

An Intelligent Generation Method for Kizil Rhombic Mural Patterns Based on Diffusion Models and Its Application in Digital Fashion Design

Qixing Lei¹, Jianjun Wang², Cheng Chang³

¹School of Plastic Arts, Shanghai Art & Design Academy, 851 Jiahang Rd, Jiading District, Shanghai 201808, China

²Institute of Cave Temple Archaeology and Art, Qiuci Research Institute of Xinjiang Uygur Autonomous Region, Kuqa City, 842000, China

³School of General Education, Shanghai Technical Institute of Electronics & Information, Shanghai 201411, China

Corresponding author: Cheng Chang (e-mail: 15921608904@163.com)

ABSTRACT Structured geometric patterns play an important role not only in cultural heritage preservation and digital design but also in engineering applications where periodic layouts inspire functional surface architectures and electromagnetic wave manipulation. The Kizil rhombic murals, characterized by distinctive lattice configurations and rich visual semantics, present significant challenges for intelligent regeneration because conventional generative methods cannot simultaneously preserve stylistic authenticity and geometric regularity. This study develops a conditional diffusion-based framework that combines latent diffusion modeling with ControlNet-guided structural constraints to generate high-fidelity Kizil rhombic patterns. A dedicated high-resolution dataset is established, and the model is optimized to jointly capture chromatic characteristics, artistic motifs, and rigid rhombic-grid organization through Canny edge conditioning. Quantitative evaluation using Fréchet Inception Distance, Inception Score, and Structural Adherence Score demonstrates that the proposed approach achieves superior visual quality and structural consistency compared with existing generative methods. The generated patterns are further integrated into a digital fashion design workflow for virtual garment creation, confirming their practical applicability in creative industries. Beyond cultural heritage regeneration, the proposed framework provides a data-driven strategy for generating structured periodic patterns with potential reference value for geometry-oriented design problems in electromagnetic surfaces, antenna-inspired pattern engineering, and intelligent visual manufacturing.

INDEX TERMS diffusion model, kizil mural patterns, digital fashion design, cultural heritage, pattern generation, electromagnetic surface design

I. INTRODUCTION

THE Kizil Caves, a significant Buddhist grotto complex located on the ancient Silk Road, house a treasure trove of murals dating from the 3rd to the 8th centuries. Among these, the rhombus-grid patterns are particularly distinctive, featuring a repeating diamond-shaped lattice that frames various narrative scenes, decorative motifs, and celestial figures [1]. These patterns are not merely decorative; they serve as a unique artistic language, embodying the cultural and religious syncretism of the region. They possess a delicate balance of geometric rigidity and artistic fluidity, characterized by specific color palettes (lapis lazuli blue, mineral green, earth red), calligraphic brushwork, and a wealth of symbolic content [2,3]. In the modern era, the digitization and innovative application of such invaluable cultural heritage have become topics of paramount importance.

However, the manual recreation of these patterns is a labor-intensive and highly skilled process, while conventional digital tools often fail to capture their nuanced aesthetic and stylistic essence, leading to sterile and decontextualized reproductions [4,5].

The rise of artificial intelligence, particularly deep learning-based generative models, has opened new avenues for artistic creation and cultural preservation [6]. Early generative adversarial networks (GANs) demonstrated the ability to produce novel images, but they often suffered from training instability, mode collapse, and a limited ability to control the structure of the output, making them ill-suited for generating patterns that must adhere to a specific geometric framework [6,7]. More recently, diffusion models have emerged as the state-of-the-art in high-fidelity image generation. These models operate by progressively adding noise to an image and then learning to reverse the process,

allowing them to generate highly detailed and coherent images from a random noise vector. Their inherent stability and the quality of their outputs have made them powerful tools for a wide range of creative applications [8,9]. However, generating images that must adhere to precise, non-local geometric structures—such as the lattice framework of the Kizil murals—poses a distinct challenge that standard generative models are not inherently equipped to handle.

Prior work on generating structured patterns, such as textiles or architectural details, has explored several avenues. For instance, some approaches have utilized procedural methods or specialized GAN architectures for texture synthesis, but these often struggle to balance global structural coherence with high-fidelity, diverse details and can be difficult to control [10]. Other methods have employed conditional GANs (e.g., pix2pix) to translate a structural map into a stylized output, yet these often require large paired datasets and may suffer from a lack of stylistic diversity in the generated results [11]. While these methods show promise, they often fall short when faced with the unique dual requirement of the Kizil patterns: maintaining the rigid, repeating geometry of the rhombus grid while simultaneously rendering the fluid, calligraphic brushwork and rich symbolic motifs contained within each cell.

