Integrative AI for the Understanding of Ancient Javanese Architectures

Arnadi Murtiyoso [1], Gabriele Mazzacca [2], Fabio Remondino [2], and Deni Suwardhi [3]

[1] Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, 6700 Strasbourg, France - arnadi.murtiyoso@insa-strasbourg.fr
[2] 3D Optical Metrology Unit (3DOM), Fondazione Bruno Kessler (FBK), 38121 Trento, Italy – (gmazzacca, remondino)@fbk.eu
[3] Spatial System and Cadastre Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Bandung Indonesia - deni.suwardhi@itb.ac.id

Keywords: Temples, AI, Javanese, Machine Learning, Visual Query, Gaussian Splatting

Abstract

The use of digital techniques has seen an increasing amount of use in recent years for heritage documentation. The development of artificial intelligence (AI) also contributed to this rise, with many different applications to help facilitate the heritage recording process. A by-product of these developments is the increasing amount of available data, in tandem with the ever-increasing need for training data for AI purposes. This paper aims to re-use old datasets and repurpose them using modern methods. The objective is therefore to see if older datasets may be used to improve the quality of AI-based methods, while also investigating the use of new technologies such as Visual Language Models (VLM) to perform semantic queries and Gaussian splatting on these datasets. For this purpose, datasets from a previous documentation project involving Javanese “*candi*” architecture is used in this paper since this particular subject has not seen too many AI-based documentation research in the literature and is thus an interesting example to evaluate the generalisation of AI methods. Results show that old datasets can very well be used with modern techniques with promising results. In terms of semantic segmentation, machine learning yielded an overall accuracy of 0.89 while deep learning yielded 0.79. Several interesting inferences were also observed in the VLM query results, while Gaussian splatting showed very strong potential for visualisation-based applications to further enhance the reusability of these old datasets.

1. Introduction

1.1. Rationale

The documentation of cultural heritage is an essential process in heritage conservation. In recent decades, the merit of 3D technology for this purpose has been acknowledged. Indeed, 3D reconstruction technologies enable a precise and faithful representation of tangible heritage which may be stored for archival purposes. These geometric representations of heritage, whether in the form of point clouds or meshes, may also be imbued with semantic attributes, which opened the way for multiple analysis to be conducted.

In recent years, the use of artificial intelligence (AI) has also seen a surge in multiple applications related to heritage documentation (De Luca, 2020). Both image-based and point-cloud based semantic segmentation employs machine learning to infer object classes and thus facilitate the data processing step, paving the way to integrated information systems in the form of Heritage Building Information Models (HBIM) and digital twins (Sammartano et al., 2023). AI has also seen other applications, including 3D rendering using radiance fields where the 3D information is stored as Multiple Layer Perceptrons (MLP) as is the case in Neural Radiance Fields (NeRF) or 3D Gaussians (Mazzacca et al., 2023; Murtiyoso and Grussenmeyer, 2023). Learning-based algorithms have also shown great promise in supporting multi-view stereo (MVS) methods to create better dense point clouds from image-based data.

1.2. Aim of the work

The aim of this work is to re-use old datasets and apply modern processing methods to improve the geometric quality of the derived documentation materials as well as add semantic information useful for better understanding heritage contents. To this end, integrative AI methods are used notably to understand Javanese temples (Murtiyoso et al., 2019) surveyed some 10 years ago. The choice of Javanese temples is multi-fold: (ii) re-use existing datasets for boosting the use of archival data, (ii) increase documentation literature on this architectural domain and (iii) evaluate the generalisation of AI methods.

2. Re-use of 3D surveying heritage data

Heritage archives are gaining more and more importance nowadays, in particular with the advent of AI methods able to process single or sparse sets of images and derive interesting 3D data useful for documentation, valorisation and communication purposes. This work considers datasets representing cultural heritage from the island of Java, Indonesia. In particular, several datasets of archaeological remains in Central Java dated to the 8th century will be used, mostly built from volcanic rocks and adorned with rich bas reliefs depicting Hindu and Buddhist stories. 3D reconstruction techniques were used to document these sites for several years, with the dual purpose of archival and 3D documentation of the sites (Murtiyoso et al., 2019b; Suwardhi et al., 2015).

The first dataset used comes from the Sewu Temple complex in Central Java, Indonesia. The Sewu complex is dated to the 8th century and is regarded as one of the biggest Buddhist religious complexes in Southeast Asia. The site was built in a symmetrical 100x100 meters square mandala pattern, with one main temple and a few hundred smaller shrines. While most of the smaller shrines are ruined, several including the large main temple are still preserved or have been reconstructed.

In 2019, a CIPA tropical school was organised in the Sewu temple complex, gathering international participants (Murtiyoso et al., 2021). During this one-week hands-on training, the area of the temple complex was scanned using terrestrial laser scanner (TLS), lidar UAV (Unmanned Aerial Vehicle) and aerial photographs, both nadiral and oblique. A topographic network was also established to help with georeferencing. Furthermore, terrestrial laser scans and close-range images were taken for five smaller but representative temples, of which 3 are shrines ("*Perwara*") and 2 are medium-sized guardian ("*Apit*") temples.

