

AI-Driven Digital Representation of Minnan Stone Lion Sculptures

Bo Shu^{1,2}, Xinyao Liu¹, Haoran Wang¹, Gerilutu Bao^{1*}, Yue Yuan^{2*}

¹ School of International Communication and Arts, Hainan University, Haikou, China

² Minnan Stone Carving Art and Industry Development Research Center, Xiamen Institute of Technology, Xiamen, China

Email Address

2413908845@qq.com (Gerilutu Bao), yuanyue@xit.edu.cn (Yue Yuan)

*Correspondence: 2413908845@qq.com

Received: 7 November 2025; Accepted: 3 December 2025; Published: 19 December 2025

Abstract:

With the rapid development of digital technologies applied to traditional cultural research, achieving efficient transformation of traditional sculptures from two-dimensional(2D) representations into three-dimensional(3D) forms—without compromising stylistic integrity or cultural recognizability—has become one of the most prominent topics in the field. This study takes the Minnan stone lion, a form that embodies distinctive regional aesthetics, as the research object and proposes an AI-based workflow for converting 2D images into 3D models. Through on-site surveys, image collection, and literature review, a multimodal image database of Minnan stone lions was established. A LoRA fine-tuned diffusion model was then employed to achieve stable 2D image generation in the Minnan style. Based on these results, the study integrates image-to-video and NeRF/3DGS-inspired reconstruction methods to produce digital models that are structurally coherent and rich in detail. The results were evaluated in terms of morphological similarity, detail fidelity, and cultural recognizability. Experiments demonstrate that this workflow can effectively balance traditional aesthetic style and geometric precision without relying on complex 3D modeling experience, offering a reusable framework for the digital preservation and artistic reinterpretation of traditional stone sculptures. Furthermore, the applicability of the “Diffusion Model + LoRA + NeRF/3DGS” combination to Minnan stone lion imagery is validated, providing a methodological reference for the digital expression and research of intangible cultural heritage.

Keywords:

Minnan Stone Lion Sculpture, Artificial Intelligence, Diffusion Model, LoRA, Digital Preservation

1. Introduction

The traditional stone lions of southern Fujian, as iconic folk stone carvings, are widely distributed at the entrances of ancestral halls, temples, shrines, and traditional neighborhoods. Their exaggerated forms, rhythmic mane patterns, and narrative decorative motifs stand in stark contrast to the solemn, realistic styles of northern stone lions. Southern Fujian stone lions emphasize dynamic movement and decorative

elements, serving as crucial material carriers for studying regional sculptural language and folk belief spaces. However, after centuries of weathering, early stone lion carvings in southern Fujian have suffered severe deterioration and face risks of relocation or demolition. Traditional restoration methods relying on photographs and hand-drawn records struggle to preserve complete information, while existing 3D data collection technologies require specialized equipment and complex modeling processes [1]. For stone lions with intricate structures and numerous open spaces, the modeling costs and time-consuming procedures make large-scale digital reconstruction impractical.

In recent years, AIGC technologies—particularly diffusion model-based image generation—have rapidly advanced in cultural heritage digitization. Yet most research remains at the level of 2D stylized representations or partial 3D reconstructions, lacking systematic experimental approaches for “specific objects.” The key challenge now lies in how to leverage diffusion models and 2D-to-3D conversion techniques to achieve rapid transformation from flat images to three-dimensional models while preserving the original stylistic features and details of the subjects. In this context, the study centers on the “AI Experiment for 2D-to-3D Conversion of Minnan Stone Lion Statues”, exploring rapid digital transformation of physical objects. Starting from Minnan stone lion statues—a distinctive regional cultural symbol—this research combines field surveys with digital technologies to develop an experimental workflow for converting 2D images into 3D models. Key focuses include preserving stylistic integrity, ensuring structural rationality, and maintaining result reusability. Specifically, the study first establishes a multimodal data repository of Minnan stone lions through field photography and literature analysis, capturing typical specimens, design variations, and intricate details. Building upon this foundation, the research introduces diffusion models and specialized LoRA fine-tuning algorithms to generate 2D images from multiple perspectives. Subsequently, employing the latest image-to-video generation model to expand 2D images into multi-angle rotating sequences, the study then applies 3D reconstruction workflows from NeRF/3DGS to transform these images into structurally complete 3D models [2,3,4]. NeRF/3DGS is referenced as an inspiration for multi-view-consistency-driven reconstruction rather than a strict scene-by-scene optimization implementation. The core proposition of this study is that a novel methodology—combining “object-specific datasets, style-controlled 2D image generation, and multi-perspective 3D reconstruction”—enables efficient conversion of Minnan stone lion sculptures from flat images to 3D models [5]. This approach eliminates the need for costly traditional 3D scanning and extensive manual modeling, offering a groundbreaking solution for the digital preservation, exhibition, and reuse of traditional sculptural cultural heritage [6].

