Buddhist Face Segmentation with 3D Point Clouds

Yuehan Pan[1], Junqi Luo [2,3], Gefei Kong [2,4], Hongchao Fan [2]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, China – p.yuehan@gmail.com
[2] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway – (gefei.kong, hongchao.fan)@ntnu.no
[3] College of Surveying and Geo-Informatics, Tongji University, Shanghai, China – harden@tongji.edu.cn
[4] Heidelberg Institute for Geoinformation Technology, Heidelberg, Germany

Keywords: Point Clouds, Semantic Segmentation, Regularization, Buddhist statues, SAM

Abstract

Three-dimensional (3D) semantic segmentation of Buddhist heads is essential for digital heritage applications such as virtual restoration, conservation, and art-historical analysis. However, existing segmentation methods face challenges due to the complex geometry and degraded surfaces of heritage statues. This paper proposes an efficient method for face segmentation from 3D point clouds of Buddhist heads, leveraging geometric features and topological relationships. The approach comprises three stages: (1) symmetry-based rotation for orientation normalization, (2) vertical gridding and color mapping for depth-aware parameterization, and (3) hybrid segmentation using a Face Topology Graph (FTG) and the Segment Anything Model (SAM) with point and box prompts. Experiments on 50 Buddhist heads demonstrate the method's efficiency and robustness, achieving an average IoU of 0.73 across seven facial components. The proposed workflow provides a scalable solution for semantic 3D modeling of heritage artifacts, supporting accurate analysis and interactive visualization.

1. Introduction

Three-dimensional (3D) modeling of Buddhist heads with high-fidelity geometric detail and rich semantic annotation plays a crucial role in diverse applications, including archaeological and art-historical analysis, virtual restoration, conservation planning, and digital museum display. Accurate digital representations not only safeguard cultural heritage but also provide a basis for advanced computational analysis and interactive visualization.

In archaeological research, experts often rely on precise measurements and comparative analysis of specific components—such as eyes, nose, mouth, cheeks, ears, and ornamental elements—to investigate stylistic variation, production techniques, and regional identities of Buddhist statues (Li, 2018). For instance, subtle differences in gaze orientation or lip curvature can reveal significant insights into iconographic evolution and cultural transmission across dynasties and geographical areas (Zhang, 2022). Capturing these nuanced characteristics requires 3D models that accurately represent the geometric, topological, and morphological details of all components, down to the smallest ornamental feature (Pan et al., 2025). Beyond archaeological interpretation, such semantic-rich models facilitate iconographic comparison and stylistic categorization, thereby enhancing our understanding of historical and artistic contexts (Morgan, 2022).

For the purpose of modelling Buddhist heads in 3D with semantic information, we need to segment 3D point clouds of Buddhist head into small parts whereas each part is corresponded to a semantic component on Buddhist head. Semantic segmentation for 3D point clouds has been a major research focus in the recent years. Existing approaches—ranging from geometric feature-based algorithms to machine learning and deep learning methods—offer varying trade-offs between accuracy, generalizability, and computational cost.

Geometry-based segmentation methods primarily rely on local geometric properties, such as curvature, surface normals, and continuity, to partition point clouds into meaningful regions. Techniques like region growing and surface fitting have been widely applied in domains such as mapping and industrial measurement, where objects exhibit relatively simple geometric structures (Poux et al., 2022; Li and Shan, 2022). However, these methods encounter significant challenges when applied to Buddhist heads, which are characterized by intricate facial features, ornamental details, and degraded surfaces caused by centuries of erosion. The variability in geometry—combined with the presence of fractures, missing components, and uneven scanning density—limits the effectiveness of rule-based approaches that assume homogenous properties.