A secondary dataset was taken from a 2018 survey of another ancient Javanese temple, namely the Sari temple (Murtiyoso et al., 2019b). Unlike Sewu, the Sari dataset consists of only one single structure. However, it dates to roughly the same era as Sewu in terms of construction and indeed is located a few dozen kilometres from the Sewu site, both located in the archeologically dense Prambanan plains along the Opak river valley near the city of Yogyakarta (Degroot, 2009). Sari was also identified as a Buddhist site. Instead of a traditional worship place, it was however identified as a possible monastery, thus providing an interesting contrast in terms of building use while still adhering to the same architectural style as Sewu. The Sari dataset available for this study was in the form of 3D point clouds acquired by TLS and photogrammetry. Similarly to the Sewu setup, a geodetic network was also established around Sari to ensure proper geometric registration and georeferencing.

A third dataset available to this study involves point clouds and orthophotos of panels located on the walls of Borobudur. Built also during the same era as Sewu and Sari, Borobudur is the largest Buddhist temple in Indonesia and a listed UNESCO World Heritage site. The Borobudur dataset stems from a survey performed in 2015 (Suwardhi et al., 2015) and presents thus an interesting view into the state of the temple one decade ago. The Borobudur dataset, due to its nature, will mostly be used for tests involving VLM queries.

3. Processing methodology

Starting from the available images and 3D point clouds, the methodology wants to apply 3D semantic segmentation methods and visual queries to enrich the available datasets and evaluate generalization aspects of current AI tools. For 3D classification, Random Forest 4 Point Cloud Classification (RF4PCC) (Grilli and Remondino, 2020) and PointTransformer (Wu et al., 2024) tools were applied. This 3D semantic segmentation test was performed for the Sewu and Sari datasets, with basic semantic segmentation analysis performed, i.e. computation of statistics such as F1 score, precision, and recall.

For the semantic segmentation experiment, two of the temples within Sewu dataset, which consists of point clouds from five temples in total, were manually classified into 11 classes. These classes were derived from the *vāstu* order as elaborated by Perdana and Kurniawan (2022). The classes as represented in the "Group 2" point cloud of the Sewu dataset is shown in Figure 1.

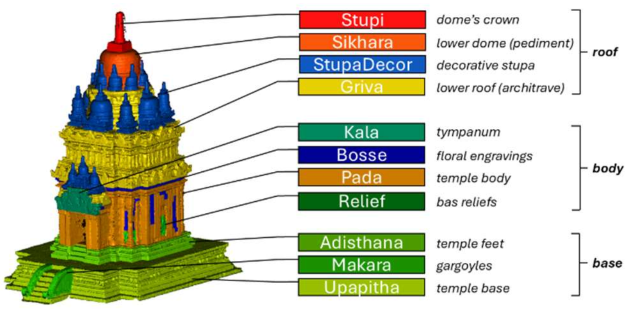

Figure 1. Classification of the temple architecture according to the *vāstu* order elaborated by Perdana and Kurniawan (2022), here applied to one of the temples in the Sewu dataset.

The Borobudur data were also processed with Visual Language Models (VLM), i.e. an AI model designed to understand and generate information that combines both visual (a given image) and textual (a given query) data. A VLM (Hu et al., 2023; Liu et al., 2023) is a system that processes and interprets images or visual inputs (such as photos, videos, or drawings) alongside language-based information (like text). These models are trained to understand the relationships between the two modalities, enabling them to perform tasks that require both visual and linguistic understanding. Beside VLM, Multi-Modal Large Language Models are also gaining importance (Yuan et al., 2025) and they were evaluated on the panels of the temples.

The VLM test was deployed on the Borobudur orthophotos, using the following prompts:

1. *Can you find and count the faces in the image?*
2. *Can you identify the elephant in the image?*
3. *Can you segment all the figures carved in stone?*
4. *Can you guess the narrative presented by the image?*
5. *Can you classify the different sitting positions of the "Buddha" class?*
6. *Can you distinguish the faces between "Buddha" and "Person"?*

The two last prompts are interesting options because in the majority of its panels, Borobudur tells stories from well-known Buddhist narratives (*Jātaka*, *Avadana*, etc.). Prompt #4 may therefore be able to help users to quickly understand the context of the panel and eventually link them to an information system to easily localize a panel among the thousands of bas-relief in Borobudur. Prompt #5 may also help users for the same purpose, as the depictions of Buddha in the temple usually follows a preset sitting positions (*mudrā*), usually linked to which façade and which floor the depiction is located in (Hidayat et al., 2014).

A third and final analysis was to employ 3D Gaussian splatting (3DGS) (Kerbl et al., 2023) to the Sari dataset. 3DGS creates novel viewpoint from training images by representing the scene as explicit anisotropic Gaussians, which are "splatted" to create a 3D render. This means that this technique can create photorealistic 3D rendering in a fairly quickly and is thus an interesting method as far as data visualization is concerned.