Despite the power of standard diffusion models, applying them directly to the generation of Kizil rhombus-grid patterns presents a significant challenge. A model trained on these images might learn the stylistic elements—the colors, textures, and motifs—but it would struggle to consistently reproduce the crucial rhombus-grid structure [12,13]. The generated images would likely be an amorphous collage of Kizil-style elements rather than a coherent, structured pattern suitable for applications like textile design. This highlights a critical research gap: the need for a specialized generative framework that can simultaneously learn the intricate, culturally specific style of the Kizil murals and enforce a predefined geometric structure.

This paper aims to fill this gap by proposing a novel method that combines fine-tuning a latent diffusion model with a conditional control mechanism. Our contribution is twofold: first, we develop and validate a robust methodology for generating high-fidelity, structurally coherent Kizil rhombus-grid patterns; second, we demonstrate the practical value of this methodology by integrating the generated patterns into a complete digital fashion design workflow, thereby creating a tangible link between ancient cultural heritage and contemporary creative industries. This work not only offers a new technical solution for digital cultural preservation but also explores a new paradigm for AI-powered design innovation.

II. METHODOLOGY

The proposed methodology is designed as a comprehensive framework, encompassing data acquisition, model development, and practical application. The core of our approach lies in a specially adapted diffusion model that prioritizes

both stylistic authenticity and structural precision.

A. DATASET CURATION AND PREPROCESSING

The foundation of our model is a high-quality, specialized dataset constructed from scratch due to the lack of existing public datasets for Kizil rhombus-grid patterns. The process began with the collection of over 5,000 high-resolution digital images of the Kizil Cave murals, primarily sourced from publicly available academic archives, including the Digital Silk Road Project and the Berlin Turfan Collection's online database. All images used were either in the public domain or available under Creative Commons licenses permitting academic research, ensuring full compliance with copyright regulations. The collected images originated from a wide range of grottoes (including Caves 17, 38, 69, and 114) to capture stylistic variations, though they were captured under varied scene and lighting conditions, which we later addressed during the color correction phase. From this initial pool, we established strict screening criteria for creating our final dataset. To be included, an image patch had to: (1) have a resolution of at least 1024×1024 pixels; (2) feature a clearly identifiable rhombus-grid structure covering at least 70% of the patch area; and (3) be free from severe, irreparable damage that obscured the core motifs. This screening process yielded approximately 2,200 candidate patches. Following this, we manually selected and cropped the final 1,500 distinct image patches based on their clarity, the integrity of the pattern, and their representation of diverse motifs and color schemes, resulting in an elimination rate of approximately 32% from the candidate pool.

To prepare the data for model training, a multi-stage preprocessing pipeline was implemented. First, perspective correction and geometric transformation were applied to ensure that the rhombus grids were aligned and not distorted. Second, color correction was carefully performed to restore the colors to their original state as closely as possible, compensating for centuries of fading and photographic inconsistencies.

Third, to provide the model with stylistically coherent examples, minor non-original damages (e.g., cracks, water stains) were carefully inpainted. This process was conducted by junior researchers under the supervision of an art history expert to ensure historical accuracy. We utilized the Patch-Match algorithm—a content-aware fill method known for preserving local textures—to perform the inpainting. The objective was not to create a flawless modern reproduction, but to remove distracting artifacts that were not part of the original artists' intent, thereby allowing the model to learn from cleaner stylistic features while preserving the authentic mural texture. Finally, to expand the dataset and prevent overfitting, we applied data augmentation techniques. These included random horizontal flipping and minor adjustments to brightness and contrast. Additionally, slight rotations within a $\pm 5^\circ$ range were applied. This specific range was determined through empirical testing; we observed that rotations exceeding 5° , after being applied and subsequently

resized to the target 512×512 resolution, introduced significant interpolation artifacts and aliasing that visibly distorted the sharp lines of the rhombus grid. Therefore, the $\pm 5^\circ$ limit was established as a trade-off to maximize data diversity while preserving the essential geometric clarity of the patterns.

The final dataset consisted of 1,500 processed images, each resized to a resolution of 512×512 pixels, which served as the ground truth for training our generative model. This dataset was then partitioned according to a standard 80/10/10 split, resulting in a training set of 1,200 images, a validation set of 150 images for monitoring performance during fine-tuning, and a test set of 150 images, which was held out and used as the reference distribution for the final quantitative evaluations (FID, IS, and SAS). While this dataset size is modest, it is effective for our chosen methodology of fine-tuning a large, pre-trained latent diffusion model.