To overcome these limitations, recent research has focused on learning-based segmentation approaches, which can be broadly divided into structured-data methods and point-based methods. Structured-data approaches convert unstructured point clouds into regular representations, such as multi-view images or voxel grids, allowing the use of convolutional neural networks (CNNs) for feature extraction (Jhaldiyal and Chaudhary, 2023; Gezawa et al., 2022). While these methods leverage mature image-based models, they suffer from projection distortions and information loss, making them less suitable for complex heritage objects. Point-based methods, by contrast, operate directly on raw point clouds, preserving geometric fidelity. Pioneering models such as PointNet and PointNet++ introduced permutation-invariant architectures and hierarchical feature learning (Qian et al., 2022), while later works incorporated graph-based learning and attention mechanisms for better local-context modeling (He et al., 2024). Despite their superior performance, deep learning models demand large annotated datasets and computational resources—conditions rarely met in cultural heritage contexts, where data is highly heterogeneous and annotation costs are prohibitive.

To address these challenges, this study proposes a novel, lightweight method for the face segmentation of Buddhist heads. Instead of applying complex existing methods (e.g., deep learning-based methods), the proposed method catches the geometric features of Buddhist head, fostering efficient and accurate segmentation. It fully utilizes the spatial characteristics of Buddhist head, namely, the nose is the highest tip on Buddhist face and other components such as mouth, forehead, left eye, right eye, left cheek, and right cheek have fixed topological relations to the nose. Therefore, the nose will be

segmented in the first step and then a face topological graph (FTG) will be established with the nose as centre of the FTG. In the next step, other facial components are segmented by using Semantic Anything Model (SAM)-based method with the prompt of FTG.

All the aforementioned steps require the face to be oriented in a strict frontal perspective, whereby, the two eyes are at the same level of height and have the same distances to the very front plan or the backward projection plan. It depicts that the proposed method is efficient to align the point clouds of Buddhist head to the correct position. The proposed approach is implemented and tested using 3D point clouds of 50 Buddhist heads. Quantitative evaluation is performed on 10 Buddhist heads with manually labelled point clouds, while qualitative evaluation is conducted through visual inspection.

2. Methodology

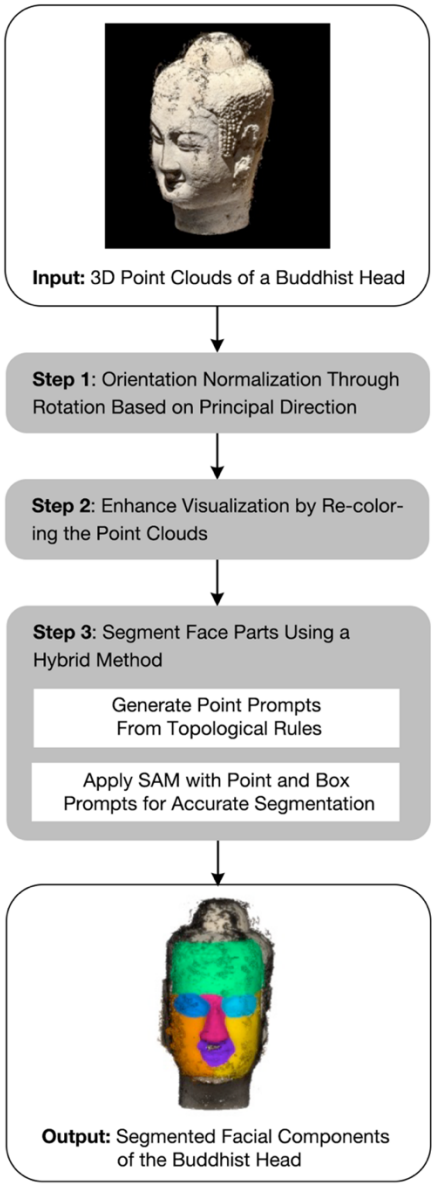

Figure 1. Overall Workflow of the Proposed Buddhist Face Segmentation Method

The entire workflow of the proposed method is illustrated in Figure 1. It mainly consists of three modules: (1) regularization of Buddhist head by main direction extraction based on symmetry analysis and then rotating 3D point clouds so that the main direction is aligning with the y-axis, (2) generation of depth map by point cloud gridding and color mapping for the purpose of detecting seed points of facial components, and (3) face segmentation based on topological rules and large vison model (LVM) by inputting the initial segmentation results (seed points) in the second step.