2. Materials and Methods

This chapter introduces the methodological framework used in this study. The entire workflow consists of three stages: Construction of a multimodal Minnan stone lion database through field research, image collection, and literature review; Stable 2D image generation using a LoRA fine-tuned diffusion model; Transformation of 2D images into 3D structures through a reconstruction workflow that combines image-to-video, multi-view synthesis, and NeRF/3DGS-like modeling, followed by multi-dimensional evaluation of results.

2.1 Multimodal Database Construction

At the data level, this study constructs a multimodal database of Shishi, Minnan, which covers “typical individuals + representative details” through field shooting, archival sorting and open data acquisition.

Field Research and Collection Documentation: Through field surveys of ancestral halls and temples in Southern Fujian, we conducted multi-angle photography and baseline mapping of representative stone lion specimens. We systematically organized images and descriptions of Minnan stone lions from museums and cultural heritage institutions across the Taiwan Strait, establishing foundational image and text databases. Representative stone lion images and 3D models were collected and categorized by design type, posture, and structural components, forming a multimodal dataset for model training and evaluation.

Multimodal data processing and annotation: The collected photos undergo denoising, format correction, and labeling. Annotation standards are established using frameworks such as “individual ID—viewing angle—local components—pattern type”. For samples with distinctive features and suitable conditions, laser scanning and multi-view photogrammetry are conducted to obtain corresponding 3D mesh data, forming a “2D-3D correspondence” subset of sample data. These 3D meshes are collected only as a small alignment/validation subset and are not required inputs for the proposed 2D-to-3D workflow.

Experimental evaluation and application verification: Quantitative analysis of 2D to 3D effect is carried out around the dimensions of morphological similarity, detail restoration, structural manufacturability and subjective aesthetic evaluation, and the best data is tested.

Data structure optimization and knowledge graph construction: taking “lion type, building function, pattern characteristics, folk custom meaning” as core nodes, constructing Minnan stone lion knowledge graph, sorting out the variation and commonality of stone lion image under different regions and different building system, providing support for model prompt word design and result interpretation, and trying to cover most regions of prompt word data.

2.2. Diffusion Models and LoRA Fine-Tuning

At the model level, this study adopts diffusion models such as FLUX/SDXL as the foundation [7,8]. By summarizing the modeling rules and folk semantics of Minnan stone lions, it employs the LoRA fine-tuning strategy to infuse the ‘Minnan stone lion’ style. This approach enables the model to maintain its general generative capabilities while enhancing its expressive and control capabilities for specific objects .

Data preparation: Select a curated set of high-quality images from the Minnan Stone Lion database, covering diverse types (e.g., front-facing lions, column-head lions) and perspectives (front, side, oblique, overhead). Perform standardized image cropping and color correction, then remove samples with excessive background interference, severe occlusion, or extreme lighting conditions. For each image, generate English/Chinese prompt terms and negative terms, extracting high-frequency keywords such as “Minnan stone lion,” “guardian lion,” “temple gate,” “curled mane,” and “carved stone base” to establish a foundation for subsequent inference-stage prompt design.