This approach leverages the powerful, generalized features learned from a vast dataset, requiring only a smaller, specialized dataset to adapt it to the specific stylistic nuances of the Kizil mural patterns.

B. THE CONDITIONAL DIFFUSION MODEL FRAMEWORK

Our generation framework is built upon a latent diffusion model, specifically the architecture popularized by Stable Diffusion, which operates in a lower-dimensional latent space for computational efficiency. However, to address the critical challenge of structural control, we integrated the ControlNet architecture. ControlNet is a neural network structure that allows for the addition of spatial conditioning inputs to large, pre-trained diffusion models, enabling more precise control over the generated output.

The overall framework, illustrated in Figure 1, works as follows. The system takes two inputs: a text prompt (e.g., “Kizil mural rhombus pattern, lapis lazuli blue, intricate details”) and a conditioning image. The conditioning image is a simple binary mask representing the filled shapes of the desired rhombus-grid structure. To create the structural condition, we process this binary template with a Canny edge detector.

This choice is deliberate and critical, as the ControlNet v1.1 model we employ was specifically pre-trained on a massive dataset of Canny edge maps. Directly using this binary mask (which consists of solid filled regions) would create a domain mismatch with the model’s training data, as the ControlNet was trained on sparse edge maps. The Canny algorithm, with its non-maximum suppression step, ensures the creation of a sparse, single-pixel-width edge map that precisely matches the expected input format of the pre-trained ControlNet, thereby maximizing structural guidance fidelity. To ensure reproducibility, we used a consistent set of parameters for the Canny detector: a low threshold of 100, a high threshold of 200, and a Sobel aperture size of 3.

The core diffusion model consists of a pre-trained U-Net denoising network. To adapt this model to our specific task, we performed a fine-tuning process on our curated Kizil pattern dataset. During fine-tuning, the model learns the specific visual vocabulary of the Kizil murals—the color harmonies, the textures of mineral pigments, the calligraphic line work, and the characteristic motifs.

Crucially, the pre-trained ControlNet module is not frozen but is fine-tuned jointly with the main U-Net model.

This joint training strategy allows the ControlNet to adapt its general understanding of edge-based conditioning to the specific stylistic nuances and line work of the Kizil mural dataset. All layers of both the U-Net and the ControlNet were unfrozen and trained end-to-end using the same set of hyperparameters, ensuring a synergistic adaptation to the target domain. The ControlNet duplicates the weights of the U-Net’s encoding layers and is trained to accept the Canny edge map of the rhombus grid as an additional condition.

The outputs from the ControlNet’s trainable encoder blocks are fed into the main U-Net at multiple hierarchical scales. Specifically, they are added to the skip-connection features that link the main U-Net’s encoder blocks to their corresponding decoder blocks. This injection is mediated by “zero convolution” layers— 1×1 convolutional layers with both weight and bias initialized to zero. This powerful strategy ensures that at the start of training, the ControlNet adds no disruptive noise to the fine-tuned U-Net, preserving its learned style. As training progresses, these layers learn to effectively integrate the structural conditions from the ControlNet into the main generation process. This mechanism forces the U-Net to generate content that aligns with the provided structural map, ensuring that the final generated pattern strictly adheres to the rhombus-grid layout while being filled with the rich, learned style of the Kizil murals. This dual-conditioning approach—semantic guidance from the text prompt and structural guidance from ControlNet—provides a high degree of control and artistic fidelity.

C. DIGITAL FASHION DESIGN APPLICATION WORKFLOW

To demonstrate the practical utility of the generated patterns, we established a workflow to integrate them into digital fashion design using industry-standard software, CLO 3D. This process transforms the 2D generated patterns into realistic 3D virtual garments.

The workflow begins with the selection of a high-resolution (e.g., 2048×2048 pixels) pattern generated by our model. The “seamless tiling” feature is crucial for textile applications; we ensure this by post-processing the generated images to create patterns that can be repeated without visible seams.

Next, a base 3D garment model is chosen or designed within CLO 3D, such as a dress, a skirt, or a coat. The generated seamless pattern is then imported as a fabric texture. The software’s UV mapping tools are used to apply the 2D texture onto the 3D garment’s pattern pieces.

