

Dunhuang_Faces: A Dataset Enhanced through Image Processing for Cultural Heritage and Machine Learning



Zizhan Zhang¹, Yingchun Cao^{2*}

¹ School of Architecture and Art, Hebei University of Architecture, Zhangjiakou 075000, China

² School of Civil Engineering and Architecture, Jiangsu University of Science and Technology, Zhenjiang 212003, China

Corresponding Author Email: yc_cao@163.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420148>

ABSTRACT

Received: 9 May 2024

Revised: 17 November 2024

Accepted: 1 December 2024

Available online: 28 February 2025

Keywords:

image dataset, super-resolution image processing, Dunhuang murals, Mogao Caves, art-focused machine learning datasets, high-fidelity image restoration, digital mural analysis, cultural heritage image reconstruction

The preservation and analysis of Dunhuang mural art face significant challenges due to the severe degradation and fragility of these invaluable cultural artifacts. To address the scarcity of high-quality data for research, we introduce Dunhuang_Faces, a super-resolution-enhanced dataset specifically designed for applications in cultural heritage studies and machine learning. By leveraging advanced super-resolution models, we restored and enhanced degraded mural images, creating a dataset that maximizes the utility of the limited available data. Dunhuang_Faces comprises three subsets: Dunhuang_Faces-20k, Dunhuang_Faces-3k, and Dunhuang_Faces-500, each processed through super-resolution techniques and standardized to support a broad range of tasks. Through experimental validation, we demonstrate its effectiveness across cultural heritage digitization analyses and machine learning applications, highlighting its versatility and potential. This work provides a valuable standardized dataset for the cultural heritage field and enriches art-related datasets in machine learning with the unique style and imagery of Dunhuang murals. It establishes a technical foundation for research in both domains, offering critical resources for the preservation, analysis, and reutilization of historical artworks.

1. INTRODUCTION

In recent years, advancements in technology and the urgent need for preservation have accelerated the digitization of cultural heritage [1-3]. This digitization has opened new avenues for analysis, interpretation, and dissemination [4, 5], integrating computational methods for tasks like 3D reconstruction, material analysis, and historical documentation [6]. However, this progress has underscored the growing demand for high-quality, diverse datasets to support emerging research and applications.

Currently, most datasets in the cultural heritage domain are developed in isolation, tailored to the specific needs of individual projects or teams [7-10]. This fragmented approach results in datasets that often lack standardization, reusability, and scalability, limiting their broader applicability. Moreover, cultural heritage research teams rarely participate in open-source initiatives, leading to minimal data sharing across the community [11]. As a result, researchers frequently replicate labor-intensive tasks, such as field surveys, architectural measurements, and raw data preprocessing. The lack of standardized, accessible datasets hampers the adoption of computational methods and restricts interdisciplinary collaboration with fields like machine learning and artificial intelligence [12].

The rapid advancements in machine learning have been fueled by large-scale datasets, which form the foundation for training and evaluating advanced models [13]. For example,

ImageNet [14], with over 14 million labeled images, catalyzed the success of AlexNet, marking the start of modern deep learning. Building on this, COCO [15] added richer annotations like bounding boxes, segmentation masks, and captions, becoming a benchmark for object detection and multi-task learning. Recent datasets address evolving challenges. COCO-O [16] explores robustness under distribution shifts, while COCONut [17] enhances segmentation benchmarks with diverse categories. OpenImages [18] offers millions of diverse, real-world samples, and specialized datasets such as Cityscapes [19] and ADE20K [20] focus on urban scene parsing and indoor/outdoor segmentation. Pretraining datasets have scaled up model capabilities. ImageNet-21k [21] and JFT-300M [22] provide billions of labeled samples, enabling the training of large-scale architectures like Vision Transformers (ViTs) [23]. LVIS [24] addresses long-tail distributions by focusing on rare categories. Multimodal datasets like Visual Genome [25] and LAION-5B [26] expand research into object relationships and text-image models. For video tasks, datasets like Kinetics-700 [27] advance action recognition. Collectively, these datasets drive innovation, ensuring machine learning models remain robust, adaptable, and capable of addressing real-world challenges.

Despite the proliferation of general-purpose datasets, the cultural heritage domain remains underrepresented in machine learning research [28]. Notable datasets like WikiArt [29] and the Rijksmuseum dataset [30] have supported tasks

such as style classification and artwork analysis but fail to meet the broader and complex needs of cultural heritage research [31]. Challenges in this field arise from the fragile, perishable nature of cultural heritage and the incomplete preservation of artifacts [32-34]. Moreover, the high volume of data with limited utility complicates data collection and processing. Existing datasets are constrained by limited volume, low diversity, and inadequate representativeness, hindering the applicability of machine learning models. Addressing these limitations requires advanced image processing, optimized use of archaeological data, and the creation of specialized, curated datasets. These efforts will enhance machine learning capabilities and foster interdisciplinary research in cultural heritage.

We collected and organized mural scan data from the Dunhuang grotto temples [35], restored it using a super-resolution model based on Stable Diffusion, and introduced the Dunhuang_Faces dataset. As shown in Figure 1, this dataset comprises several subsets with high-quality annotations and diverse data for various applications. Dunhuang_Face-20k includes 19,126 face images with XML-format annotations, enabling large-scale training and analysis. Dunhuang_Face-3k focuses on 3,282 relatively well-preserved faces, while Dunhuang_Face-500 provides 517 best-preserved faces with fine-grained details for precision tasks. This comprehensive dataset not only integrates expertise from cultural heritage professionals but also incorporates insights from the machine learning community, fostering interdisciplinary research. By providing a robust foundation for tasks like object detection, restoration, and generative modeling, Dunhuang_Faces enriches the resources available to both cultural heritage scholars and computer scientists, encouraging deeper exploration of this essential but underrepresented domain.

2. RELATED WORK

2.1 Traditional research

The digitalization of cultural heritage has been explored extensively across multiple domains, leading to diverse applications. Geographic Information Systems (GIS) are widely used for spatial analysis, risk assessment, and dynamic monitoring, proving invaluable in evaluating site distribution and risks from natural disasters or urban development [36-38]. Integrating remote sensing data with GIS enhances monitoring and protection of large-scale heritage areas, supporting sustainable management [39, 40]. Digital archives and virtual museums have also gained prominence. High-resolution imaging and 3D scanning technologies enable the creation of extensive digital repositories that preserve artifacts and improve global accessibility for education, research, and public engagement [41, 42]. Virtual museums offer immersive experiences, fostering a deeper understanding of heritage [43, 44]. Digital mapping and digital twin technologies are transformative tools for heritage preservation. Laser scanning and UAV photogrammetry generate accurate 3D models for restoration and conservation [45, 46]. Digital twins enable real-time monitoring and predictive modeling, promoting dynamic site management and sustainable preservation strategies [47, 48].

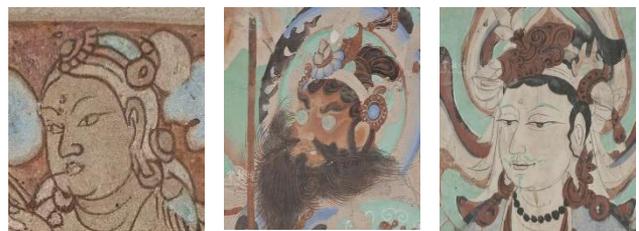
Dunhuang_Faces-20K



Dunhuang_Faces-3K



Dunhuang_Faces-500



Note: Dunhuang_Faces-20k includes all digitized faces. Dunhuang_Faces-3k consists of standardized, super-resolved images of faces with complete facial structures. Dunhuang_Faces-500 comprises standardized, super-resolved images of faces with complete features, including facial structures, headwear, and ornaments.

Figure 1. Examples from the Dunhuang_Faces dataset, illustrating the three subsets

Beyond digital preservation and management, substantial efforts have reimagined cultural heritage through innovative technologies. Multimedia technologies like augmented reality (AR) and interactive platforms have revolutionized its presentation and dissemination. AR overlays historical narratives and contextual information onto real-world settings, enriching public engagement and educational experiences [49, 50]. Multimedia platforms combining visual, auditory, and interactive elements amplify the impact of cultural storytelling [51, 52]. Data-driven approaches, including machine learning and big data analytics, are increasingly used for artifact classification, restoration prediction, and public perception analysis. Machine learning frameworks automate the categorization of digital heritage data, improving efficiency and scalability in managing complex datasets [12, 53]. These methods streamline workflows and create new possibilities for understanding and preserving cultural heritage. The digitalization of cultural heritage is an essential trend in preservation and research. However, current efforts are often hindered by the lack of diverse and comprehensive datasets. Existing resources are limited in volume and fail to capture the variety of heritage types and forms, restricting the effectiveness of digital applications. Developing standardized datasets that integrate diverse heritage-spanning tangible, intangible, local, and global elements is crucial. Such datasets would enhance interoperability across disciplines and platforms, enabling robust data-driven approaches and fostering interdisciplinary collaboration. They would also drive innovation and advance cultural heritage digitalization.