Training parameter configuration: The learning rate is maintained at approximately $1e-4$, with fine-tuning based on loss function convergence. The Rank value is set between 16 and 32 to balance style representation and model size. Training steps are determined by evaluating both loss curves and image quality, typically spanning several thousand iterations. The EMA (Exponential Moving Average) and checkpoint mechanisms are implemented to enable seamless rollback to intermediate versions with optimal performance.

Iterative Monitoring and Mid-term Evaluation: Using seed values and standard prompts at different stages to generate control samples, we observe changes in lion-head proportions, mane textures, and base forms after introducing LoRA at each stage, identifying the most suitable seed values and prompts. Through subjective evaluation and simple image classification, we assess whether the generated structures meet the recognizability and diversity standards of the “Minnan Stone Lion” style, avoiding overfitting or excessive stylistic deviations, as shown in Figure 1.

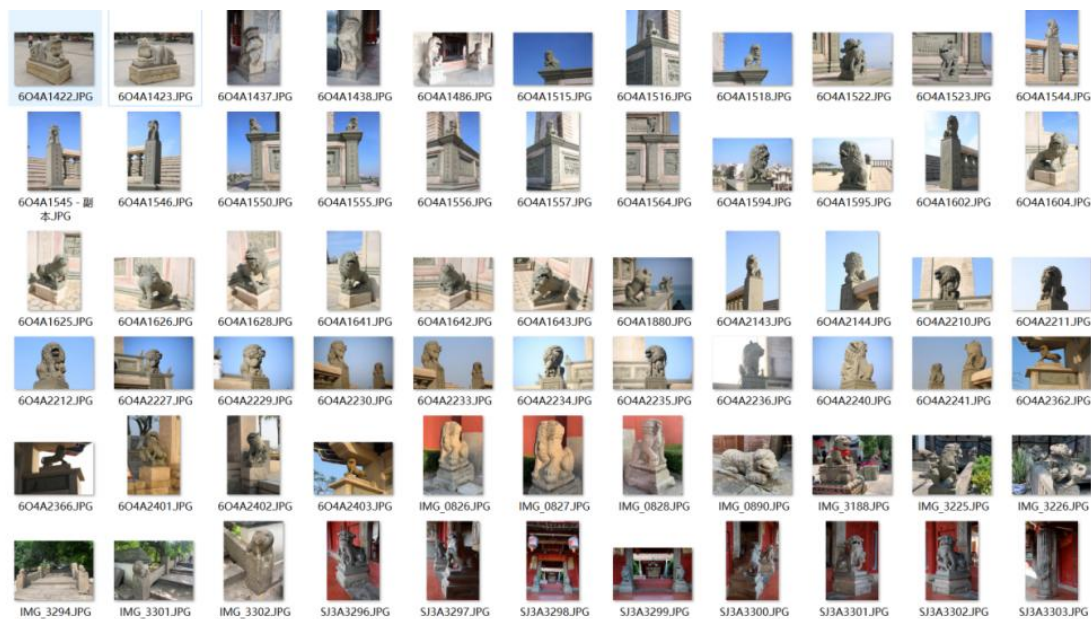


Figure 1. Examples from the Minnan stone lion dataset.

2.3. 360-Degree Rotation via Image-to-Video

At the process level, this study first utilizes a diffusion model combined with Minnan Shishi LoRA to generate high-quality static images. These are then extended into 360-degree rotating videos through an image-to-video model. In the ComfyUI node flow, the base model and Minnan Shishi LoRA are loaded to generate a high-quality static frame of Minnan Shishi from text prompts and reference images, serving as a “reference perspective” image (Figure 2). An image-to-video generation model is then applied, with a dedicated “rotation” LoRA overlay, to control the generation of a 360-degree rotating video from text prompts. This expands a single static image into a continuous sequence of multi-angle views including front, side, back, and top-down perspectives. Combined with a video super-resolution module for enhancement, the video maintains motion continuity while improving spatial resolution per frame, providing finer texture and contour information for subsequent 3D reconstruction (Figure 3).



Figure 2. Text-to-image module for Minnan stone lions (diffusion model + LoRA).