2.1 Symmetry-based point cloud rotation

In module (1), the goal is to determine the principal orientation of the Buddhist statue's point cloud and rotate it to align with y-axis, facilitating accurate Buddhist face segmentation. The process begins by slicing the input 3D point clouds of the Buddhist head along z-axis and extracting the points around the center z-values. This step reduces the data volume and speeds up the subsequent main direction estimation. Next, the sliced point cloud is projected onto xy-plane along z-axis, and its 2D boundary is extracted (see the black outline on Figure 2a and 2b). The main direction (the red line on Figure 2a.) of the input Buddha's head point cloud is then estimated based on the bilateral symmetry, which is a common characteristic of human-like faces and statues. The main direction estimation is achieved as follows.

1. Define a set of candidate angles of main direction $\Theta = \{\theta \mid \theta \in [0,180], ang \in Z\}$, and initialize the angle θ = 0 (unit: degree).
2. For each candidate θ, construct a vector along the θ direction and use it to split the traced 2D boundary into two parts on the left and right sides of the vector.
3. Reflect the left-side boundary points onto the right side, and then compute the mean distance between the reflected points and their closest points on the right points. Denote this symmetry error (from left to right) as e_{lr}.
4. Similarly, reflect the right-side boundary points to the left, and compute the corresponding symmetry error (from right to left) e_{rl}.
5. Calculate the total symmetry error $e = e_{lr} + e_{rl}$.
6. Repeat step 2—5 across all candidate angles in Θ. The θ that yields the minimum total symmetry error e is selected as the main direction of the Buddha point cloud.

Ultimately, a 2D rotation matrix is constructed using the calculated main direction. The original point cloud is then rotated and aligned with the y axis (Figure 2b.), standardizing the orientation for the following process.

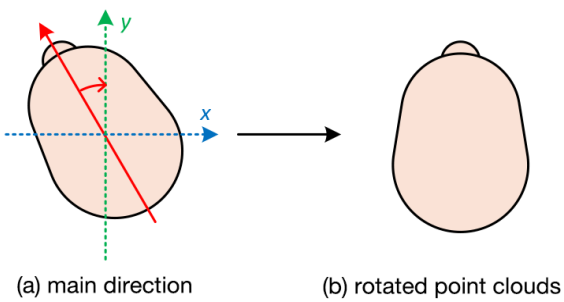

Figure 2. Rotation Based on Main Direction

2.2 Color mapping of Buddhist head via gridding

With the point cloud properly aligned with coordinate system from Module (1), the facial features are easier to extract. In module (2), first, the rotated point cloud is projected to x-y plane to extract the 2D bounding box. Afterwards, grid the bounding box along the y-axis based on a predefined resolution and obtain a series of horizontal grid bins (Figure 3a.). A finer resolution corresponds to a more detailed re-colored parameterization result. Subsequently, determine the corresponding y-axis grid bin for each point, and assign a unique color (i.e., label) to each bin to colorize the 3D points. Finally, the re-colored point cloud outputs as the head parameterization result of the Buddhist statues, including the nose, cheeks, and forehead, as shown in Figure 3b.

This vertical gridding can be, actually, regarded as a depth map with the tip of nose is the highest point while mouth and eyes are lower than the nose. This kind of approach is particularly effective for statues with front-facing, symmetric heads—like Buddhist statues—allowing for a computationally efficient and structurally meaningful head parameterization output.