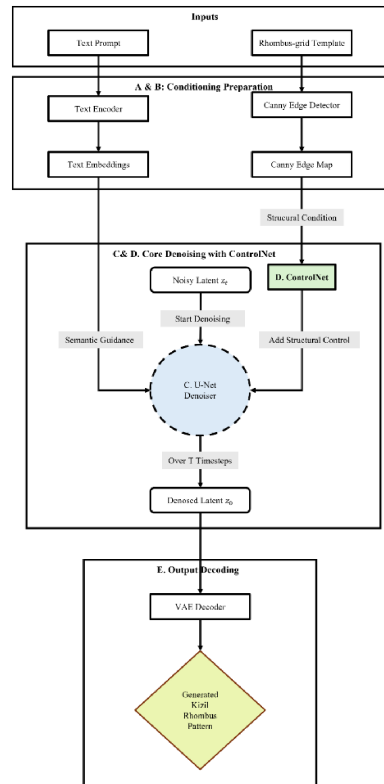


FIGURE 1. Detailed architecture of the proposed conditional latent diffusion model. The process begins with two inputs: a Text Prompt and a Rhombus-grid Template. (A) The template is converted into a sparse Canny edge map, which serves as the structural condition for ControlNet. (B) The text prompt is converted into semantic embeddings by a text encoder. (C) The core of the model is a U-Net denoiser that operates in the latent space. Over multiple timesteps, it iteratively removes noise from a random latent vector (z_t), guided by the text embeddings. (D) The ControlNet, a trainable copy of the U-Net's encoder, processes the Canny edge map. Its outputs are added to the skip connections of the main U-Net, injecting the structural guidance from the edge map into the corresponding layers of the generative process. (E) After the final denoising step, the resulting clean latent vector (z_0) is decoded back into a high-resolution image by the VAE Decoder

This stage involves critical artistic and technical decisions, including adjusting the scale, rotation, and placement of the pattern to complement the garment's silhouette and construction lines. For instance, a larger pattern scale might be used for the main body of a coat, while a smaller scale could be applied to cuffs and collars for a refined look.

To enhance realism, Physically Based Rendering (PBR) material properties are assigned to the fabric. This includes defining parameters such as roughness, metallicness, and normal maps. A subtle normal map can be derived from the pattern's texture to simulate the slight unevenness of a woven fabric or the texture of ancient mural surfaces, adding a tactile dimension to the visual appearance. Finally, the complete garment is simulated on a virtual avatar, allowing for adjustments to fit and drape. The process culminates in a high-quality rendering of the final virtual fashion piece, showcasing the Kizil-inspired pattern in a contemporary design context.

D. IMPLEMENTATION DETAILS

Model Architecture and Weights

Our framework is built upon the Stable Diffusion v1.5 pre-trained latent diffusion model. For structural control, we integrated the official ControlNet v1.1 model pre-trained for Canny edge conditioning. Both the base model and Control-

Net were implemented using the Hugging Face Diffusers library in PyTorch.

Training Hyperparameters

The following hyperparameters were applied to train both the main U-Net model and the ControlNet module, as they were fine-tuned jointly. The fine-tuning process was conducted on a single NVIDIA A100 GPU for 50 epochs with a batch size of 4. We used the AdamW optimizer with a constant learning rate of 1×10^{-5} , betas of ($\beta_1 = 0.9$, $\beta_2 = 0.999$), and a weight decay of 1×10^{-2} . To ensure the reproducibility of our results, a fixed random seed was used for all experiments, including data partitioning and model weight initialization. We did not employ an early stopping strategy; the model was trained for a fixed 50 epochs—a duration determined in preliminary runs where we observed that the validation loss and generated image quality consistently stabilized after approximately 45 epochs, making a dynamic stopping criterion unnecessary. The data augmentation techniques were intentionally conservative to maintain structural stability, a key factor for our SAS metric. As noted previously, rotations were limited to $\pm 5^\circ$ to prevent grid distortion after resizing, while horizontal flipping and minor brightness/contrast adjustments were confirmed through testing to have no degrading effect on the geometric integrity

of the lattice structure required by the ControlNet.

The diffusion process was configured with 1,000 training timesteps.

Inference Details

For generating patterns, we used the DDIM (Denoising Diffusion Implicit Models) sampler with 50 inference steps and a classifier-free guidance scale of 7.5.

Preprocessing Algorithm

As previously mentioned, the inpainting of minor damages in the dataset was performed using a semi-automated process. Specifically, we used Adobe Photoshop’s Content-Aware Fill tool, an implementation based on the PatchMatch algorithm, with manual expert oversight to ensure stylistic integrity.