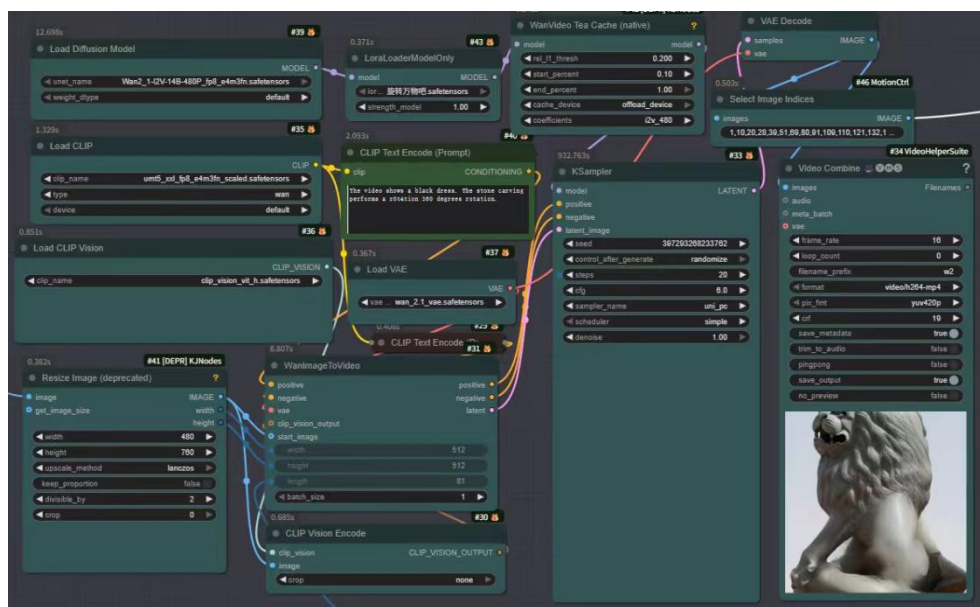


Figure 3. Multi-view generation module for Minnan stone lions (image-to-video).

2.4. NeRF/3DGS-Inspired 3D Reconstruction Workflow

Existing research has proposed generative models for rapid 3D object reconstruction from single images, which can serve as lightweight baselines for 2D-to-3D conversion.

NeRF represents scenes as continuous neural radiance fields and achieves perspective synthesis through scene-by-scene optimization via differential rendering [9]. 3D Gaussian Splatting employs three-dimensional Gaussians as explicit representations to enable real-time radiance field rendering and efficient perspective browsing. During the 3D reconstruction phase, this study did not directly adopt the classic NeRF/3DGS framework for scene-by-scene optimization of individual Minnan-style stone lion sculptures, but instead drew inspiration from its “implicit 3D modeling driven by multi-view consistency” concept. For 3D generation with multi-

view inputs, existing workflows propose models that generate high-resolution 3D content through multi-view consistency fusion. For multi-view 3D generation, recent approaches have explored high-resolution 3D content creation driven by multi-view consistency fusion. This study treats rotated video frame extraction as multi-view input and completes the 3D mesh generation of Minnan stone lions under this framework. Such models are more mechanistically similar to “conditional generative 3D modeling,” while NeRF/3DGS belongs to “scene-by-scene optimization based on differential rendering.” Although they differ in training methods and geometric accuracy, both share the fundamental approach of “reconstructing 3D structures using multi-view images.” The specific workflow is as follows.

Multi-view image dataset construction: The 360-degree rotating video is treated as a multi-view image dataset centered on the same object. Key frames are extracted and fused to generate high-resolution 3D representations. For the same Minnan stone lion statue, a multi-view image dataset is constructed by extracting key frames and estimating their camera trajectories.

Implicit 3D representation learning: The multi-view frame input is fed into a NeRF/3DGS-like reconstruction module based on img2vid weights to learn the implicit 3D representation (voxel field or dense point cloud) of Shishi.

Surface reconstruction and mesh extraction: Using surface reconstruction algorithms (e.g., Marching Cubes) to extract the initial 3D mesh model from the implicit representation, which serves as the foundation for subsequent mesh optimization and texture mapping, as shown in Figure 4.