4. Results

4.1. Semantic segmentation

Semantic segmentation tests were done on photogrammetric point clouds of five temples in the Sewu dataset. Of the five temples, two were annotated and the remaining three kept as prediction datasets. Half of each training dataset was segmented manually to be used as training, while the other half were used as validation. Eleven different classes, following the classical *vāstu* order and decorations (Perdana and Kurniawan, 2022), were considered.

The prediction results (Figure 2) gave a promising quantitative result for both of the tested models. The machine learning model RF4PCC attained an overall accuracy of 0.89 and average F1 score 0.82 (see Table 1), although visually the algorithm still struggled with some specific classes. On the other hand, the deep learning method PointTransformer generated an overall accuracy of 0.79 and an average F1 score of 0.73, slightly lower than RF4CC. Similarly, visually this method also seems to struggle with some classes, especially underrepresented ones such as "Bosse" and "Kala". It should also be noted that the very small training information and lack of generalization may also influence these results.

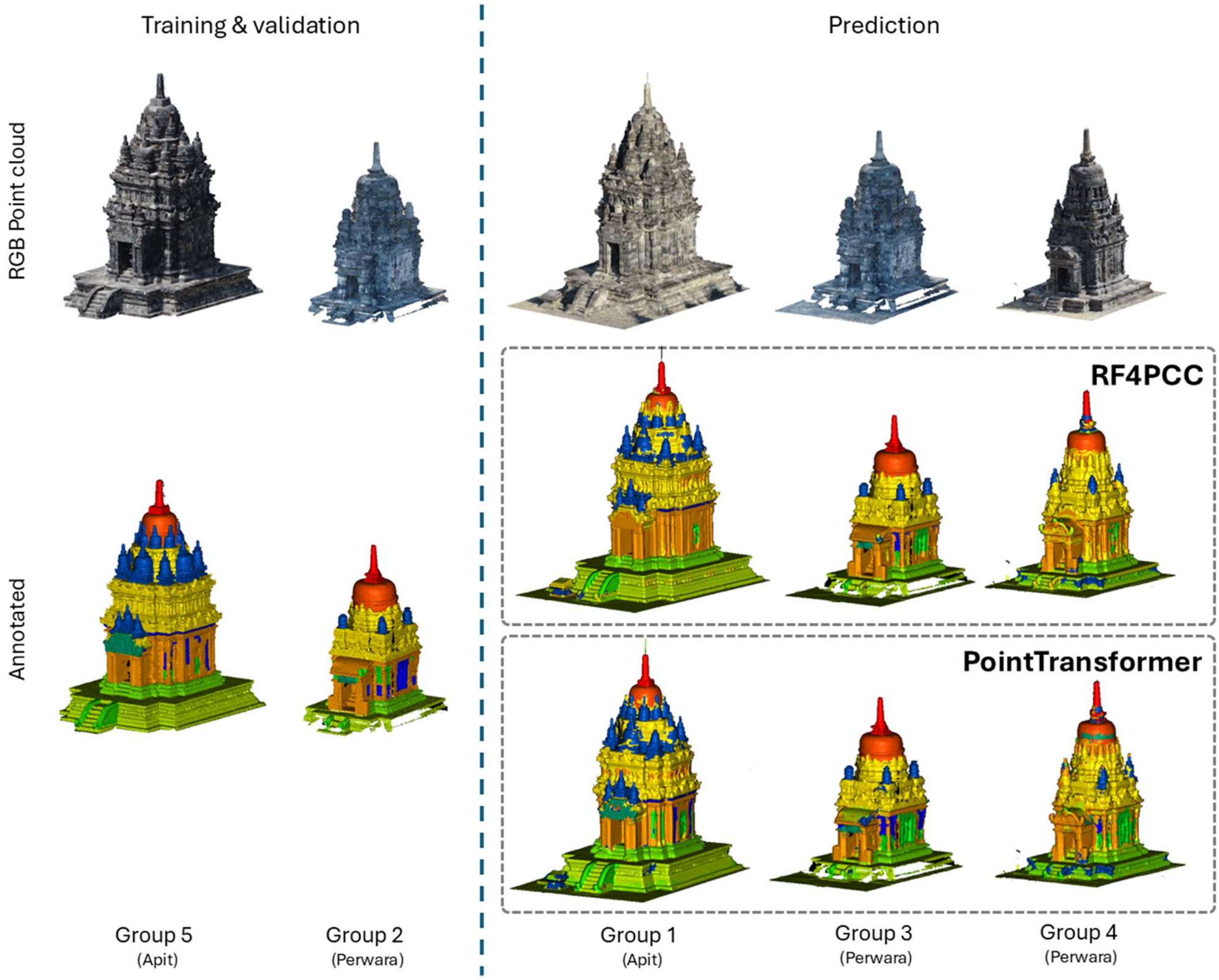

Figure 2. Results on the Sewu datasets, semantically segmented using the RF4PCC and PointTransformer algorithm.