III. EXPERIMENTS AND RESULTS

To validate the effectiveness of our proposed method, we conducted a series of quantitative and qualitative experiments. The model was trained on a single NVIDIA A100 GPU for 50 epochs, using a learning rate of 1×10^{-5} and a batch size of 4.

A. QUANTITATIVE EVALUATION

Recognizing that these metrics do not explicitly evaluate the geometric integrity of the patterns, we introduce an additional, more targeted metric: the Structural Adherence Score (SAS). To calculate the SAS, we generated 500 images from each model and processed them with a Hough Line Transform, an algorithm designed to detect straight lines. An image was classified as “structurally adherent” if the algorithm successfully detected a minimum of three distinct, parallel diagonal lines in both directions, confirming the presence of a coherent rhombus grid. To ensure the reproducibility of this metric, the specific parameters for our Hough Line Transform implementation were as follows: a Canny edge detector with low and high thresholds of 50 and 150, respectively, was first applied. The Hough accumulation threshold was set to 100, the minimum line length was set to 50 pixels, and the maximum allowed gap between points on the same line was 10 pixels. For parallelism judgment, lines were grouped based on their angle; the angle tolerance for considering lines parallel was set to $\pm 2^\circ$. This robust set of parameters ensures that only clearly defined and consistent grid structures are counted, providing a stable and reproducible basis for the SAS calculation. The SAS represents the percentage of generated images that successfully met this structural criterion. FID measures the similarity between the distribution of the generated images and the real images; a lower score indicates higher fidelity. IS measures both the quality (clarity) and diversity of the generated images, with a higher score being better. To ensure reproducibility, all FID and IS calculations were performed using a standard implementation with a pre-trained InceptionV3 model. The FID score was computed

by comparing the 2,048-dimensional features from the final average pooling layer of 500 generated images against the features of all 150 images in our independent, held-out test set. The IS was calculated on the same set of 500 generated images. The reported scores are the result of a single, deterministic evaluation run. We compared our proposed method (Fine-tuned Diffusion + ControlNet) against two baselines: a standard Generative Adversarial Network (StyleGAN2) trained on the same dataset, and the fine-tuned diffusion model without the ControlNet conditioning.

As shown in Table 1, our proposed method achieved the best scores across all metrics. It obtained the lowest FID (16.51) and the highest IS (4.52), indicating that its outputs are both high-fidelity and diverse. More importantly, the SAS metric provides direct quantitative evidence of our method’s primary contribution. Our model achieved a near-perfect SAS of 99.2%, confirming that the ControlNet conditioning effectively enforces the crucial rhombus-grid structure. In stark contrast, both the StyleGAN2 and the diffusion model without control failed to maintain this structure, scoring only 5.2% and 8.6%, respectively. This result highlights a critical weakness in standard generative models that FID and IS scores alone cannot reveal, validating the necessity of our structural guidance approach. The FID score of 16.51 is significantly lower than that of StyleGAN2 (48.72) and the diffusion model without ControlNet (21.34), indicating that our generated patterns have a statistical distribution much closer to the real mural images. Similarly, our method’s IS of 4.52 is the highest, suggesting a superior combination of image quality and diversity. The improvement over the “No Control” diffusion model highlights the benefit of structural guidance; by offloading the task of learning the grid structure to ControlNet, the main model can focus entirely on rendering high-fidelity stylistic details.

TABLE 1. Quantitative Comparison of Generative Models

Model	Fr�chet Inception Distance (FID)	Inception Score (IS)	Structural Adherence Score (SAS)
StyleGAN2	48.72	3.15	5.2%
Fine-tuned Diffusion (No Control)	21.34	4.28	8.6%
Proposed Method (Ours)	16.51	4.52	99.2%

B. ABLATION STUDIES

To better understand the specific contributions of our key design choices—namely, the fine-tuning of the diffusion model and the use of Canny edges for conditioning—we conducted a series of ablation studies. We evaluated two additional model configurations: (1) a pre-trained (not fine-tuned) diffusion model with ControlNet, to isolate the effect of fine-tuning; and (2) our fully fine-tuned model conditioned on simple binary line art instead of Canny edges, to validate our choice of conditioning mechanism.

The results of our ablation studies, presented in Table 2, validate our methodological choices. The model without fine-tuning produced structurally coherent patterns (98.9% SAS) but failed to capture the specific artistic style of the Kizil murals, resulting in a much higher (worse) FID score

of 35.89. This demonstrates that fine-tuning is critical for achieving cultural and stylistic fidelity. Furthermore, when using binary line art as a condition instead of Canny edges, we observed a slight degradation in the FID score (18.24 vs. 16.51), suggesting that the sparse and detailed information from Canny edges provides a more effective condition for rendering nuanced stylistic details.