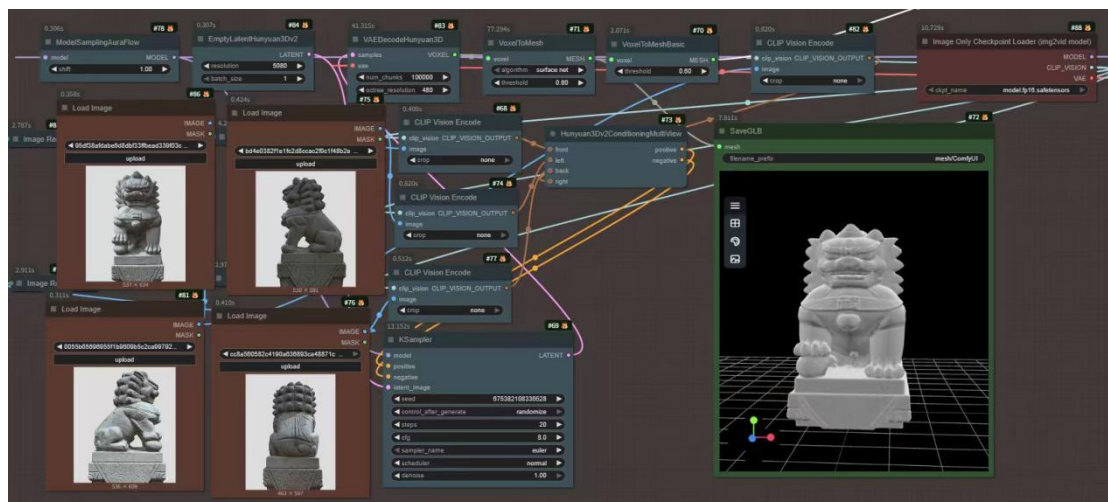


Figure 4. NeRF/3DGS-inspired reconstruction module from 360-degree imagery to a 3D model for a Minnan stone lion.

Mesh optimization and material mapping: Perform topology optimization, hole repair, and normal recalculation on the initial mesh, with moderate thickening of critical weak areas. Combine multi-view images for texture blending and UV unwinding to generate high-quality texture maps suitable for rendering and multi-platform display.

Export and Scene Integration: The optimized Shishi 3D model can be exported in standard formats (e.g., FBX/OBJ/GLB) for real-time display and interactive browsing on game engines or WebGL platforms, providing geometric foundations for potential 3D printing, sculpting, and other physical production processes.

2.5. Evaluation Metrics and Experimental Design

At the application level, this study conducted multi-group comparative experiments across dimensions including morphological similarity, detail fidelity, structural rationality, and subjective aesthetic evaluation to quantitatively and qualitatively analyze 2D-to-3D conversion effects. For morphological similarity and structural rationality, physical photographs were compared with 3D models, with scoring performed on overall proportions, center of gravity positioning, and dynamic postures. The 3D software's sectioning and measurement tools were also utilized to verify the model's volume relationships and support logic.

Detail and pattern restoration: The evaluation focuses on whether details such as mane curls, pearl-biting, lion interaction, and base edge decorations are effectively preserved, using a combination of localized magnification observation and expert scoring.

Potential manufacturability (structural manufacturability prediction): Using mesh analysis tools to evaluate key parameters like local thickness, cantilever length, and connectivity of fine components, this method provides geometric-level predictions about a model's suitability for subsequent engraving or 3D printing without actual printing.

Subjective aesthetic evaluation and cultural identification: Through the questionnaire and interview of the audience with Minnan cultural background and non-local background, the evaluation of different models in the subjective dimension of "Minnan sense", "traditional sense", "affinity" and so on is compared.