	Class	RF4PCC			PointTransformer		
		P	R	F1	P	R	F1
1	Bosse	0.65	0.54	0.59	0.33	0.75	0.46
2	StupaDecor	0.90	0.84	0.87	0.70	0.86	0.77
3	Kala	0.61	0.46	0.52	0.43	0.79	0.55
4	Relief	0.77	0.64	0.70	0.49	0.86	0.63
5	Makara	0.93	0.69	0.79	0.56	0.76	0.65
6	Adisthana	0.95	0.95	0.95	0.83	0.93	0.88
7	Upapitha	0.97	0.97	0.97	0.95	0.91	0.93
8	Griva	0.84	0.95	0.89	0.83	0.74	0.79
9	Pada	0.87	0.83	0.85	0.80	0.56	0.66
10	Sikhara	0.94	0.84	0.89	0.82	0.85	0.84
11	Stupi	0.97	0.97	0.97	0.87	0.93	0.90
	Average	**0.86**	**0.79**	**0.82**	**0.69**	**0.81**	**0.73**
	Weighted average	**0.89**	**0.89**	**0.89**	**0.81**	**0.79**	**0.79**

Table 1. Qualitative assessment parameters for the semantic segmentation process.

Figure 3 shows a comparison of the F1 scores for each class from the two tested methods. RF4PCC performs better across all classes except "Kala", where it performed slightly worse than PointTransformer. On the other hand, PointTransformer fared notably worse for several specific classes such as "Bosse" (-0.13 points with regards to RF4PCC) and "Pada" (-0.19 points). In both cases, results for the "Bosse", "Kala", and "Relief" classes performed the worst, although this may be explained by the under representation of these classes' training data.

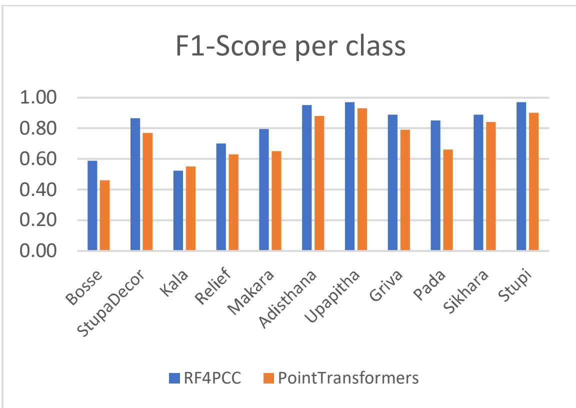

Figure 3. Bar chart showing the performance of the two tested methods on the eleven predicted classes in terms of F1 score.

(a) *(b)* *(c)*

Figure 4. First tests on performing visual query using a VLM (Haotian et al., 2023). The prompts were: (a) Can you find and count the faces in the image? (b) Can you identify the elephant in the image? (c) Can you segment all the figures carved in stone?

	
Q: You're given the image of a stone relief coming from an archaeological site in Southeast Asia. Could you interpret the scene represented in it and describe the details of the decorations?	**Q:** This relief comes from a temple in Southeast Asia. could you make a description of it and interpret the scene it represents?
A: *The stone relief appears to depict a significant scene from Southeast Asian history or mythology. The central figure, likely a deity or important historical character**,** **is seated in a cross-legged position**[a], which is common in Southeast Asian art and often associated with meditation or important moments of revelation.* *The figure's posture and the objects around them suggest a ceremonial or spiritual scene. The item in their left hand could be a symbolic object, perhaps a staff or scepter, while the flower in their right hand might represent purity or offering.* *The architectural elements surrounding the central figure are particularly interesting. The pillars and structures on either side could represent **a palace, temple, or other important building**[b], possibly indicating the figure's status or the significance of the location.* *The intricate carvings on the pillars and the decorative elements in the corners showcase the advanced artistic and craftsmanship skills of the civilization that created this relief. These details might include geometric patterns, stylized representations of nature, or symbolic motifs that were important in their culture.* *The use of **stone as the medium for this relief is typical of many Southeast Asian archaeological sites**[c], as stone was a durable material suitable for creating lasting monuments and telling important stories.* *Overall, this relief seems to be a complex scene that combines religious, royal, and cultural elements, providing valuable insights into the beliefs, social structures, and artistic traditions of the society that produced it.*	***A:*** *The stone low relief you're referring to is **characteristic of Southeast Asian Buddhist art, particularly from the classical period**[d]**.** It depicts a serene scene of monastic life, with several monks in various meditation postures.* *The relief is carved into a horizontal band, likely made of sandstone or a similar light-colored stone. The figures are rendered with great detail, showing the monks' robes, facial features, and peaceful expressions. They're arranged in different positions - some seated cross-legged in lotus positions, others standing or in walking postures.* *The composition is carefully balanced, with the central figure of a **high-ranking monk or Buddha**[e] surrounded by other monks. This arrangement emphasizes the importance of spiritual leadership and community in Buddhist practice.* *The style of the relief is naturalistic yet stylized, typical of Southeast Asian Buddhist art. The artists have skillfully captured the essence of meditation and spiritual enlightenment through their use of subtle lines and contours, as well as the expressive quality of the figures' faces and postures.* *This type of relief was common in temples and monasteries throughout the region, serving both decorative and educational purposes. It taught visitors about Buddhist teachings and monastic life while also being a form of artistic expression that reflected the cultural and religious values of the time.*