Table 1. Comparison of the Dunhuang_Faces dataset and its subsets (Dunhuang_Faces-20k, Dunhuang_Faces-3k, Dunhuang_Faces-500) with widely used general-purpose and artistic datasets

Dataset Name	Number of Classes	Resolution	Number of Images	Description
ImageNet [14]	~20,000	Variable	14 million	General object and scene images
COCO [15]	80	Variable	330,000	Everyday objects in context
CIFAR-10 [54]	10	32×32	60,000	Low-resolution object images
CIFAR-100 [54]	100	32×32	60,000	Low-resolution object images
OpenImages [18]	~600	Variable	9 million	Real-world object images with annotations
Cityscapes [19]	30	1024×2048	25,000	Street-level urban scene images
ADE20K [20]	~3,000	Variable	25,000	Indoor and outdoor scene images
WikiArt [29]	~27	Variable	80,000	Artistic paintings and styles
Rijksmuseum [30]	N/A	Variable	112,000	Historical artwork photographs
DEArt [31]	12	Variable	~70,000	European artwork images
SemArt [55]	N/A	Variable	44,000+	Image-text pairs from European fine art
PrintArt [56]	N/A	Monochrome	988	Compositionally annotated monochrome prints
NoisyArt [57]	N/A	Variable	N/A	Webly-supervised dataset with noisy labels
Art500k [58]	N/A	Variable	500,000+	Annotated with artist, genre, and art movement
ArtBench-10 [59]	10	256×256	N/A	Class-balanced dataset of 10 artwork categories
Dunhuang_Faces-30k	N/A	Variable	19,126	Large dataset of murals and face images
Dunhuang_Faces-3k	N/A	512×512	3,282	Focused on face structure and attributes
Dunhuang_Faces-500	N/A	512×512	517	High-quality images with descriptive tags

Note: The table highlights differences in class diversity, resolution, number of images, and domain focus, emphasizing the unique contribution of Dunhuang_Faces to ancient mural image research.

2.2 Datasets

Datasets are foundational to machine learning, serving as benchmarks and essential resources for training and evaluating models, as summarized in Table 1. ImageNet [14], with 14 million labeled images across thousands of categories, laid the groundwork for modern deep learning, catalyzing breakthroughs like AlexNet, which demonstrated the potential of convolutional neural networks. Building on this, COCO [15] introduced richer annotations for segmentation and image captioning, enabling the development of versatile models. Recent extensions like COCO-O [16] and COCONut [17] refined benchmarks, addressing robustness under distribution shifts and enhancing segmentation diversity. Smaller datasets such as CIFAR-10 and CIFAR-100 [54] remain widely used for classification due to their simplicity and accessibility, supporting rapid experimentation with novel algorithms. Open-Images [18] offers extensive real-world images and annotations for object detection and segmentation tasks. Task-specific datasets like Cityscapes [19] and ADE20K [20] target urban scene parsing and indoor/outdoor segmentation, respectively, with pixel-level annotations. These datasets collectively advance machine learning applications, enabling models to tackle diverse and complex challenges.

In the cultural heritage domain, various datasets address challenges such as style recognition, artifact analysis, semantic understanding, and generative modeling. These datasets are essential for domain-specific studies and interdisciplinary applications, advancing cultural preservation and computational research. For style recognition and art classification, WikiArt [29] provides over 80,000 artworks across diverse styles and movements, supporting style transfer and visual art analysis. Similarly, the Rijksmuseum dataset [30], with 110,000 digitized artworks and rich metadata, facilitates computational art preservation and historical research. DEArt [31] offers expert-curated annotations for European art, enabling tasks like pose estimation, composition

analysis, and art history investigations. Smaller datasets focus on semantic and compositional analyses. SemArt [55] pairs 44,000 European fine-art images with textual descriptions, supporting multi-modal retrieval and semantic understanding. PrintArt [56], a dataset of 988 annotated monochrome prints, enables fine-grained compositional analysis, while NoisyArt [57] simulates real-world challenges with noisy labels, promoting robust model development. Large-scale datasets expand machine learning in cultural heritage. Art500k [58], with 500,000 annotated artworks, supports classification, retrieval, and large-scale pretraining. ArtBench-10 [59], a curated dataset of 10 balanced artwork categories, benchmarks generative models and classification methods with standardized annotations. These datasets collectively support computational techniques while preserving and enhancing understanding of artistic and historical artifacts.

To enrich artistic datasets, we propose Dunhuang_Faces, a comprehensive dataset designed to support research in this field. The dataset includes multiple subsets and facilitates typological analysis, object detection, and generative applications, addressing diverse research needs. Unlike existing artistic datasets such as WikiArt and Rijksmuseum, which are often categorized by artistic styles, authorship, or other singular standards, Dunhuang_Faces focuses on the digitized portions of Dunhuang murals, standardizing human figure imagery to meet the demands of both cultural heritage and machine learning research. This dataset is not limited to supporting specific machine learning tasks; instead, it serves as a versatile resource capable of addressing various needs in traditional cultural heritage research, such as documentation and analysis, as well as computational tasks like object detection and generative modeling. By bridging these two domains, Dunhuang_Faces offers a more micro-level focus compared to other artistic datasets, emphasizing the meticulous standardization of existing data to facilitate broader interdisciplinary applications.

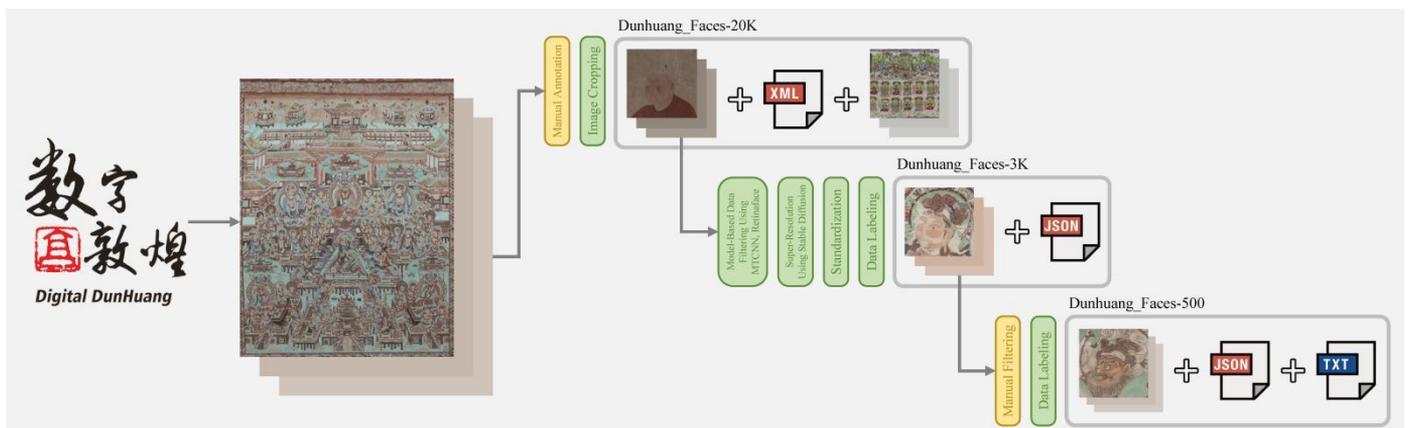
3. THE DUNHUANG_FACES DATASET

We extend our heartfelt gratitude to the Dunhuang Academy and numerous scholars for their extensive research on the grotto temples in the Dunhuang region, which has resulted in a wealth of invaluable resources. The Dunhuang_Faces dataset is derived from the Complete Collection of Dunhuang Murals [60], while additional information and references related to the Dunhuang murals have been drawn from the Digital Dunhuang platform (<https://www.e-dunhuang.com/>), published by the Dunhuang Academy. This dataset has been meticulously curated to support a variety of research tasks, ranging from face detection to historical and artistic analysis, by leveraging the rich visual and cultural elements embedded in the murals. Due to the severe damage and degradation of many murals, research data in this domain is inherently scarce. To address this limitation, we focused on collecting all valuable and intact facial images from the murals, ensuring that no potentially useful data was excluded. As a result, we did not prioritize balancing the sample sizes across different categories but instead aimed to maximize the dataset’s coverage of all available and valuable human face imagery. The dataset is divided into three subsets, each tailored to specific research objectives. Dunhuang_Faces-20k consists of 390 high-resolution images of complete murals from the Dunhuang caves, alongside 19,126 cropped face images annotated with bounding boxes in XML format. This subset is designed for large-scale tasks, such as face detection and localization, providing a comprehensive dataset for developing and benchmarking general-purpose face recognition models in cultural heritage