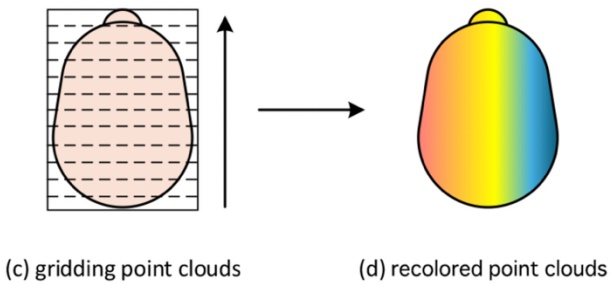

Figure 3. Re-coloring Buddhist Point Clouds

2.3 Face segmentation via topological rules and LVM

After further cleaning the recolored Buddhist head, the facial-only Buddhist point cloud will be used for the segmentation step. The module for face segmentation consists of two sub-modules: (i) nose segmentation and the establishment of face topological graph (FTG) based on recolored Buddhist heads and (ii) Semantic Anything Model (SAM)-based face segmentation with the prompt of FTG, where SAM is one of the most famous and widely-used LVM.

2.3.1 Nose extraction and FTG establishment

The colored Buddhist face point cloud significantly highlights tip of the nose as the point with the color at the end of the color bar which is corresponding to the highest points on the depth map. By defining it as the seed point, nose area is easily extracted as it has a clear dividing line with other facial parts.

After nose extraction, a Face Topology Graph (FTG) can be established based on facial topological features. First, as illustrated in Figure 4a, we project Buddhist face point cloud to 2D along y-axis to generate the 2D Buddhist face image *I*. Subsequently, starting from the nose node on *I* (i.e., the 2D point for tip of the nose), forehead node is set close to the top of the point cloud and much higher than the nose node, while mouth node is set at the bottom and much lower than the nose. Two eye nodes (left and right) are located between the forehead and nose nodes, and closer to the forehead. On the contrary, two cheek nodes are at the left and right of the nose nodes, or a littler lower and between the mouth and nose nodes.

Ultimately, the seven nodes of FTG are obtained: nose, mouth, forehead, left eye, right eye, left cheek, and right cheek. All other nodes connect to the centred nose node by edges, completing the establishment of FTG, as denoted in Figure 4b.

2.3.2 SAM-based face segmentation

SAM is a powerful foundation model for image segmentation. We use it here to segment face parts of Buddhist face image *I*. SAM can receive two types of prompts: (1). sparse point, box, and text prompts; and (2). dense mask prompt. FTG already provides the initial point prompts to SAM, guiding SAM to segment face parts. However, these point prompts are generally not enough: although brow bone clearly divides forehead and eyes, there is no clear and complete dividing lines between cheeks and eyes. Cheeks and mouth as well. Additionally, forehead and cheeks generally show a smooth distribution and lack features, from color to texture.

According to this analysis, additional box prompt is necessary, as it presents the location information to each face parts, which is significant and fixed for a face. Hence, the box prompts of forehead, two eyes, and the mouth are generated based on FTG and the size of *I*. Blue boxes in Figure 4c presents an example of the generated box prompts. Two cheek box prompts are passed as they can be easily segmented after the segmentation of all other face parts.