Table 2. Results of the query asking to interpret the input images. Several interesting statements generated by the reply are designated by (a), (b), (c), (d) and (e).

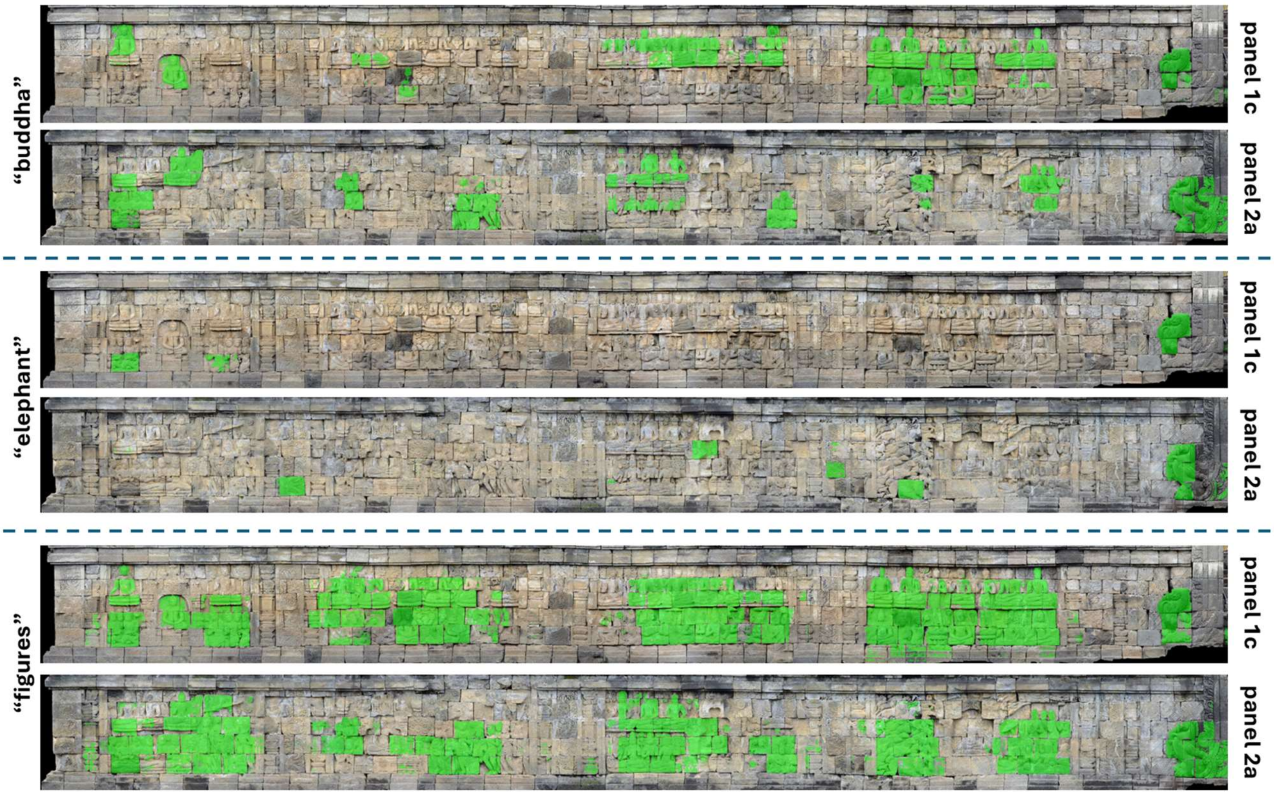

Figure 5. Segmented orthoimages of Borobudur panels using the Sa2VA model (Yuan et al., 2025).

Figure 6. Results of the image segmentation projected into the 3D point cloud of the panels for the class "buddha".

However, when the trained model was used to predict classes on the other two datasets (Sari and Borobudur), both models failed to generate meaningful results. This is despite the fact that, at least in the case of Sari, both datasets were built using the same architectural style, dated more or less to the same era, and are located within the same geographical region. A more likely cause to this problem is the scale of the objects which are very different between the datasets. This problem is even more evident with the Borobudur dataset, where the site is significantly larger and higher resolution data consists more of relief panels than architectural elements.