3. Results and Discussion

This chapter provides a comprehensive analysis of the "2D Image-Multi-Perspective Sequence-3D Model" workflow output proposed in this study. The evaluation focuses not on singular algorithmic metrics, but rather on practical implementation, assessing whether the generated 3D digital models of Minnan stone lions achieve the integrated effects of "distinctive style, logical structure, and readable details" in digital presentation contexts. The analysis will center on three key dimensions: First, the stability of overall contours and proportional relationships (including volumetric perception, center of gravity, and structural logic); Second, the retention of critical local features (such as facial expressions, mane rhythm, and claw and base decorations that significantly influence the recognition of "Minnan characteristics"); Third, the analysis of common distortion types and their causes in areas like perforated structures, thin-layer decorations, and high-frequency patterns. Through comparative observation of typical samples and problem identification, this chapter further clarifies the applicability and limitations of the workflow in digital expression and creative assistance, providing a foundation for subsequent conclusions and future research directions.

3.1. 3D Model Generation and Overall Form Analysis

In the experimental phase, the research team selected representative Minnan stone lion cases (including temple gate guardian lions, column-head lions, and pedestal lions) from the database. Using the "Minnan stone lion LoRA + 2D-to-3D workflow" approach, they generated 3D models (Figure 5) and compared them with unadjusted universal models. The results demonstrated that the universal model often exhibited issues like "small face" and "slender torso" in head-to-body proportions and volume,

while the LoRA-tuned model accurately captured the Minnan stone lion's distinctive features of “large head, plump body, and exaggerated dynamic posture.” The 3D-reconstructed models not only closely resembled the original prototypes in overall form but also better aligned the center of gravity and support relationships with the structural logic of traditional stone carvings.

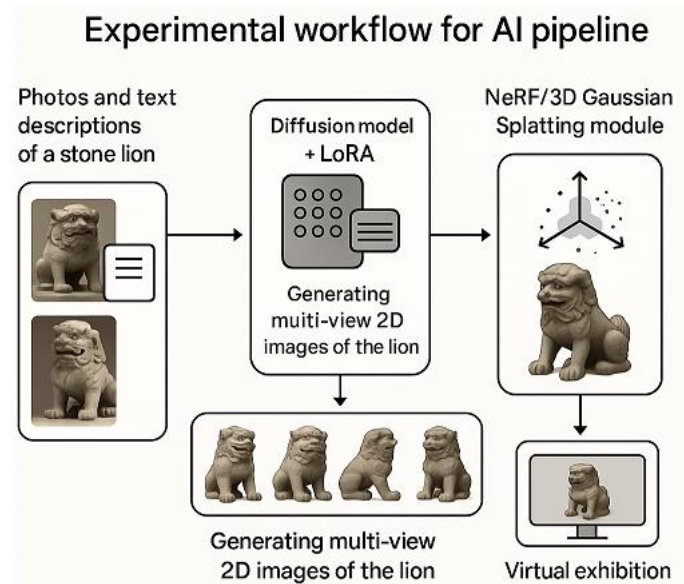
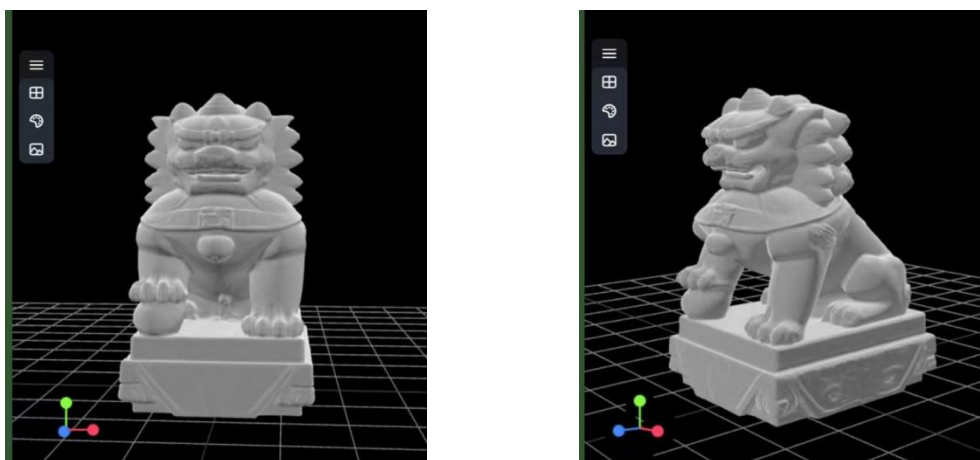


Figure 5. Schematic of the 2D-to-3D workflow for Minnan stone lions.