research. Dunhuang_Faces-3k includes 3,282 cropped face images with intact facial features. This subset is enriched with detailed attribute annotations, including facial identity, orientation, dynasty, and cave ID. It is particularly suitable for tasks such as facial attribute recognition, historical analysis, and transfer learning, where higher-quality annotated data is required. Dunhuang_Faces-500 focuses on 549 face images that are both well-preserved and accompanied by detailed attribute labels and comprehensive textual descriptions. This subset supports fine-grained studies, such as facial feature analysis, style-based generative modeling, and the contextual interpretation of artistic and historical details, making it an invaluable resource for exploring intricate cultural and artistic characteristics of Dunhuang murals. The Dunhuang_Faces Dataset provides a unique foundation for bridging computational methods and cultural heritage studies. By combining high-quality annotations, historical context, and artistic details, it offers unprecedented opportunities for interdisciplinary research and promotes the integration of technology into the study and preservation of cultural heritage.

3.1 Data creation

The Dunhuang_Faces Dataset was meticulously constructed through a well-defined and systematic workflow, as illustrated in Figure 2, to ensure high quality and consistency across all data subsets. Advanced standardization techniques were employed throughout the process, including resolution enhancement and bounding box normalization, to improve the usability and compatibility of the dataset for both cultural heritage studies and computer vision applications.



Note: The process begins with high-resolution mural scans from the Digital Dunhuang platform. Dunhuang_Faces-20k is generated by extracting and labeling faces with XML annotations. From this subset, Dunhuang_Faces-3k is curated through manual filtering and super-resolution processing, with detailed JSON annotations. Finally, Dunhuang_Faces-500 is created by selecting the highest-quality images, enriched with additional textual descriptions (TXT) and JSON metadata for fine-grained analysis.

Figure 2. Workflow for constructing the Dunhuang_Faces dataset

3.1.1 Data annotation

The first step in constructing the Dunhuang_Faces Dataset was the annotation of human faces depicted in the Dunhuang cave mural images. High-resolution mural images were meticulously analyzed to locate and label human faces with precise bounding box annotations. The labeling process ensured comprehensive coverage, accounting for variations in size, orientation, and position of faces across the murals. Based on these bounding box annotations, the faces were cropped from the original mural images, resulting in 19,126 cropped face images paired with corresponding XML annotation files. This effort culminated in the creation of the

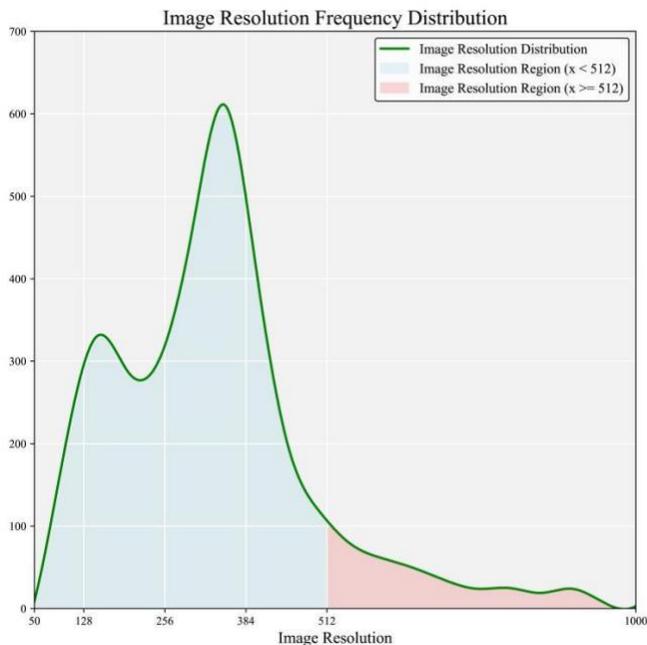
Dunhuang_Faces-20k subset, which consists of 390 complete mural images, 19,126 cropped face images, and detailed bounding box annotations. This subset serves as a foundational resource for tasks such as face detection and localization, enabling robust model development and evaluation in cultural heritage and computer vision research.

3.1.2 Data selection

To construct Dunhuang_Faces-3k, we employed a semi-automated workflow that combined advanced deep learning models with manual refinement. Specifically, two state-of-the-art face detection models, MTCNN and RetinaFace, were

utilized to analyze and filter the cropped face images from the Dunhuang_Faces-20k subset. These models were instrumental in identifying and prioritizing face images with relatively intact key facial features, such as eyes, nose, and mouth, even in cases of visible damage or degradation. The selection criteria for this subset focused on retaining images where key facial features were structurally complete, ensuring suitability for recognition, typological tasks, and studies on the structural characteristics of facial features in mural figures. The results were then meticulously reviewed and refined by experts to ensure the final subset met high standards of quality and consistency. This process yielded 3,282 cropped face images that retained sufficient detail for tasks requiring advanced analysis.

For Dunhuang_Faces-500, the selection process was conducted entirely manually. This subset was curated based on stricter criteria, prioritizing images with fully preserved facial features as well as intact headwear and ornaments, making it suitable for tasks such as style-based generative modeling, fine-grained analysis, and detailed studies of mural character design. The final selection of 549 images represents the highest quality samples in the dataset, making this subset particularly well-suited for fine-grained studies and detailed research.



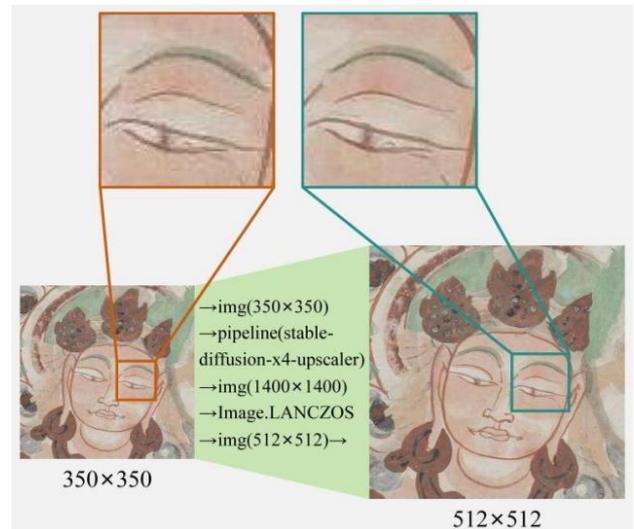
Note: The majority of images have resolutions concentrated around 350×350 pixels, with a smaller number above 512×512. Images with resolutions below 512×512 require super-resolution enhancement to ensure consistency and quality.

Figure 3. Pre-processing resolution distribution of extracted face images

3.1.3 Standardization and data enhancement

Due to the complex preservation conditions of the murals, much of the information they originally carried has been gradually lost over time. To fully exploit the existing facial data from Dunhuang murals, we applied advanced image processing techniques to enhance and maximize their value. A total of 3,282 facial images with intact facial structures were collected for analysis. As shown in Figure 3, the resolution of these images varied significantly, primarily due to the original scales of the figures depicted in the murals. Most face resolutions were concentrated within the range of 256×256 to

400×400 pixels. For images with resolutions below 512×512 pixels, a super-resolution model was employed to mitigate pixelation and jagged artifacts, effectively restoring visual clarity. For images with resolutions exceeding 512×512 pixels, careful down sampling was performed to ensure no degradation of image quality occurred.



Note: The process involves inputting a low-resolution image (350×350 pixels) into a super-resolution pipeline using the stable-diffusion-x4-upscaler model, upscaling it to an intermediate resolution (1400×1400 pixels), and then applying a down sampling operation (using the Lanczos filter) to produce a standardized output of 512×512 pixels with enhanced detail and clarity.