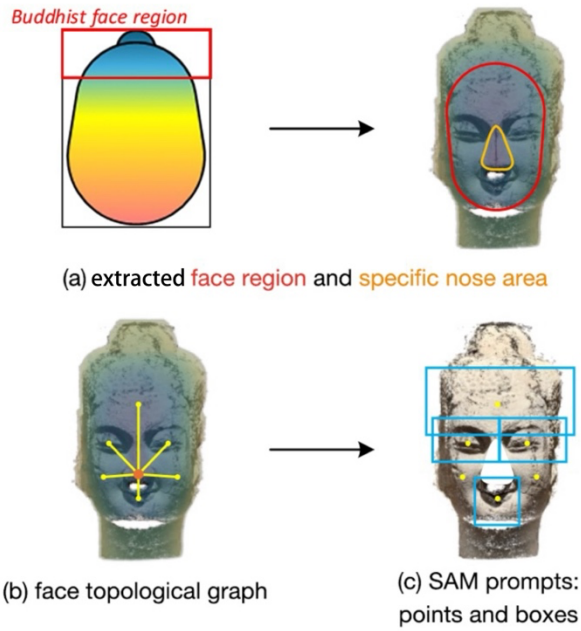

Figure 4. Face Segmentation Based on Topological Rules and SAM

3. Experiments and Results

The proposed method has been implemented and tested with 3D point clouds of 50 Buddhist heads. The 3D point clouds of each Buddhist head were generated by using Pix4D with about 100 to 150 images taken by a smart phone. The images were captured in culture heritage sites and museum in Dazu, Urumqi and Beijing in 2024 and 2025. Among others, 10 Buddhist heads have been semantically labelled manually in CloudCompare for the purpose of quantitative evaluation. In this section, results in the experiment will be demonstrated and evaluated.

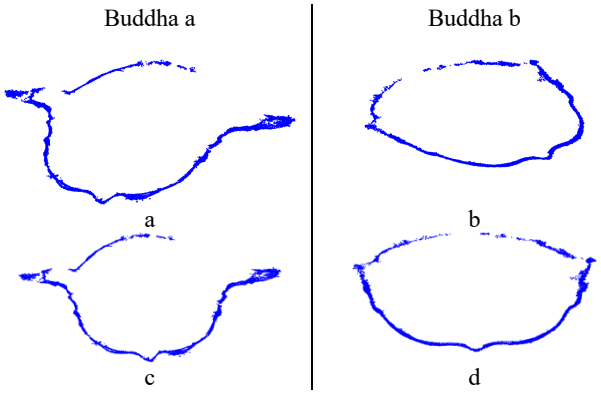

Figure 5. Result of main direction-based rotation

Figure 5 shows the results of projecting 3D point clouds of two Buddhist heads onto x-y plane and rotating their main axis aligning with y axis. It is the output from module (1) by taking point clouds of two Buddhist heads as examples, while Figure 5(a) and Figure 5(b) represent Buddhist head before rotation and Figure 5(c) and Figure 5(d) represent the results after the rotation. As one can see, the input point clouds have been successfully aligned with y-axis based on the estimated main directions. The alignment normalizes facial orientation across different statues, providing a consistent basis for parameterization.

