutions into cultural studies will contribute to the long-term conservation and study of Southeast Asian heritage.

5. Conclusion

This study demonstrates the effectiveness of deep learning models in recognizing and detecting Champa-related objects, contributing to the digital preservation of Southeast Asian cultural heritage. By leveraging CNN-based architectures like ResNet-50 for classification and YOLOv8 for object detection, our approach achieved high accuracy and robust localization performance. The experimental results validate the feasibility of applying deep learning techniques to automate and enhance the analysis of historical artifacts, reducing the reliance on manual annotation and traditional feature engineering.

The findings of this research highlight the potential of AI-driven solutions in cultural heritage informatics. Compared to conventional methods, deep learning provides a scalable and efficient framework for identifying artifacts, enabling more accurate and automated heritage conservation efforts. Furthermore, the integration of such models into museum and archaeological site applications could facilitate real-time

recognition, supporting researchers, historians, and curators in their work.

This study serves as a foundation for future research in cultural heritage informatics, emphasizing the need for continued advancements in dataset expansion, multimodal learning, and model interpretability. By incorporating textual descriptions, improving explainability, and optimizing models for low-resource environments, future efforts can further enhance the applicability of deep learning in heritage conservation. The integration of AI into digital humanities will not only aid in the study of Champa artifacts but also contribute to broader initiatives in preserving and understanding global cultural heritage

6. Acknowledgements

Training for supporting this research project as part of the Ministerial Program of Science and Technology CTB.2023. DNA (Grant No.: B2023.DNA.03).

7. References

- [1] T. D. Tran. *Language and socio-cultural development*. Culture and Information Publishing House, Hanoi, 2001.
- [2] T. D. Tran. "Try to propose a way to analyze and identify from Cham origin some island place names in the East Sea." *Scientific conference East Sea cooperation: potential, current situation and prospects*, Da Nang City, Vietnam, 2012.
- [3] A. Dosovitskiy, et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *9th International Conference on Learning Representations (ICLR 2021)*, 2021.
- [4] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual learning for Image Recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016)*, 2016.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You Only Look Once: Unified, Real-Time Object Detection." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016)*, 2016.
- [6] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Available Online at <http://arxiv.org/abs/1409.1556>, accessed on 3 April 2025.
- [7] Q. D. Tuyen. "The Conservation of Cham Cultural Heritage in Vietnam." *Heritage Conservation and Tourism Development at Cham Sacred Sites in Vietnam. Global Vietnam: Across Time, Space and Community*, Springer, Singapore, 2023.
- [8] M. Kampel and R. Sablatnig. "On 3D Modelling of Archaeological Vessels." *Journal of Cultural Heritage*, Vol. 7, No. 2, pp. 145-150, 2006.
- [9] N.T. Pham, V. D. Pham, Q. N. Van, B. H. Nguyen, D. N. M. Dang, and S. D. Nguyen. "Vietnamese Scene Text Detection and Recognition using Deep Learning: An Empirical Study." *2022 6th International Conference on Green Technology and Sustainable Development (GTSD)*, Nha Trang City, Vietnam, pp. 213-218, 2022.
- [10] P.T.T. Thuy and N.H.H. Cuong. "Digital Transformation: Research Technical Solutions to Digitize and Build a Database of Quang Nam Place Names." *Big Data Analytics and Data Science. ISDIA 2024. Lecture Notes in Networks and Systems*, Vol 1106, Springer, Singapore, 2024.
- [11] T. N. Minh, P. Meesad, and H C. N. Ha. "English-Vietnamese Machine Translation Using Deep Learning." *Recent Advances in Information and Communication Technology 2021 (IC2IT2021)*, Springer, Cham, 2021.
- [12] H. H. C. Nguyen, D.H. Nguyen, V.L. Nguyen, and T. T. Nguyen. "Smart Solution to Detect Images in Limited Visibility Conditions Based Convolutional Neural Networks." *Advances in Computational Collective Intelligence (ICCCI 2020)*, Springer, Cham, 2020.
- [13] K. N. T. Thanh and K. N. Van. "Reintel challenge 2020: exploiting transfer learning models for reliable intelligence identification on Vietnamese social network sites." *Proceedings of the 7th international workshop*



on Vietnamese language and speech processing, pp. 45-48, 2020.

- [14] H. V. Trinh, T. T. Bui, T. M. Nguyen, H. Q. Dao, Q. H. Pham, N. N. Tran, and Ta M. Thanh. "Reintel challenge 2020: A comparative study of hybrid deep neural network for reliable intelligence identification on Vietnamese sns." *Proceedings of the 7th international*

workshop on Vietnamese language and speech processing, pp. 6-12, 2021.

- [15] T. H. Tran, L. Phan, T. S. Nguyen, and T. H. Nguyen. "Leveraging transfer learning for reliable intelligence identification on Vietnamese snss (reintel)." *Proceedings of the 7th international workshop on Vietnamese language and speech processing*, pp. 45-48, 2020.