TABLE 2. Ablation Study on Model Components and Conditioning

Model Configuration	FID	IS	SAS	Qualitative Assessment
Proposed Method (Fine-tuned + Canny)	16.51	4.52	99.2%	Stylistically authentic & structurally correct.
No Fine-tuning + Canny	35.89	4.15	98.9%	Structurally correct, but lacks specific Kizil artistic style.
Fine-tuned + Line Art	18.24	4.48	99.1%	Structurally correct, but stylistic details are less refined.

C. QUALITATIVE EVALUATION

Visual inspection is crucial for evaluating artistic and stylistic success. Our model successfully generated a diverse array of patterns that consistently adhered to the rhombus-grid structure while faithfully reproducing the characteristic color palettes (such as lapis lazuli blue, mineral green, and earth red) and intricate motifs inherent in the Kizil murals. These generated patterns exhibited a high level of detail and stylistic consistency, showcasing the model's ability to create novel compositions that blend learned elements in aesthetically pleasing ways, rather than mere replications of the training data.

To further validate the quality, we conducted a formal user study with 30 participants, recruited via convenience sampling from local university design programs and digital art studios. The participants were stratified into two groups: 15 senior fashion design students and 15 professional digital artists. The study was conducted under a single-blind protocol, where participants were not informed of the origin of the images (i.e., whether they were real or AI-generated). Each participant was shown a set of 40 images (20 real, 20 generated) one at a time on a calibrated monitor. Crucially, the presentation order of these 40 images was uniquely randomized for each participant to mitigate order effects. Participants then rated each image on a 5-point Likert scale across three criteria: Authenticity (resemblance to historical Kizil art), Aesthetic Appeal, and Suitability for Fashion Application. The results in Table 3 are highly encouraging. The generated patterns scored very closely to the real ones in terms of Authenticity (4.41 vs. 4.78), indicating our model successfully captured the historical style.

Intriguingly, the generated patterns scored slightly higher on Aesthetic Appeal and significantly higher on Suitability for Fashion Application. Participants commented that the generated patterns often felt more “refined,” “consistent,” and “usable” for design, as they were free from the physical degradation present in the real murals and exhibited novel

combinations of motifs that were highly inspirational.

TABLE 3. User Study Results (Mean Scores on a 1–5 Scale)

Image Source	Authenticity	Aesthetic Appeal	Suitability for Fashion
Real Patterns	4.78	4.55	4.31
Generated Patterns (Ours)	4.41	4.62	4.75

D. FASHION DESIGN APPLICATION RESULTS

For the practical application, we integrated several patterns generated by our method into a digital fashion design workflow using CLO 3D software. The seamless patterns were successfully applied to various virtual garment models, such as dresses and coats. These applications demonstrated the effective translation of the 2D generated patterns onto 3D virtual fabrics. The patterns draped naturally over the garment forms, and their scale and placement were adjusted to complement the silhouettes, showcasing the aesthetic and functional potential in contemporary fashion design. The results effectively illustrate the tangible utility of our method as a tool for designers, enabling them to draw inspiration from cultural heritage and translate it into innovative, market-ready concepts with high efficiency and creative freedom.

IV. DISCUSSION

The results of our experiments strongly validate the proposed methodology. The quantitative metrics clearly position our conditional diffusion model as superior to both GAN-based approaches and non-conditional diffusion models for this specific task. The low FID score is a testament to the model's ability to learn the deep stylistic distribution of the Kizil patterns, while the high IS score confirms the generation of diverse and high-quality outputs. The crucial innovation of our work is the successful integration of ControlNet as a mechanism for enforcing structural constraints. As the comparison with the “No Control” diffusion model shows, simply fine-tuning a powerful model is insufficient for tasks requiring precise geometric arrangements.

Our results clearly demonstrate that the ControlNet module is the primary component responsible for this success. As shown in our ablation study, it ensures near-perfect structural consistency (98.9% SAS) even without fine-tuning the base model. The fine-tuning process, in turn, is critical for achieving stylistic fidelity and diversity, which is reflected in the significant improvement of the FID and IS scores. This effective decoupling—where ControlNet handles the structure and fine-tuning handles the style—is the key advantage of our framework.

The qualitative findings from the user study are perhaps even more insightful. The high rating for “Authenticity” suggests that our method achieves more than just superficial mimicry; it captures the nuanced character of the original artworks. The fact that generated patterns were rated higher in “Aesthetic Appeal” and “Suitability for Fashion” points

to an interesting dynamic: the AI is not just a replicator, but also a refiner.