4.2. Visual query

Figure 4 shows the results of the tests with a VLM on Borobudur panels. This panel in particular is located on the north wall of the temple on the first gallery. It is numbered 76 in a set of panels depicting the story of the monk Rudrayana (part of an episode in the *Avadana* corpus). The query results can then be projected from the images to the point clouds to perform quantitative analyses or used for unsupervised 3D classification processes. In this figure, the first three queries related to object detection and image segmentation was tested. Figure 4 (a) showed violet dots on the results of the query, where the AI model was able to count the number of faces on the panel except for one. This one failure to identify a person is however due to the deteriorated state of the panel in this particular part, rendering the identification to fail. That being said, the human attributes apart from the face is still intact, meaning that the AI model was able to limit its search to a very specific query.

In Figure 4(b), the AI was asked to identify an animal, notably an elephant and segment the image containing this instance. It correctly identified the elephant in the panel, however the segmentation part of the image still showed that a part of another class (a person) was included in the segmentation. Similarly in 4(c), the AI was asked to segment the figures in the image. While it succeeded in doing its attributed task, one figure to the far left of the image was unidentified, probably due to insufficient features.

Table 2 shows the result of a different type of prompts, specifically those asking for a qualitative interpretation of the scene. Overall, the AI gave a very general description of the scene and was not able to pinpoint a specific interpretation without more context. Some of the most interesting observations are denoted in the table:

(a) *"... is seated in a cross-legged position...":* here the AI managed to identify a specific sitting style that is typical of Buddhist carvings. However, it shies away from giving details of the type of *mudrā* as identified in (Hidayat et al., 2014). This means that this result cannot yet be used to localise the panel with regards to Borobudur's sections.

(b) *"... a palace, temple, or other important building..."*: the AI managed to identify the carvings of a structure to the right side of the image, usually indeed interpreted as a temple structure.

(c) *"... stone as the medium for this relief is typical of many Southeast Asian archaeological sites..."*: the AI provided an additional context correctly, although again it shies away on giving detailed information on where exactly this style of carving originated from.

(d) *"... characteristic of Southeast Asian Buddhist art, particularly from the classical period."*: the AI correctly predicts that the panel depicts a Buddhist (in contrast to other similar styles such as Hindu) scene even without context from the prompt, as well as the architectural era (classical).

(e) *"... high-ranking monk or Buddha ..."*: the AI correctly identified an important figure in the middle of the scene, although it failed to recognise that the central figure does not, in fact, depict the Buddha.

More results obtained using the Sa2VA model for VLM are presented in Figure 5, focusing on the detection of figures (both animal and anthropomorphic), Buddhas, and elephants. Here Prompt #6 were used to obtain the results. The analysis was conducted on both orthophotos and 3D point clouds, with panels 2a and 1c of the Borobudur dataset selected as representative examples. Figure 6 shows the projected labels from the segmented orthophoto into the 3D point cloud for the class "buddha".

Despite targeted efforts to identify various iconographic representations of the Buddha, the model exhibited limited success, even with simpler object categories. This underperformance is attributed to several challenging conditions inherent in the dataset: the weathered state of the stone surfaces, the presence of chipped areas, and the visual disruption caused by seams between stone blocks. These factors collectively hindered the model's ability to accurately interpret and segment the visual content. They represent significant challenges to automated visual analysis in heritage contexts. Here the projection of labels towards the 3D data is a way to propose a method that could lean towards a zero-shot approach (no specific 3d training required) for 3D segmentation.

4.3. 3D Gaussian splatting

The Sari dataset was used to perform a 3DGS reconstruction in order to see the potential of this novel technology. Figure 7 shows the results of this process, generated in about 15 minutes using an NVIDIA RTX 3000 Ada Generation Laptop GPU computer and using DSLR images in the dataset. In this figure, the rendering produces a photorealistic texture as can also be seen on the zoomed-in inset. The "splatting" of the Gaussians can however be observed by the artefacts on the 3D render. While 3DGS was not originally conceived to create point clouds, it is possible to convert the predicted Gaussians into a geometric point cloud by applying density thresholds. This was done in this case.

The Sari dataset was the subject of several analysis detailed in Murtiyoso et al. (2019a) in which various photogrammetric sensors were tested against a reference TLS point cloud. To test the geometric potential of the 3DGS method, we decided to reuse this analysis but add the 3DGS point cloud as an additional discussion point (Figure 8); in this regard it is possible to determine whether 3DGS is an interesting option for cultural heritage 3D reconstruction as far as the geometry is concerned.

To this end, an identical setup was taken but this time using the 3DGS point cloud as the input. The cloud-to-mesh analysis to compute signed Euclidean distance of the main facade was performed as a TLS mesh was available. In Murtiyoso et al. (2019a), the same analysis was performed to compare point clouds generated by DSLR, 360, and drone cameras.

Figure 7. 3D rendering of the Sari dataset by using the 3DGS method, with an inset on a specific carving to show the textural quality of the rendering.