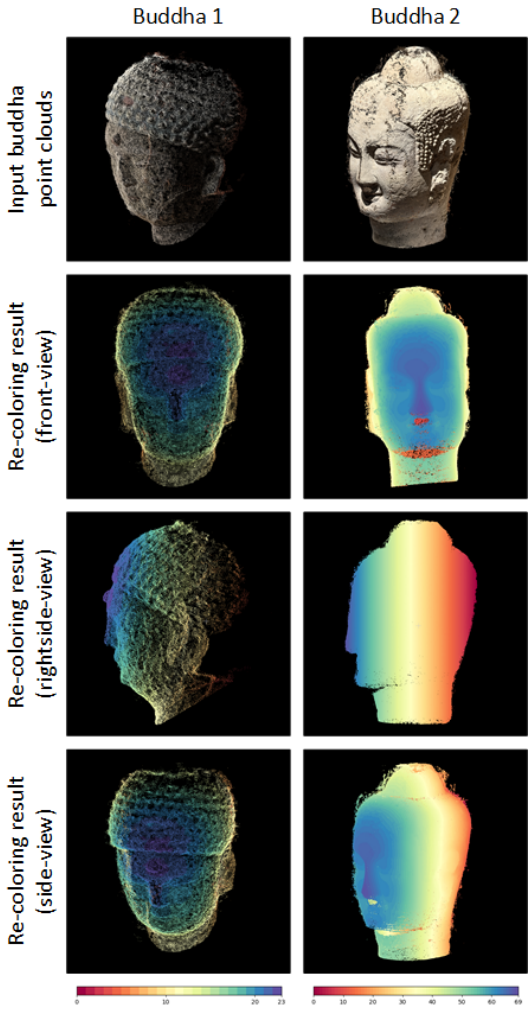

Figure 6. Result of head re-coloring

For the next step in module 2, with the input of point clouds after the regularization, Figure 6 shows that the re-colorized point clouds. According to the algorithm of re-colouring presented in Section 2.2, the output is actually a 2.5D model with regularized grid and height information in each grid. When we observe the Buddhist head from the side, it is substantially obvious that the tip of nose is the highest place, and the forehead is the second highest place. In the next step, this spatial characteristic will be used to identify the grid of nose tip at first and then build up the FTG for further process.

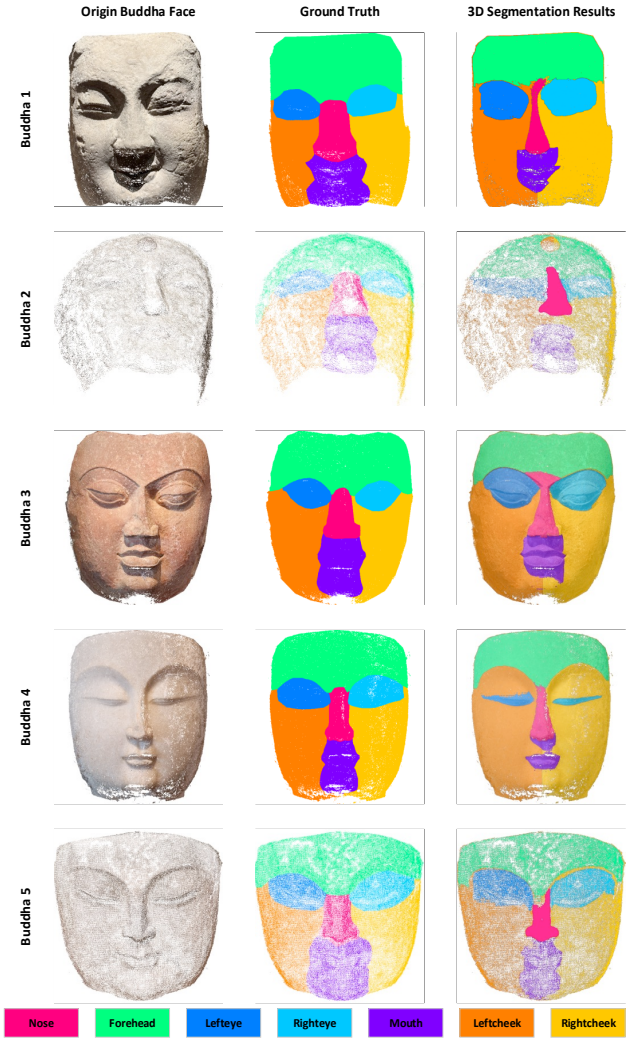

Figure 7. Result of 3D segmentation

For all the 50 Buddhist heads, we did qualitative evaluation by visually comparing the segmentation results with the photos of these 50 Buddhist heads. Figure 7 shows the final results of face segmentation results for five selected Buddhist head from module (3). In the first column, the input data of 3D point clouds of the five Buddhist heads are listed. The middle column shows the ground truth data acquired by manually labelling. And the third column shows the results yield from the proposed approach of face segmentation.

Overall, the precision of the face segmentation is very high, as all seven components have been correctly targeted and segmented from the 3D point clouds. However, the accuracy is not that high. When comparing the segmentation results in the third column with the ground truth data in the second column, it can be observed the difference: (i) the boundaries of some components were not exactly identified and there is wrong