By learning from a dataset cleaned of physical degradation, the model can generate idealized versions of the patterns that are more immediately applicable to the pristine world of fashion design. This highlights the potential for AI to act as a bridge, translating the raw, fragmented beauty of ancient artifacts into a polished, coherent form ready for contemporary creative reuse.

A crucial finding from our user study is that participants often preferred the generated patterns for being more “refined” and “usable,” precisely because they are free from the physical degradation of the original murals. This preference raises an important philosophical question at the heart of our work: is the goal to create an “idealized” version of the art, or to authentically reproduce its current, damaged state?

Our approach, which involves inpainting imperfections in the training data, deliberately leans toward digital restoration over literal replication. We argue that the purpose of this specific research is not to create a static, historically exact archive of the murals as they exist today. Rather, our aim is to regenerate the underlying artistic language of the Kizil patterns for the purpose of creative reuse and reinterpretation. By presenting the model with “clean” stylistic information, we enable it to function as a bridge, translating the fragmented beauty of an ancient artifact into a coherent and applicable form for contemporary designers.

However, we acknowledge the ethical responsibility inherent in this choice. These AI-generated patterns should not be viewed as historically authentic reproductions. They are modern interpretations inspired by a historical source. It is critical that this distinction is maintained, ensuring that the “idealized” digital version serves to promote and celebrate the original cultural heritage rather than supplanting the complex reality of the physical artifact.

Despite the promising results, this study has several limitations. First, regarding the methodology, our user study provides valuable qualitative insights but was conducted with a limited sample size of 30 participants, which is not large enough to establish statistical significance. The expert group was composed of design students and digital artists; future work would be strengthened by including art historians to better evaluate authenticity and a non-expert control group to gauge broader aesthetic perceptions. Second, regarding the dataset, its scale of 1,500 images, while sufficient for fine-tuning, is relatively small, and a larger dataset could enable the generation of more varied motifs. Third, while our method provides excellent control over the overall grid structure, control over the specific motifs within each rhombus remains an area for future exploration. Finally, our practical application was demonstrated in a digital-only workflow, and future work should explore the physical realization of these designs through digital textile printing. The broader implications of this research are significant. It provides a robust and replicable framework for the intelligent generation of any structured cultural

pattern, from Islamic geometric tiles to Celtic knotwork. This technology empowers designers, artists, and cultural historians, providing a tool that respects historical accuracy while enabling boundless creative exploration. It represents a new paradigm for cultural heritage preservation, one that moves beyond static digital archives and toward dynamic, interactive, and generative systems that can breathe new life into ancient art forms, ensuring their relevance and appreciation for generations to come.

V. CONCLUSION

This research successfully developed and validated a novel framework for the intelligent regeneration of structured cultural heritage patterns, demonstrating a powerful synergy between latent diffusion models and conditional control mechanisms. Our core contribution is demonstrating that the key to this challenge lies in a two-part approach: using a strong conditional mechanism like ControlNet is essential and primarily responsible for guaranteeing geometric integrity, while fine-tuning the base diffusion model is crucial for capturing the nuanced artistic style of the source material. The results show that this approach not only preserves the stylistic fidelity of the Kizil murals but also produces novel, aesthetically compelling patterns that are immediately applicable to modern creative industries like digital fashion.

More broadly, this work reframes the role of AI in cultural heritage from a simple tool for digital archiving to a dynamic partner in creative reinterpretation. By providing a method to regenerate the underlying artistic language of historical artifacts, we open up new paradigms for how designers, historians, and the public can interact with and draw inspiration from the past. This methodology provides a robust blueprint for regenerating other structured art forms, from Islamic tiling to Celtic knotwork, ensuring that ancient artistic traditions remain a vibrant and evolving part of our contemporary visual culture.

VI. FUTURE WORK

While this study establishes a strong foundation, several ambitious research avenues remain. For more granular creative control, our method could be extended to cell-specific content generation. We hypothesize that this could be achieved by replacing the single Canny edge map with a multi-class segmentation map, where each rhombus is assigned a unique ID. This structural input could be paired with a parsed text prompt (e.g., “Cell 1: a floral motif; Cell 2: a mythical beast”), allowing the model to use localized cross-attention to populate the grid with specific narrative or decorative elements, effectively turning the pattern into a controllable canvas.