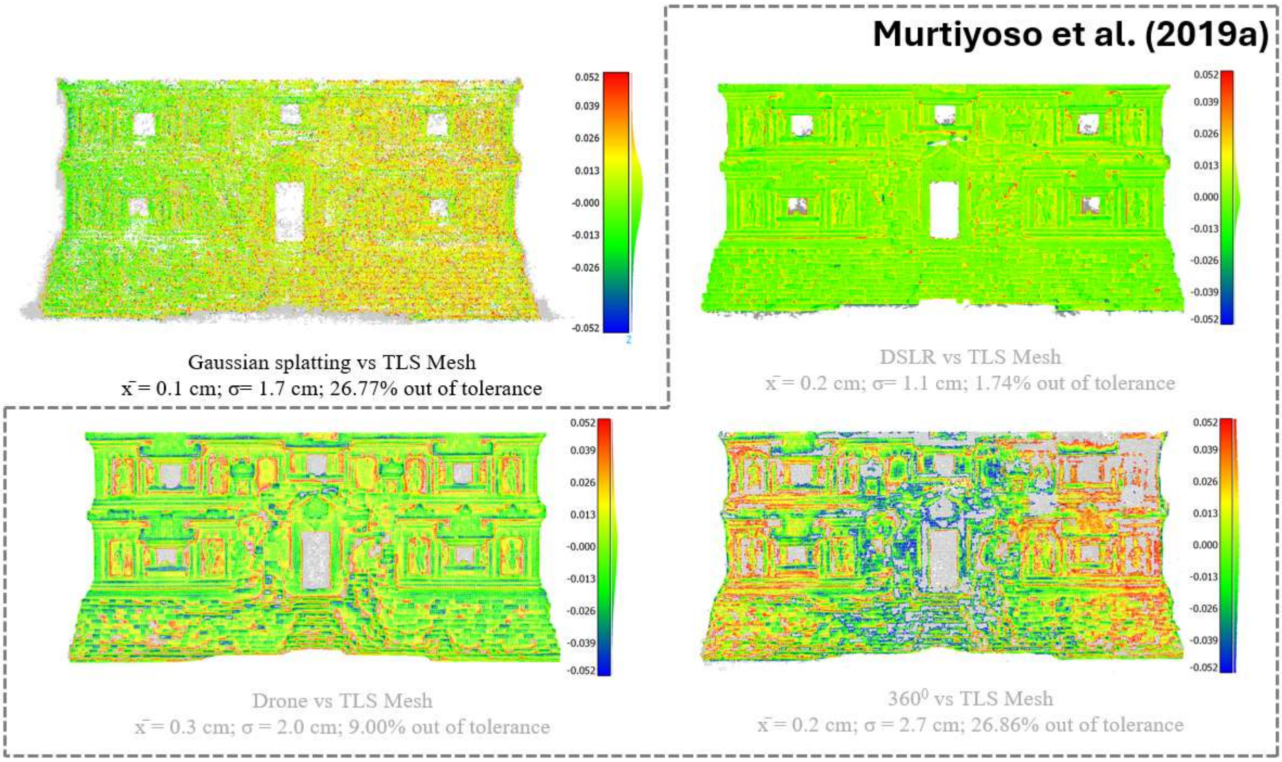

Figure 8. A cloud-to-mesh analysis of the Sari’s main facade 3DGS point cloud, compared to a TLS-derived mesh. This analysis is here showcased against an older one performed under the same parameters described in Murtiyoso et al. (2019a).

Note that in the Sari dataset, the DSLR data was acquired using a Canon EOS 5DS R with a 24 mm fixed lens. The drone data was captured using a DJI Phantom 4, while a Ricoh Theta V was used for the 360^0 (spherical) acquisition. The TLS data itself was taken using a Faro Focus M70.

In Figure 8, we see that when the same analysis was applied, the 3DGS point cloud is able to generate a mean error of 0.1 cm, which is very much comparable to the other sensors tested. The standard deviation amounted to 1.7 cm; this indeed reflects the high amount of noise still present. However, this also shows that the error is generally more distributed than the other photogrammetric results, namely those with a similar or worse standard deviation values (drone with 2.0 cm and 360 with 2.7 cm). It is also interesting to note that 3DGS generated the highest rate of outlier points, amounting to 26.77% of the points. This is much higher than DSLR at 1.74% and drone at 9%, but rather similar to the 360^0 camera dataset at 26.86%. This observation is actually visually observable; indeed, one of the bottlenecks in the use of 3DGS for 3D reconstruction is the conversion process from Gaussians into point cloud. This problem has already been observed before with other radiance fields-based methods (Murtiyoso and Grussenmeyer, 2023), although contrary to NeRF, 3DGS uses a more explicit representation and parametrisation in the future should be more straightforward.

5. Conclusions and perspectives

In this paper, three types of AI implementation were tested using older datasets to help increase the understanding of Javanese architecture. Semantic segmentation was tested using machine learning (RF4PCC) and deep learning (PointTransformer) on the Sewu dataset. Both methods show promise, although generalisation still remains a problem since the trained model was not able to be transferred to other temples despite proximity in style, historical era, and geographical location (but different scales).