Table 1. Statistics of different regions segmentation from the Buddha face

	$mIoU$	IoU_{Nose}	IoU_{Forehead}	IoU_{Lefteye}	IoU_{Righteye}	IoU_{Mouth}	$IoU_{\text{Leftcheek}}$	$IoU_{\text{Rightcheek}}$
Buddha 1	0.712	0.626	0.727	0.583	0.744	0.683	0.783	0.843
Buddha 2	0.755	0.541	0.928	0.811	0.866	0.505	0.803	0.832
Buddha 3	0.727	0.725	0.717	0.710	0.676	0.609	0.830	0.823
Buddha 4	0.559	0.646	0.737	0.403	0.373	0.527	0.619	0.606
Buddha 5	0.734	0.799	0.803	0.747	0.707	0.621	0.743	0.716
Buddha 6	0.758	0.898	0.851	0.771	0.799	0.395	0.777	0.816
Buddha 7	0.870	0.852	0.799	0.979	0.872	0.770	0.854	0.961
Buddha 8	0.775	0.893	0.874	0.486	0.385	0.937	0.935	0.915
Buddha 9	0.678	0.583	0.728	0.591	0.674	0.779	0.708	0.681
Buddha 10	0.733	0.741	0.751	0.616	0.704	0.634	0.879	0.804
Average	0.730	0.730	0.792	0.670	0.680	0.646	0.793	0.800

classification always in the place of the end of bridge of the nose; (ii) similarly, there is also wrong classification in the lower part of mouth. In further, there are usually less segmentation in the eye areas, (iii) otherwise, forehead and cheeks are well segmented.

For quantitative assessment, we compared the segmentation results of 10 Buddhist heads with the ground truth data that were labelled manually using the software of CloudCompare. Whereas we use the metric of Intersection of Union (IoU) for the evaluation. The calculation of IoU can be applied with the Equation 1.

$$IoU = \frac{\left|Mask_{\text{seg}} \cap Mask_{\text{gt}}\right|}{\left|Mask_{\text{seg}} \cup Mask_{\text{gt}}\right|} \tag{1}$$

Table 1 presents the IoU scores for different regions of the Buddha face, including the IoUs of 10 individual test samples as well as their average. While the highest value of IoU is 0.979 for the left eye segmentation of Buddha number 7, the lowest value of IoU is 0.373 for the right eye segmentation of Buddha number 4. And the average accuracy of the segmentation is about 0.73 for all 10 Buddhist heads. However, the segmentation for cheeks and forehead is significantly better than the results of other components. The reason is that the surface of these three components is large and smoothly curved, where the other four components have more complicated geometric details.

We also inspected the performance of the proposed method. The experiment is conducted on a laptop with Intel Core i7-10750H CPU @ 2.60 GHz 5 GHz with a 32 GB RAM and a Nvidia Quadro T1000 GPU (8 GB). The average time costs of three individual stages are shown in Table 2. The sizes of the 3D point clouds of the 50 Buddhist heads are between 80MB to 120MB with 2.5 million to 3 million points. The total process takes less than one minute. The average time for the three modules are 19.67 seconds, 0.05 seconds and 39.64 seconds, respectively. It depicts that the proposed method is substantially efficient.

Table 2. Time costs of every stage in the proposed method

	Point Cloud Rotation	Color Mapping	Face Segmentation
Time Cost (s)	19.67	0.05	39.64

4. Conclusions and Outlook

This paper presents an AI-driven, automatic and training-free method for face segmentation from 3D point clouds of Buddhist statues. The proposed method contains three main steps, (1) main direction alignment, (2) vertical gridding, and (3) semantic segmentation using facial topological rules and LVM. By leveraging the inherent symmetry of head structures, the proposed method accurately aligns and highlights key facial features such as the nose, forehead, and cheeks. Furthermore, benefiting from the proposal of FTG based on the significant topological relationship of face parts and the introduction of novel LVM model, SAM, the proposed method finally achieves automatic and reasonably accurate face segmentation. The results of face segmentation from 3D point clouds can be used for further step of 3D modelling with semantic information that can support quantitative calculations and interactive operations in various applications.