Figure 4. Workflow for super-resolution enhancement of face images

As shown in Figure 4, we employed a systematic approach to enhance image resolution and quality, utilizing the stable-diffusion-x4-upscaler, a state-of-the-art super-resolution model based on Stable Diffusion technology. This model is renowned for its ability to restore fine details and effectively reduce noise, making it particularly suitable for improving the quality of mural face images. The implementation utilized the pretrained stable-diffusion-x4-upscaler model from the Hugging Face Diffusers library, with all computations conducted on an NVIDIA 4070 Ti GPU in mixed-precision (float16) mode to accelerate processing. To maximize the effectiveness of the upscaling process, a tiered super-resolution workflow was designed based on the original resolution of the images:

- For images with resolutions below 256×256 pixels, a two-stage upscaling process was applied. The first stage upscaled the images to approximately 256×256 pixels, followed by a second stage to reach 512×512 pixels, with intermediate steps using Lanczos resizing to mitigate artifacts and preserve finer details.
- Images with resolutions between 256×256 and 512×512 pixels were processed in a single upscaling stage to refine details while maintaining original proportions.
- Images exceeding 512×512 pixels were skipped for upscaling but carefully down sampled using Lanczos filters to standardize the dataset without compromising quality.

The upscaling process used a consistent text prompt ("A high-quality detailed image") to guide the Stable Diffusion model toward producing visually consistent and artifact-free results. Intermediate GPU memory was cleared during batch processing to optimize performance and avoid memory overflow. By meticulously adjusting the upscaling parameters

at each stage, we successfully preserved intricate details in the mural faces and significantly improved visual clarity. As illustrated in Figure 4, pixelation and jagged defects, particularly in the eye regions, were substantially reduced, resulting in a noticeably enhanced visual output.

Table 2. Similarity metric

Resolution	PSNR	SSIM	LPIPS
~512	32.2113	0.7741	0.1529
512~	32.9783	0.8626	0.0576

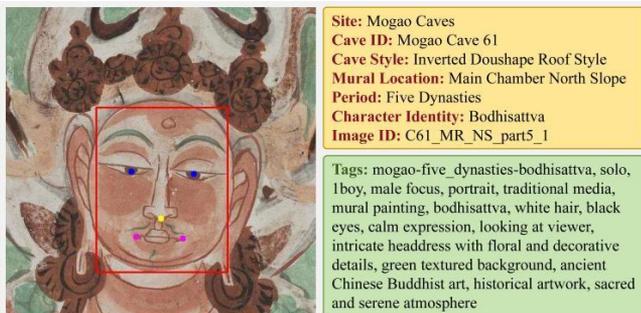
Table 2 demonstrates that, regardless of whether the images were of high or low resolution, the processed outputs exhibit a high degree of similarity to the originals, with PSNR values exceeding 32 dB and SSIM consistently reaching high levels. Additionally, the LPIPS scores provide further confirmation of the preservation quality, with values of 0.1529 for images below 512×512 resolution and 0.0576 for images above 512×512 resolution. These results confirm that the mural information was preserved with minimal distortion or loss during processing. Furthermore, as shown in Table 3, additional image quality metrics reveal notable improvements, particularly in the edge sharpness index (ESI), which increased by 1.908 times for low-resolution images after processing.

Table 3. Similarity metric

Resolution	Image	ESI↑	SNR↑	ED↑	Entropy↑	HFC↓
~512	low-res	1.908	10.9939	0.2908	6.7492	0.4506
	high-res		11.0105	0.3344	6.7613	0.4462
Resolution	Image	ESI↑	SNR↑	ED↑	Entropy↑	HFC↑
	512~	low-res	11.5143	0.3408	6.611	0.4377
	high-res	11.5343	0.3816	6.6344	0.4316	

3.1.4 Attribute labeling and text annotation

The final step in constructing the Dunhuang_Faces Dataset was to enrich the subsets with detailed annotations. For the Dunhuang_Faces-3k subset, additional attribute labels were included, such as facial identity, orientation, dynasty, and cave ID. These annotations serve as valuable metadata, enabling research tasks such as facial attribute recognition, historical categorization, and contextual analysis.



Note: Each image includes annotated facial landmarks, attribute labels (cave ID, style, location, period, and character identity), and descriptive tags capturing artistic, cultural, and contextual details. These annotations provide rich metadata for fine-grained analysis and research applications.

Figure 5. Example of data in the Dunhuang_Faces-500 subset

As shown in Figure 5, for the Dunhuang-500 subset, the

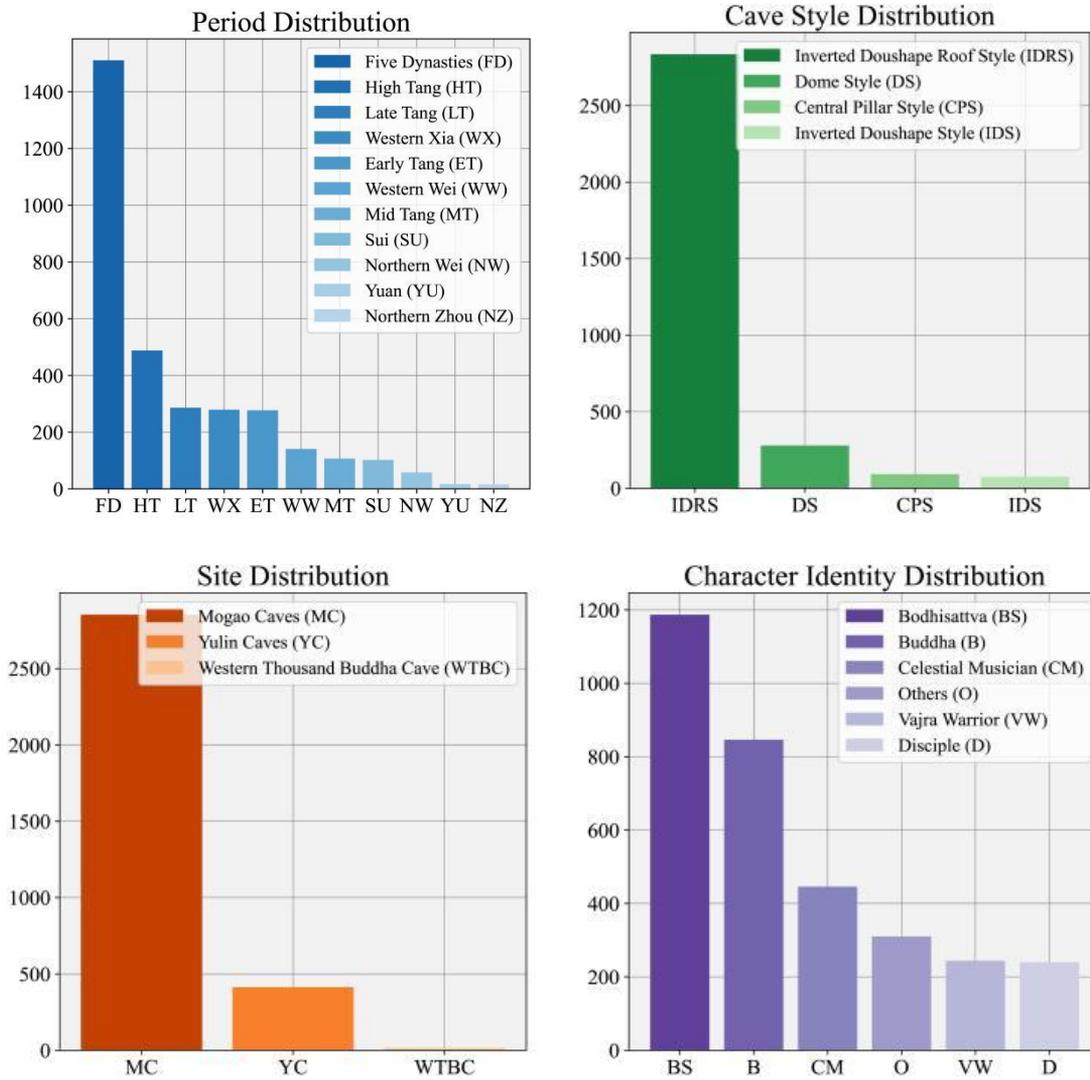
Metrics such as signal-to-noise ratio (SNR), edge density (ED), and entropy also demonstrated moderate enhancements, reflecting improvements in overall image quality, the clarity of mural outlines, and the amount of preserved information. However, a slight reduction in edge frequency content (EFC) was observed for low-resolution images, likely due to the noise reduction or smoothing effects introduced by the super-resolution process. Despite this, the high-frequency information remained largely unaffected. By leveraging the Stable Diffusion-based super-resolution model, we achieved substantial improvements in both visual quality and image fidelity, ensuring that facial details were preserved without significant alterations.