A VLM-based query was performed on the Borobudur dataset, mainly concerning panels of bas reliefs. Our results show that while the VLM showed promise for object identification-related tasks, it remains frustratingly generic when used for interpretation-related tasks.

Finally, the Sari dataset, already subject to a geometric analysis in a previous project, was reprocessed using the 3DGS method. The resulting point cloud is still nowhere near an acceptable level for heritage archiving purposes, with notable problems related to data noise and the number of outliers. However, it generated a very good photorealistic visuals in a relatively quick processing time (especially compared to conventional MVS-based dense point cloud process). With future links to game engines, this method may have an excellent potential for immersive-based products such as augmented or virtual reality, or gamification.

Overall, this paper addressed the question of how AI can be applied on older datasets and still generate interesting results. With the ever-increasing rate of development in the AI domain, the results presented in this paper, if anything, showed a trend towards better results both in terms of geometric and non-geometric quality. Older datasets play an important role in this case since not only can they be used as a snapshot of a particular epoch for the heritage site, but they may also be used to complement modern AI-based techniques, such as has been demonstrated in a small scale with this paper.

Acknowledgements

This paper was published within the context of the Digital Twins for Cultural Heritage (Twin4CH) Junior Professorship Chair, financed by the French National Agency for Research (*Agence nationale de la recherche* – ANR).

References

De Luca, L., 2020. Towards the Semantic-aware 3D Digitisation of Architectural Heritage, in: Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia HeritAge Contents. ACM, New York, NY, USA, pp. 3–4.

Degroot, V., 2009. Candi, Space and Landscape: A Study on the Distribution, Orientation and Spatial Organization of Central Javanese Temple Remains, Mededeling. ed. Sidestone Press, Leiden.

Grilli, E., Remondino, F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. ISPRS Int J Geoinf 9, 379.

Hidayat, I.K., Sunarto, P., Guntur, T., 2014. Mengenal Relief, Mudra dan Stupa Candi Borobudur untuk Anak-Anak Usia 9-12 Tahun melalui Edugame. Journal of Visual Art and Design 6, 58–68.

Hu, J., Yao, Y., Wang, C., Wang, S., Pan, Y., Chen, Q., Yu, T., Wu, H., Zhao, Y., Zhang, H., others, 2023. Large multilingual models pivot zero-shot multimodal learning across languages. arXiv preprint arXiv:2308.12038.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Trans Graph 42.

Liu, H., Li, C., Wu, Q., Lee, Y.J., 2023. Visual Instruction Tuning, in: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 34892–34916.

Mazzacca, G., Karami, A., Rigon, S., Farella, E.M., Trybala, P., Remondino, F., 2023. NeRF for heritage 3D reconstruction, in: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 25–30.

Murtiyoso, A., Grussenmeyer, P., 2023. Initial assessment on the use of state-of-the-art NeRF neural network 3D reconstruction for heritage documentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-2–2023, 1113–1118.

Murtiyoso, A., Grussenmeyer, P., Suwardhi, D., 2019a. Technical Considerations in Low-Cost Heritage Documentation. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W17, 225–232.

Murtiyoso, A., Grussenmeyer, P., Suwardhi, D., Fadilah, W.A., Permana, H.A., Wicaksono, D., 2019b. Multi-Sensor 3D Recording Pipeline for the Documentation of Javanese Temples, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 829–834.

Murtiyoso, A., Suwardhi, D., Grussenmeyer, P., Fadilah, W.A., Fauzan, K.N., Trisyanti, S.W., Macher, H., 2021. Heritage documentation and knowledge transfer: a report on the CIPA Tropical School in Candi Sewu (Indonesia). The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-M-1–2021, 493–497.

Perdana, A.B., Kurniawan, K.R., 2022. Vāstu Order as an Alternative Concept for Analysing Javanese Temple Architecture | Tatanan Vāstu sebagai Konsep Alternatif untuk Menelaah Arsitektur Candi Jawa. SPAFA Journal 6, q790t38g6i.

Sammartano, G., Avena, M., Fillia, E., Spanò, A., 2023. Integrated HBIM-GIS Models for Multi-Scale Seismic Vulnerability Assessment of Historical Buildings. Remote Sens (Basel) 15.

Suwardhi, D., Menna, F., Remondino, F., Hanke, K., Akmalia, R., 2015. Digital 3D Borobudur - Integration of 3D Surveying and Modeling Techniques, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 417–423.

Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H., 2024. Point Transformer V3: Simpler, Faster, Stronger, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 4840–4851.

Yuan, H., Li, X., Zhang, T., Huang, Z., Xu, S., Ji, S., Tong, Y., Qi, L., Feng, J., Yang, M.-H., 2025. Sa2VA: Marrying SAM2 with LLaVA for Dense Grounded Understanding of Images and Videos. arXiv preprint arXiv:2501.04001.