Experiments have been carried out with 3D point clouds of 50 Buddhist heads whereas the 3D point clouds were generated from dense image matching by inputting 100-150 images of each Buddhist heads. Qualitative evaluation was conducted by visual inspection based on checking the segmentation results against photos of Buddhist heads. Quantitative evaluation was done by calculating the factor of Intersection of Union (IoU) to compare the segmentation results with the ground truth data which is obtained by manual labelling in CloudCompare. Both types of evaluation depict that the overall results of segmentation are satisfactory because there is no missing and wrong segmentation.

Nevertheless, over- and under-segmentation along component boundaries exist for all components on almost every Buddhist head. The reason lies in the ambiguity of the process when defining the exact boundaries among facial components. The results of segmentation on left, right cheeks and forehead are much better than those on eyes, noses and mouths. The main reason is that the surfaces of cheeks and forehead are more smooth and homogenous in terms of geometric characteristics, while eyes, nose and mouths reveal complicated structures and small details.

The metric of quality assessment can only coarsely indicate the performance of the semantic segmentation because of the following two factors: firstly, the amount of 3D points on different types of semantic components on the same Buddhist head varies quite largely, and secondly, the density of 3D point clouds is different on different Buddhist heads. This makes the IoU not comparable, in fact. In the future, the exact outline of each segment needs to be extracted and compared with that of ground truth data by using parameters of shape and size of polygons.

The LVM model, SAM, demonstrates strong segmentation capabilities using only text and initial FTG prompts.

Nevertheless, there is much space to improve the performance of the proposed method. In the current study, the FTG is very simply established by only indicating the topological centre of the graph. In the next stage, the FTG can be organized so that it can describe topological relations between any two components. This will certainly increase the accuracy of segmentation. On the other hand, the performance can be improved by using better text prompts.

Acknowledgement: this work is supported by the project Digitalization for Culture Heritage [Grant number UTF-2024/10124], funded by the Norwegian Directorate for Higher Education and Skills

References

Gezawa, A. S., Bello, Z. A., Wang, Q., Yunqi, L., 2022. A voxelized point clouds representation for object classification and segmentation on 3D data. *The Journal of Supercomputing*, 78(1), 1479-1500. https://doi.org/10.1007/s11227-021-03899-x

He, S., Ding, H., Jiang, X., & Wen, B., 2024. Segpoint: Segment any point cloud via large language model. In European Conference on Computer Vision (pp. 349-367). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72670-5_20

Jhaldiyal, A., Chaudhary, N., 2023. Semantic segmentation of 3d lidar data using deep learning: a review of projection-based methods. *Applied Intelligence*, 53(6), 6844-6855.

Li, L., 2018. Examining image-making techniques, sectarian disputes, and the establishment of Buddhist aesthetics through the Buddha's thirty-two marks. *Art Exploration*, 32(5), 11. (original article in Chinese) http://doi.org/10.13574/j.cnki.artsexp.2018.05.005

Li, Z., Shan, J., 2022. RANSAC-based multi primitive building reconstruction from 3D point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 185, 247-260.
Morgan, C. (2022). Current digital archaeology. Annual Review of Anthropology, 51(1), 213–231. https://doi.org/10.1146/annurev-anthro-041320-114101

Pan, Y., Hou, M., Su Y., Li, H., Wei, S., Fan, H., 2025. A CityGML ADE for modeling Buddhist statues in 3D with semantic information. *Journal of Cultural Heritage*, 76 111-120. https://doi.org/10.1016/j.culher.2025.09.001

Poux, F., Mattes, C., Selman, Z., Kobbelt, L., 2022. Automatic region-growing system for the segmentation of large point clouds. Automation in Construction, 138, 104250.

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., Ghanem, B., 2022. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in neural information processing systems*, 35, 23192-23204.

Zhang, M., 2022. Analysis of the Buddha's meditative kneeling posture. *Cultural Relics*, (9), 86–96. (original article in Chinese) http://doi.org/10.13619/j.cnki.cn11-1532/k.2022.09.006