After processing, the images were standardized to a resolution of 512×512 pixels, forming the Dunhuang_Faces-3k and Dunhuang_Faces-500 subsets. This standardization not only ensured consistency across the dataset but also enhanced its compatibility with machine learning pipelines, where uniform input dimensions are essential. By leveraging the stable-diffusion-x4-upscaler within a carefully designed enhancement workflow, we achieved a significant improvement in the quality of the face images while preserving their historical and artistic integrity. This approach ensured that the dataset retained its value as a resource for both cultural heritage research and computational applications.

dataset was further enriched with detailed textual descriptions for each face image. These descriptions provide an in-depth analysis of official features and contextual elements, including details about headwear, facial expressions, and the likely social or religious identities depicted in the murals. Designed to capture the intricate cultural and historical significance of each face, these annotations offer valuable insights into the artistic styles and societal roles represented in the imagery. By integrating these detailed textual descriptions with attribute labels such as identity, orientation, and dynasty, this subset enables more fine-grained studies of Dunhuang mural art. This comprehensive layer of annotation bridges the gap between cultural heritage research and computer vision, supporting advanced tasks such as semantic analysis, historical reconstruction, and multimodal learning.

3.2 Data statistics and analysis

The Dunhuang_Faces Dataset is organized into three subsets of varying sizes and levels of curation, each designed to support diverse research needs. The largest subset, Dunhuang_Faces-20k, comprises 19,126 face images, representing all digitized mural faces from the Dunhuang grottoes. Each image is annotated with bounding box information stored in corresponding XML files, making this subset the most extensive collection for large-scale studies and serving as the foundation for the other two subsets.



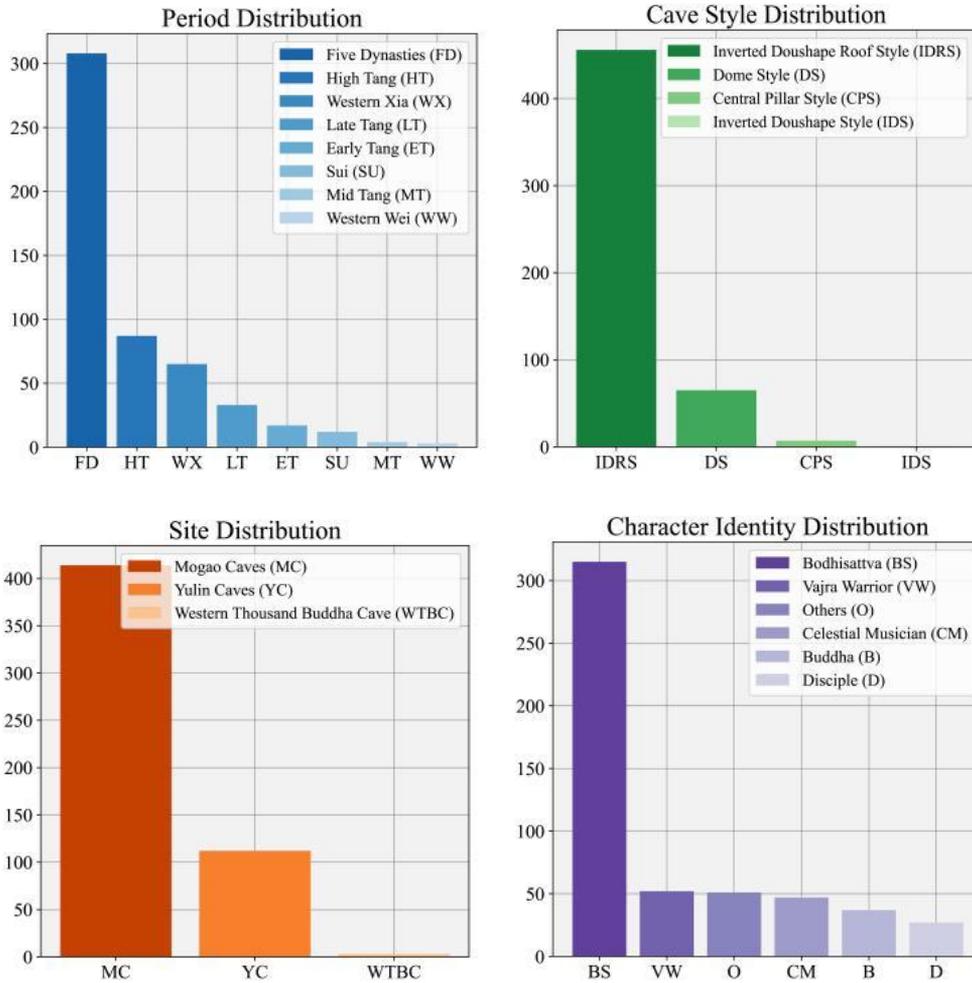
Note: Period includes Five Dynasties (FD), High Tang (HT), Early Tang (ET), Mid Tang (MT), Late Tang (LT), Western Wei (WW), Northern Wei (NW), Northern Zhou (NZ), Sui (SU), Yuan (YU), and Western Xia (WX). Cave style includes Inverted Doushape Roof Style (IDRS), Inverted Doushape Style (IDS), Central Pillar Style (CPS), and Dome Style (DS). Site includes Mogao Caves (MC), Yulin Caves (YC), and Western Thousand Buddha Cave (WTBC). Character identity includes Bodhisattvas (BS), Buddhas (B), Celestial Musicians (CM), Vajra Warriors (VW), Disciples (D), Others (O), and Donors (DN).

Figure 6. Attribute distributions in the 3K subset

As shown in Figure 6, the Dunhuang_Faces-3k subset, consisting of 3,282 images, was curated using the MTCNN and RetinaFace models with a confidence threshold of 0.9, followed by manual corrections to ensure accuracy. This subset includes detailed multi-level attribute annotations stored in JSON files, capturing aspects such as historical periods, cave architectural styles, mural sites, and character identities. Unlike single-class categories, these attributes are hierarchically structured, reflecting the rich cultural and artistic dimensions of the Dunhuang murals. The images span a wide historical timeline, with distributions across periods such as the Northern Wei (58 images), Western Wei (141 images), Northern Zhou (16 images), Sui (102 images), Early Tang (277 images), High Tang (488 images), Mid Tang (107 images), Late Tang (286 images), Five Dynasties (1511 images), Western Xia (279 images), and Yuan (17 images), illustrating the evolution of mural art over centuries. Architectural styles represented include Inverted Doushape Roof Style (2,836 images), Inverted Doushape Style (76 images), Central Pillar Style (92 images), and Dome Style

(279 images), showcasing the diversity of artistic expressions across Dunhuang mural sites such as the Mogao Caves (2,853 images), Western Thousand Buddha Cave (16 images), and Yulin Caves (413 images). Character identities include Buddha (846 images), Bodhisattvas (1,187 images), Celestial Musicians (446 images), Disciples (240 images), and Vajra Warriors (244 images), providing a rich resource for studying intricate iconography.

As shown in Figure 7, the Dunhuang_Faces-500 subset contains 549 high-quality images, manually curated from the 3k subset, and includes additional tag annotations stored in TXT files. Tailored for generative model research and other tasks requiring high-quality data, this subset provides an invaluable resource for fine-grained analysis. Together, these three subsets, with their comprehensive annotations, hierarchical structures, and diverse distributions, form a robust foundation for exploring the historical, artistic, and cultural significance of Dunhuang murals. Additionally, they support advanced computer vision research, generative modeling, and cultural heritage preservation efforts.



Note: It is showing similar patterns to the Dunhuang_Faces-3k subset. The subset is dominated by the Five Dynasties period, the Inverted Doushape Roof Style, Mogao Caves as the primary site, and depictions of Bodhisattvas and Buddhas. These distributions reflect the consistent focus on key cultural and artistic elements across both subsets.

Figure 7. Attribute distributions in the 500 subsets

4. EXPERIMENTS

To illustrate the versatility and applicability of our dataset, we conducted two distinct experiments. The first experiment evaluates its use in the traditional domain of cultural heritage, emphasizing tasks such as historical analysis and preservation. This experiment utilized the Dunhuang_Faces-20k subset, specifically using the XML-format bounding box annotations for face detection and localization tasks. The second experiment explores its potential to advance research in machine learning, focusing on applications like generative modeling and fine-grained facial attribute recognition. This experiment used the Dunhuang_Faces-500 subset, leveraging the attribute annotations in JSON format and descriptive tag files. Together, these experiments demonstrate the dataset’s value in bridging the interdisciplinary gap between cultural heritage studies and machine learning research.