Furthermore, the transition from screen to physical fabric warrants a dedicated research effort beyond simple color space conversion. Future work should explore material-aware pattern generation, where the AI model is conditioned

not only on style and structure but also on the physical properties of a target textile.

This could involve integrating PBR parameters, such as fabric weave and roughness maps, directly into the model's training process, allowing it to generate patterns that are aesthetically optimized for the texture and drape of a specific material, like silk or cotton.

AUTHOR CONTRIBUTIONS

Qixing Lei and Cheng Chang designed the study; all authors conducted the study; Qixing Lei and Jianjun Wang collected and analyzed the data. Jianjun Wang and Cheng Chang participated in drafting the manuscript, and all authors contributed to critical revision of the manuscript for important intellectual content. All authors gave final approval of the version to be published. All authors participated fully in the work, took public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or completeness of any part of the work were appropriately investigated and resolved.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This survey was conducted in compliance with Ethics Committee of Shanghai Technical Institute of Electronics

& Information. Participants were informed of the study's purpose and data usage prior to participation, and responses were collected anonymously. No personally identifiable information was stored.

AVAILABILITY OF DATA AND MATERIALS

The datasets used and/or analysed during the current study were available from the corresponding author on reasonable request.

REFERENCES

- [1] M. Zin, "The Kanaganahalli Stupa: An Analysis of the 60 Massive Slabs Covering the Dome," New Delhi, India: Aryan Books International; 2018.
- [2] Z. Zhou, L. Shen, and H. Zhang, "The Wall Painting Techniques and Materials of Kizil Grottoes," in Aoki S, Taniguchi Y, Rickerby S, et al., editors. *Conservation and Painting Techniques of Wall Paintings on the Ancient Silk Road*. Cultural Heritage Science. Singapore: Springer, 2021, pp. 235-251, doi: 10.1007/978-981-33-4161-6, [Online]. Available: <https://link.springer.com/book/10.1007/978-981-33-4161-6>.
- [3] Y. Hayashi, "Research of Blueness: Lapis Lazuli in the Kizil Grottoes Murals of the Silk Road," University of Helsinki [dissertation]. Helsinki, Finland: University of Helsinki 2019, [Online]. Available: <http://hdl.handle.net/10138/301910>.
- [4] D. Bruton, "Theorizing Digital Cultural Heritage: A Critical Discourse," *Information, Communication & Society*, vol. 14, no. 7, pp. 1077-1078, 2011, doi: 10.1080/1369118X.2010.542826.
- [5] F. Spettu, "Towards a new approach to BIM for Cultural Heritage: A framework for the reverse engineering of Milan Cathedral [dissertation]," Milan, Italy: Politecnico di Milano, 2020.
- [6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative Adversarial Nets," *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*; December 8 - 13, pp. Montreal, Canada: MIT Press; 2014. p. 2672-2680, 2014, [Online]. Available: <https://dl.acm.org/doi/10.5555/2969033.2969125>.
- [7] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*; 5-10 December 2016; Barcelona, Spain. Red Hook, NY, USA: Curran Associates Inc.; 2016. p. 2234-2242, [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/3157096.3157346>.
- [8] J. Ho, A. Jain, and P. Abbeel, "Denosing diffusion probabilistic models," *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20)*; 6-12 December 2020; Vancouver, Canada. Red Hook, NY, USA: Curran Associates Inc.; 2020. Article 574; p. 6840-6851, [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/3495724.3496298>.
- [9] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics," *Proceedings of the 32nd International Conference on Machine Learning - Volume 37 (ICML'15)*; 6-11 July 2015; Lille, France. JMLR.org; 2015. p. 2256-2265, [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/3045118.3045358>.
- [10] U. Bergmann, N. Jetchev, and R. Vollgraf, "Learning texture manifolds with the Periodic Spatial GAN," *Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17)*; 6-11 August 2017; Sydney, Australia. JMLR.org; 2017. p. 469-477, [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/3305381.3305430>.
- [11] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-To-Image Translation with Conditional Adversarial Networks," *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125-1134, 2017, doi: 10.1109/CVPR.2017.632.
- [12] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; 1-6 October 2023; Paris, France. IEEE; 2023. p. 3813-3824, doi: 10.1109/iccv51070.2023.00355.
- [13] Y. Li, H. Liu, Q. Wu, F. Mu, J. Yang, J. Gao, et al., "GLIGEN: Open-Set Grounded Text-to-Image Generation," *Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 17-24 June 2023; Vancouver, Canada. IEEE; 2023. p. 22511-22521, doi: 10.1109/cvpr52729.2023.02156.