3.2. Detail and Ornament Fidelity

The study conducted comparative analyses on specific local details, focusing on mane texture patterns, lion interaction elements, and base decorative motifs. While traditional manual modeling achieves high precision, it requires extensive manual labor. Models generated using generic approaches often exhibit blurred details and significant deviations from Minnan architectural aesthetics. In contrast, the workflow-generated models demonstrate greater consistency in mane direction, texture density, and eye-tooth morphology. Although less detailed than traditional manual methods, they preserve rich local details while maintaining stylistic coherence, and can reliably produce desired models as shown in Figure 6.



A. Three-dimensional reconstructed model of the stone lion (front view).

B. Three-dimensional reconstructed stone lion model (side view).

Figure 6. Reconstructed Minnan stone lion models (A: front view; B: side view).

3.3. Subjective Aesthetic Evaluation and Cultural Identifiability

Through surveys and interviews targeting audiences with Minnan cultural backgrounds and non-native backgrounds, this study compared the evaluations of different models in subjective dimensions such as “Minnan identity,” “traditionalism,” and “affinity.” The results showed that the 3D model using Minnan Shishi LoRA scored higher in “geographical style recognition” and “formal familiarity,” with respondents more readily associating it with spatial imagery like Minnan ancestral halls and temples. This indicates that object-specific LoRA demonstrates distinct advantages in conveying cultural symbols.

3.4. Discussion

Based on the experimental and analytical findings, the workflow developed in this study—combining a diffusion model with LoRA and NeRF/3DGS-inspired 3D reconstruction for converting Minnan-style stone lion plans into 3D models—achieves a balanced approach between style preservation, local detail quality, and structural rationality [10]. Notably, the 3D reconstruction phase avoids traditional scene-by-scene optimization for individual lions. Instead, it adopts a multi-view consistency-driven implicit modeling strategy, using 360-degree rotation video frames as input to generate a mesh representation through a pre-trained 3D generative model in a single pass. This methodology contrasts with NeRF’s scene-by-scene differential rendering optimization paradigm while matching 3D Gaussian Splatting in rendering efficiency and interactive display performance.

The experimental results show that the 3D model generated by the workflow is superior to the general model without fine-tuning in terms of overall scale, fur and facial expression details, and structural rationality. This preliminary study validates the feasibility and advantages of the “diffusion model + LoRA + NeRF/3DGS-like 3D reconstruction” approach for the specific case of Minnan stone lions.

4. Conclusions and Outlook

Through sustained research and implementation, this project has achieved significant milestones in digital representation and modeling of Minnan stone lions. At the data infrastructure level, systematic field surveys and image acquisition were conducted using representative Minnan sites and museum collections as experimental subjects. Leveraging AIGC technology, we established a comprehensive database integrating 2D images, 3D meshes, and textual descriptions, providing a robust foundation for analyzing lion morphology, training models, and future research. Methodologically, we developed a node-based, reproducible workflow for 2D-to-3D conversion through diffusion model style learning, multi-view image generation, and NeRF/3DGS-inspired 3D reconstruction. This workflow has been validated through multiple case studies, establishing an operational and scalable AI-assisted modeling framework. The project has produced structurally complete 3D digital models with high-definition details, successfully rendered in 3D environments, and tested for interactive applications. These models now serve as practical templates for classroom teaching and cultural design, while providing reliable geometric references and stylistic guidance for virtual exhibitions, AR-based tours, and related design practices. Meanwhile, by leveraging “Minnan Stone Lion Culture + AI Modeling” as a starting point, this initiative has facilitated interdisciplinary integration across art, design, architecture, and computer science. Through course practices and workshop formats,

it explores cross-disciplinary project-based teaching models to enhance students' understanding of the synergy between local culture and cutting-edge digital technologies. Ultimately, a research framework featuring “customized object database construction + LoRA training model fine-tuning + 2D-to-3D system conversion” has been developed. This approach demonstrates strong transferability and scalability, and can be adapted for digital research of other intangible cultural heritage sculptures like memorial arch carvings, ancestral hall wood carvings, and wind lion statues. The solution provides replicable pathways and exemplary models for broader digital representation of cultural heritage.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Author Contributions

Bo Shu, Xinyao Liu, these two authors have contributed equally to this work.