4.1 Analysis of mural composition

This experiment demonstrates the practical application of the dataset in visualizing the typical compositional styles of Dunhuang Buddhist cave murals. Using data from the Dunhuang_Faces-20k subset, we analyzed the compositional

patterns of the murals to gain insights into their artistic and structural characteristics. The mural images and corresponding facial bounding box annotations were accessed seamlessly through the associated XML files, which streamlined the workflow for compositional analysis and ensured high accuracy. The bounding boxes are strictly aligned with the edges of the faces and headwear, making them a reliable representation of the figures’ positions within the murals. The size of the bounding boxes reflects the scale and posture of the figures in the mural, enabling us to investigate compositional aspects such as symmetry and spatial relationships. By providing pre-annotated regions, the Dunhuang_Faces Dataset significantly simplifies data collection and preprocessing, allowing researchers to efficiently extract and align features for consistent and precise analysis.

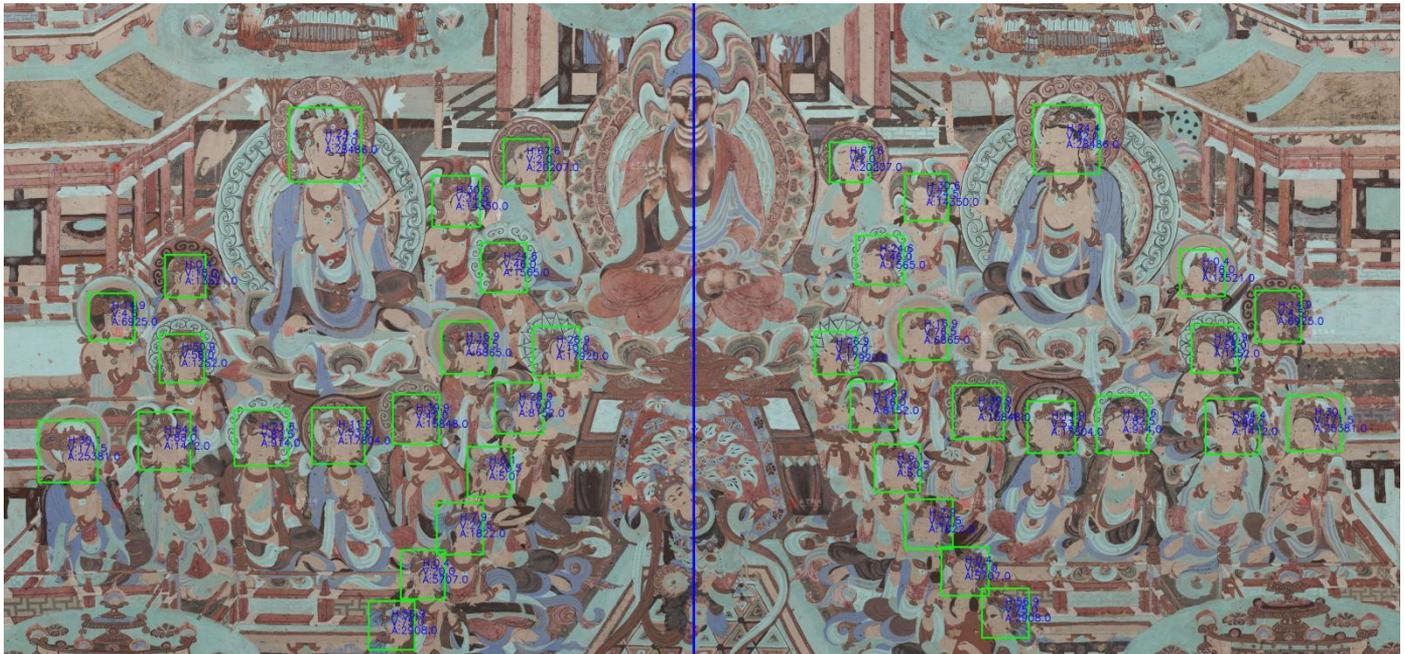
The symmetry axis c was determined by extracting the x -coordinate ranges (x_{min} and x_{max}) of all bounding boxes, computing the center x -coordinate for each box, and then averaging these center points:

$$c = \frac{1}{N} \sum_{i=1}^N \frac{x_{min_i} + x_{max_i}}{2}$$

We defined a tolerance range of 200 pixels to pair the bounding boxes on both sides of the symmetry axis. Bounding boxes with coordinate differences within 200 pixels were identified as symmetric faces. As shown in Figure 8, we calculated the horizontal error, vertical error, and area error for the symmetric bounding boxes. We calculated and analyzed the symmetry error metrics as follows: a horizontal error with a mean of 27.12 pixels and a maximum of 67.60 pixels, a vertical error with a mean of 45.66 pixels and a maximum of 171.50 pixels, and an area error with a mean of 10, 054.95 pixels² and a maximum of 28, 486.00 pixels². Given the original image resolution of 8447×4946 pixels, the horizontal and vertical errors are relatively small, while the area error is comparatively larger. These results demonstrate that the

figures in the analyzed mural exhibit a rigorously symmetrical arrangement, with low horizontal and vertical errors substantiating the structural precision of the composition. The larger area errors, on the other hand, highlight notable variations in individual figure attributes, such as expressions, postures, and proportions. This interplay of strict structural symmetry and artistic diversity is a defining characteristic of Mogao Cave murals.

Through digital analysis, we demonstrated the strict compositional structure and diverse individual representations present in the Dunhuang murals. This experiment under-scores the potential of the Dunhuang_Faces Dataset to facilitate quantitative studies of cultural heritage artifacts and enhance our understanding of traditional artistic practices.



Note: Green bounding boxes highlight detected faces with corresponding symmetry error metrics: horizontal error (H), vertical error (V), and area error (A) for symmetric pairs. The visualization reveals a strict symmetry in the mural composition, unified with diverse individual character representations, reflecting the balance between structural precision and artistic variability.

Figure 8. Visualization of compositional analysis

4.2 Five dynasties bodhisattva image generation and restoration

To evaluate the applicability of the processed dataset, we conducted an experiment by fine-tuning a Low-Rank Adaptation (LoRA) model to generate and restore Five Dynasties Bodhisattva mural imagery. This experiment highlighted the dataset’s utility for tasks such as style-specific generation, line art coloring, and mural restoration. For this experiment, we used the Dunhuang_Faces-500 subset and filtered 40 Five Dynasties Bodhisattva images based on attribute labels in the JSON files. Additionally, we utilized the descriptive tag files in TXT format to further refine the selection of specific styles and characteristics. These images represented various Bodhisattva depictions, including frontal, left-profile, and right-profile views, all standardized to a resolution of 512×512 pixels to match the requirements of the LoRA training pipeline. The LoRA model was fine-tuned using the v1-5-pruned Stable Diffusion architecture, focusing on generating Five Dynasties Bodhisattva imagery and restoring mural features. The model was optimized for tasks such as style-specific generation, line art coloring, and mural restoration, ensuring high stylistic fidelity and preserving the

artistic characteristics of the original imagery.

The fine-tuned LoRA model demonstrated promising results across the evaluated tasks. As shown in Figure 9, in the style generation task, the model accurately captured key characteristics of Five Dynasties Bodhisattva depictions, including facial features, head ornaments, and earrings, achieving a visual style closely resembling the historical imagery. While some outputs exhibited a tendency toward high color saturation, the overall stylistic fidelity was preserved. Quantitative evaluation using the Fréchet Inception Distance (FID) yielded a score of 48.78. Considering the limited training data of only 40 images due to mural preservation constraints, this score is within an acceptable range, demonstrating the model’s high fidelity in replicating Five Dynasties Bodhisattva imagery.

In the mural restoration task, as shown in Figure 9, we selected an image with damage to the eye and head ornaments for a localized restoration experiment. The model successfully reconstructed the eye details and enhanced the intricate features of the head ornaments. Evaluation metrics, including PSNR and SSIM, averaged 44.06 dB and 0.9804, respectively, indicating that the restoration process effectively preserved the original information and quality of the image. These results

demonstrate that training a LoRA model on images from the dataset and applying it to repair similarly styled damaged images is effective, offering valuable insights and research resources for real-world mural restoration tasks. In addition to restoration, as shown in Figure 9, the trained LoRA model was also applied to tasks such as coloring line art images and generating stylized outputs using ControlNet-based methods or img-to-img workflows, achieving stylistic generation consistent with the Five Dynasties Bodhisattva imagery. By providing line drawings as input, the model accurately filled in vibrant, historically inspired colors while maintaining the characteristic features of Five Dynasties Bodhisattva depictions, including facial expressions, ornaments, and shading. This experiment highlights the model’s ability to extend beyond restoration tasks, demonstrating its versatility in generating stylistically faithful, fully colored images from minimal input.



Note: The tasks showcase the dataset’s potential for generating stylistically consistent images, restoring damaged mural details, and coloring line art with Dunhuang-specific styles.

Figure 9. Examples of text-to-image generation (top row), localized damage restoration (middle row), and stylized generation (bottom row) using the *Dunhuang_Faces-500* dataset and a fine-tuned LoRA model

5. CONCLUSION

This study introduced *Dunhuang_Faces*, a novel dataset designed to address the scarcity of high-quality data in mural art research, with a particular focus on the preservation and analysis of Dunhuang murals. Considering the severe degradation and fragility of many mural images, we employed a state-of-the-art super-resolution model to restore and enhance existing data, resulting in a specialized dataset for facial imagery within Dunhuang murals. This approach maximized the utility of the limited available data, filling a critical gap in art image datasets, particularly for representations of Dunhuang figures. The *Dunhuang_Faces* dataset comprises three subsets: *Dunhuang_Faces-20k*, *Dunhuang_Faces-3k*, and *Dunhuang_Faces-500*. Each subset was meticulously curated and standardized to ensure usability across a wide range of applications. These datasets provide a robust foundation for both traditional cultural heritage studies and contemporary machine learning tasks.

By bridging the gap between traditional art preservation and modern computational techniques, *Dunhuang_Faces* contributes significantly to the growing body of resources available for cultural heritage research. The dataset supports advancements in machine learning applications while offering valuable insights into the conservation and understanding of historical mural art. In the domain of cultural heritage, the dataset enables tasks such as mural restoration, historical analysis, and digital archiving, facilitating the preservation of these invaluable artifacts. In the machine learning field, *Dunhuang_Faces* serves as a versatile resource for face detection, style transfer, and generative modeling, supporting the development of novel algorithms tailored to cultural heritage applications. By integrating detailed annotations, descriptive tags, and structured data, this dataset bridges the needs of traditional cultural research and modern computational tasks, fostering interdisciplinary collaboration. We hope that this dataset will inspire further exploration and foster new methodologies for preserving and analyzing invaluable cultural artifacts such as the Dunhuang murals. In the future, we plan to expand the dataset by adding annotations for additional facial regions, such as headwear and cheeks, and introducing more attributes that reflect the artistic and cultural significance of the murals. These enhancements aim to support a broader range of research tasks and deepen the dataset’s value for interdisciplinary applications.

REFERENCES

- [1] Santos, P., Ritz, M., Fuhrmann, C., et al. (2017). Acceleration of 3D mass digitization processes: Recent advances and challenges. *Mixed Reality and Gamification for Cultural Heritage*, pp. 99-128. https://doi.org/10.1007/978-3-319-49607-8_4
- [2] Aicardi, I., Chiabrand, F., Lingua, A.M., Noardo, F. (2018). Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach. *Journal of Cultural Heritage*, 32: 257-266. <https://doi.org/10.1016/j.culher.2017.11.006>
- [3] Buragohain, D., Meng, Y., Deng, C., Li, Q., Chaudhary, S. (2024). Digitalizing cultural heritage through metaverse applications: Challenges, opportunities, and strategies. *Heritage Science*, 12(1): 295.

- [4] Ivanova, K., Dobрева, M., Stanchev, P., Totkov, G. (2012). Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation. Plovdiv University Publishing House, Plovdiv, Bulgaria.
- [5] Jin, P., Liu, Y. (2022). Fluid space: Digitisation of cultural heritage and its media dissemination. *Telematics and Informatics Reports*, 8: 100022. <https://doi.org/10.1016/j.teler.2022.100022>
- [6] Gregor, R., Sipiran, I., Papaioannou, G., Schreck, T., Andreadis, A., Mavridis, P. (2014). Towards automated 3D reconstruction of defective cultural heritage objects. In *Eurographics Workshop on Graphics and Cultural Heritage*, pp. 135-144. <https://doi.org/10.2312/gch.20141311>
- [7] Windhager, F., Federico, P., Schreder, G., Glinka, K., Dörk, M., Miksch, S., Mayr, E. (2018). Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE Transactions on Visualization and Computer Graphics*, 25(6): 2311-2330. <https://doi.org/10.1109/TVCG.2018.2830759>
- [8] Galantucci, R.A., Fatiguso, F. (2019). Advanced damage detection techniques in historical buildings using digital photogrammetry and 3D surface analysis. *Journal of Cultural Heritage*, 36: 51-62. <https://doi.org/10.1016/j.culher.2018.09.014>
- [9] Meinecke, C. (2022). Labeling of cultural heritage collections on the intersection of visual analytics and digital humanities. In *2022 IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*, pp. 19-24. <https://doi.org/10.1109/VIS4DH57440.2022.00009>
- [10] Towarek, A., Halicz, L., Matwin, S., Wagner, B. (2024). Machine learning in analytical chemistry for cultural heritage: A comprehensive review. *Journal of Cultural Heritage*, 70: 64-70. <https://doi.org/10.1016/j.culher.2024.08.014>
- [11] Roued-Cunliffe, H. (2020). *Open Heritage Data: An Introduction to Research, Publishing and Programming with Open Data in The Heritage Sector*. Facet Publishing.
- [12] Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S. (2020). Machine learning for cultural heritage: A survey. *Pattern Recognition Letters*, 133: 102-108. <https://doi.org/10.1016/j.patrec.2020.02.017>
- [13] Wang, M., Fu, W., He, X., Hao, S., Wu, X. (2020). A survey on large-scale machine learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(6): 2574-2594. <https://doi.org/10.1109/TKDE.2020.3015777>
- [14] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on Computer Vision and Pattern Recognition*, pp. 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [15] Lin, T.Y., Maire, M., Belongie, S., et al. (2014). Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland*, pp. 740-755.
- [16] Mao, X., Chen, Y., Zhu, Y., Chen, D., Su, H., Zhang, R., Xue, H. (2023). Coco-o: A benchmark for object detectors under natural distribution shifts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6339-6350.
- [17] Deng, X., Yu, Q., Wang, P., Shen, X., Chen, L.C. (2024). Coconut: Modernizing coco segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21863-21873.
- [18] Kuznetsova, A., Rom, H., Alldrin, N., et al. (2020). The open images dataset V4: Unified image classification, object detection, and visual relationship detection at scale. *International journal of computer vision*, 128(7): 1956-1981. <https://doi.org/10.1007/s11263-020-01316-z>
- [19] Cordts, M., Omran, M., Ramos, S., et al. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213-3223.
- [20] Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A. (2017). Scene parsing through ADE20k dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 633-641.
- [21] Ridnik, T., Ben-Baruch, E., Noy, A., Zelnik-Manor, L. (2021). Imagenet-21k pretraining for the masses. *arXiv preprint arXiv:2104.10972*. <https://doi.org/10.48550/arXiv.2104.10972>
- [22] Sun, C., Shrivastava, A., Singh, S., Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 843-852.
- [23] Ruan, B.K., Shuai, H.H., Cheng, W.H. (2022). Vision transformers: State of the art and research challenges. *arXiv preprint arXiv:2207.03041*. <https://doi.org/10.48550/arXiv.2207.03041>
- [24] Gupta, A., Dollar, P., Girshick, R. (2019). Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5356-5364.
- [25] Krishna, R., Zhu, Y., Groth, O., et al. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123: 32-73. <https://doi.org/10.1007/s11263-016-0981-7>
- [26] Schuhmann, C., Beaumont, R., Vencu, R., et al. (2022). Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35: 25278-25294.
- [27] Carreira, J., Noland, E., Hillier, C., Zisserman, A. (2019). A short note on the kinetics-700 human action dataset. *arXiv preprint arXiv:1907.06987*. <https://doi.org/10.48550/arXiv.1907.06987>
- [28] Alkemade, H., Claeysens, S., Colavizza, G., et al. (2023). Datasheets for digital cultural heritage datasets. *Journal of Open Humanities Data*, 9(17): 1-11. <https://doi.org/10.5334/johd.124>
- [29] Rakovitsky, A., Knott, J. (2020). *WikiArt Analysis Using Facial Emotion Analysis*.
- [30] Dijkshoorn, C., Jongma, L., Aroyo, L., Van Ossenbruggen, J., Schreiber, G., Ter Weele, W., Wielemaker, J. (2018). The Rijksmuseum collection as linked data. *Semantic Web*, 9(2): 221-230. <https://doi.org/10.3233/SW-170257>
- [31] Reshetnikov, A., Marinescu, M.C., Lopez, J.M. (2022). DEArt: Dataset of European art. In *European Conference on Computer Vision*, pp. 218-233. https://doi.org/10.1007/978-3-031-25056-9_15
- [32] Striova, J., Camaiti, M., Castellucci, E.M., Sansonetti, A. (2011). Chemical, morphological and chromatic behavior of mural paintings under Er: YAG laser