Conceptualization: Bo Shu; Methodology: Bo Shu; Software: Bo Shu, Xinyao Liu; Validation: Bo Shu, Xinyao Liu; Formal analysis: Bo Shu, Xinyao Liu; Investigation: Bo Shu, Xinyao Liu; Resources: Haoran Wang, Gerilutu Bao, Yue Yuan; Data Curation: Haoran Wang; Writing – original draft preparation: Xinyao Liu, Haoran Wang; Writing – review and editing: Gerilutu Bao, Yue Yuan; Visualization: Xinyao Liu; Supervision: Bo Shu, Gerilutu Bao, Yue Yuan; Project administration: Bo Shu, Yue Yuan; Funding acquisition: Bo Shu.

Funding

This research was supported by the following projects:

(1) Fujian Provincial University Humanities and Social Sciences Research Base - Minnan Stone Carving Art and Industry Development Research Center 2024 Open Research Grant Project, titled “Integration of Tradition and Technology: Digital Research on Minnan Stone Carving Design and Processing” (Project No.: XGSDKF202405).

(2) Hainan University Scientific Research Startup Fund Project, titled “Visualization Research on the Design Application of Virtual Reality Technology in Public Art and Spatial Environments” (Project No.: kyqd(sk)2312).

Acknowledgments

The author would like to acknowledge the valuable support and insights provided by colleagues and collaborators during the research process.

References

- [1] Remondino, F. Heritage Recording and 3D Modeling with Photogrammetry and 3D Scanning. *Remote Sens.* 2011, 3, 1104-1138. Available online: <https://www.mdpi.com/2072-4292/3/6/1104> (accessed on 15 November 2025).
- [2] Kerbl, B.; Kopanas, G.; Leimkühler, T.; Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* 2023, 42, 139:1–

- 139:14. Available online: <https://arxiv.org/abs/2308.04079> (accessed on 15 November 2025).
- [3] Hu, E.J.; Shen, Y.; Wallis, P.; et al. LoRA: Low-Rank Adaptation of Large Language Models. Status (preprint; arXiv). 2021. Available online: <https://arxiv.org/abs/2106.09685> (accessed on 15 November 2025).
- [4] Guo, Y.; Yang, C.; Rao, A.; et al. AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning. Status (preprint; arXiv). 2023. Available online: <https://arxiv.org/abs/2307.04725> (accessed on 15 November 2025).
- [5] Tang, J.; Chen, Z.; Chen, X.; Wang, T.; Zeng, G.; Liu, Z. LGM: Large Multi-View Gaussian Model for High-Resolution 3D Content Creation. Status (preprint; arXiv). 2024. Available online: <https://arxiv.org/abs/2402.05054> (accessed on 15 November 2025).
- [6] Zhang, J. Folk Culture of Kang Tou Stone Lions of China. In Proceedings of the 2014 International Conference on Social Science (ICSS-2014); 2014; <https://doi.org/10.2991/icss-14.2014.112>.
- [7] Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In Advances in Neural Information Processing Systems (NeurIPS); 2020. Available online: <https://arxiv.org/abs/2006.11239> (accessed on 15 November 2025).
- [8] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 10684–10695.
- [9] Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Proceedings of the European Conference on Computer Vision (ECCV); 2020. Available online: <https://arxiv.org/abs/2003.08934> (accessed on 15 November 2025).
- [10] Zhang, L.; Rao, A.; Agrawala, M. Adding Conditional Control to Text-to-Image Diffusion Models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–6 October 2023; pp. 3836–3847. Available online: https://openaccess.thecvf.com/content/ICCV2023/html/Zhang_Adding_Conditional_Control_to_Text-to-Image_Diffusion_Models_ICCV_2023_paper.html (accessed on 15 November 2025).



© 2025 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)