- irradiation. *Applied Physics A*, 104: 649-660. <https://doi.org/10.1007/s00339-011-6303-6>
- [33] Ogura, D., Nakata, Y., Hokoi, S., Takabayashi, H., Okada, K., Su, B., Xue, P. (2019). Influence of light environment on deterioration of mural paintings in Mogao Cave 285, Dunhuang. In *AIP Conference Proceedings*, 2170(1): 020013. <https://doi.org/10.1063/1.5132732>
- [34] Durmus, D. (2021). Characterizing color quality, damage to artwork, and light intensity of multi-primary LEDs for museums. *Heritage*, 4(1): 188-197. <https://doi.org/10.3390/heritage4010011>
- [35] Yu, T., Lin, C., Zhang, S., et al. (2022). Artificial intelligence for Dunhuang cultural heritage protection: The project and the dataset. *International Journal of Computer Vision*, 130(11): 2646-2673. <https://doi.org/10.1007/s11263-022-01665-x>
- [36] Krizhevsky, A., Hinton, G. (2009). Learning multiple layers of features from tiny images. University of Toronto.
- [37] Garcia, N., Vogiatzis, G. (2018). How to read paintings: Semantic art understanding with multi-modal retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.
- [38] Carneiro, G., Da Silva, N.P., Del Bue, A., Costeira, J.P. (2012). Artistic image classification: An analysis on the PrintArt database. In *Computer Vision—ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, pp. 143-157. https://doi.org/10.1007/978-3-642-33765-9_11
- [39] Chiaro, R.D., Bagdanov, A.D., Bimbo, A.D. (2021). NoisyArt: Exploiting the noisy web for zero-shot classification and artwork instance recognition. In *Data Analytics for Cultural Heritage*, pp. 1-24. https://doi.org/10.1007/978-3-030-66777-1_1
- [40] Mao, H., Cheung, M., She, J. (2017). Deepart: Learning joint representations of visual arts. In *Proceedings of the 25th ACM International Conference on Multimedia*, pp. 1183-1191. <https://doi.org/10.1145/3123266.3123405>
- [41] Liao, P., Li, X., Liu, X., Keutzer, K. (2022). The artbench dataset: Benchmarking generative models with artworks. *arXiv preprint arXiv:2206.11404*. <https://doi.org/10.48550/arXiv.2206.11404>
- [42] Hosse, K., Schilcher, M. (2003). Temporal GIS for analysis and visualization of cultural heritage. In *Proceedings of CIPA XIX International Symposium*, Commission V, WG5, Antalya.
- [43] Elfadaly, A., Attia, W., Qelichi, M.M., Murgante, B., Lasaponara, R. (2018). Management of cultural heritage sites using remote sensing indices and spatial analysis techniques. *Surveys in Geophysics*, 39: 1347-1377. <https://doi.org/10.1007/s10712-018-9489-8>
- [44] Agustina, I.H. (2021). GIS approach to spatial analysis of heritage settlement: Case study of Magersari Kasepuhan Palace, Indonesia. *Journal of Engineering Science and Technology*, 16(2): 1614-1629.
- [45] Hadjimitsis, D., Agapiou, A., Alexakis, D., Sarris, A. (2013). Exploring natural and anthropogenic risk for cultural heritage in Cyprus using remote sensing and GIS. *International Journal of Digital Earth*, 6(2): 115-142. <https://doi.org/10.1080/17538947.2011.602119>
- [46] Agapiou, A., Lysandrou, V., Alexakis, D.D., Themistocleous, K., Cuca, B., Argyriou, A., Sarris, A., Hadjimitsis, D.G. (2015). Cultural heritage management and monitoring using remote sensing data and GIS: The case study of Paphos area, Cyprus. *Computers, Environment and Urban Systems*, 54: 230-239. <https://doi.org/10.1016/j.compenvurbsys.2015.09.003>
- [47] Koller, D., Frischer, B., Humphreys, G. (2010). Research challenges for digital archives of 3D cultural heritage models. *Journal on Computing and Cultural Heritage (JOCCH)*, 2(3): 1-17. <https://doi.org/10.1145/1658346.165834>
- [48] Wijesundara, C., Sugimoto, S. (2018). Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space. *Library and Information Science Research E-Journal*, 28(2): 58-80. <https://doi.org/10.32655/LIBRES.2018.2.2>
- [49] Styliani, S., Fotis, L., Kostas, K., Petros, P. (2009). Virtual museums, a survey and some issues for consideration. *Journal of Cultural Heritage*, 10(4): 520-528. <https://doi.org/10.1016/j.culher.2009.03.003>
- [50] Besoain, F., González-Ortega, J., Gallardo, I. (2022). An evaluation of the effects of a virtual museum on users' attitudes towards cultural heritage. *Applied Sciences*, 12(3): 1341. <https://doi.org/10.3390/app12031341>
- [51] Georgopoulos, A., Oikonomou, C., Adamopoulos, E., Stathopoulou, E.K. (2016). Evaluating unmanned aerial platforms for cultural heritage large scale mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41: 355-362. <https://doi.org/10.5194/isprs-archives-XLI-B5-355-2016>
- [52] Llabani, A., Abazaj, F. (2024). 3D documentation of cultural heritage using terrestrial laser scanning. *Journal of Applied Engineering Science*, 22(2): 267-271. <https://doi.org/10.5937/jaes0-50414>
- [53] Luther, W., Baloian, N., Biella, D., Sacher, D. (2023). Digital twins and enabling technologies in museums and cultural heritage: An overview. *Sensors*, 23(3): 1583. <https://doi.org/10.3390/s23031583>
- [54] Dang, X., Liu, W., Hong, Q., Wang, Y., Chen, X. (2023). Digital twin applications on cultural world heritage sites in China: A state-of-the-art overview. *Journal of Cultural Heritage*, 64: 228-243. <https://doi.org/10.1016/j.culher.2023.10.005>
- [55] Paolini, P., Di Blas, N. (2014). Storytelling for cultural heritage. *Innovative Technologies in Urban Mapping: Built Space and Mental Space*, pp. 33-45. https://doi.org/10.1007/978-3-319-03798-1_4
- [56] Brusaporci, S., Graziosi, F., Franchi, F., Maiezza, P., Tata, A. (2021). Mixed reality experiences for the historical storytelling of cultural heritage. In *From Building Information Modelling to Mixed Reality*, pp. 33-46. https://doi.org/10.1007/978-3-030-49278-6_3
- [57] Valtolina, S., Franzoni, S., Mazzoleni, P., Bertino, E. (2005). Dissemination of cultural heritage content through virtual reality and multimedia techniques: A case study. In *11th International Multimedia Modelling Conference*, pp. 214-221. <https://doi.org/10.1109/MMMC.2005.36>
- [58] Tang, T., Zhang, H. (2023). An interactive holographic multimedia technology and its application in the preservation and dissemination of intangible cultural heritage. *International Journal of Digital Multimedia Broadcasting*, 2023(1): 6527345. <https://doi.org/10.1155/2023/6527345>

[59] Llamas, J.M., Leronés, P., Medina, R., Zalama, E., Gómez-García-Bermejo, J. (2017). Classification of architectural heritage images using deep learning techniques. *Applied Sciences*, 7(10): 992.

<https://doi.org/10.3390/app7100992>
[60] Duan, W.J. (2010). Complete Collection of Dunhuang Murals. Tianjin People's Fine Arts Publishing House, Tianjin, China.