

DiffTile: High-Resolution Periodic Pattern Generation via Splice, Focus and Enhance

Xicheng Lan, Chenyu Niu , Shenghan Xu , Wenshuo Gao , Luyao Zhang , Jiaying Liu , *Fellow, IEEE*,
and Shuai Yang , *Member, IEEE*

Abstract—High-resolution periodic patterns are widely used in the printing and dyeing industry. However, the manual design of such periodic patterns has considerable time and workforce costs. Recent advances in powerful image generation models, which are represented by latent diffusion models (LDMs), provide an automated solution. However, owing to the inherent randomness of the generation process and the computational complexity of LDMs, generating strictly periodic and high-resolution patterns is still challenging. In this paper, a novel text-driven framework for high-resolution periodic pattern generation is proposed. For periodic pattern generation, we propose a training-free Splice-and-Focus mechanism, which constrains latent features and modifies the attention mechanism to generate natural and strictly periodic patterns. On this basis, we provide two types of generation modes: direct periodic pattern generation and periodization of arbitrary existing patterns. For high-resolution pattern generation, we propose a ReLife pipeline that integrates super-resolution and image detail enhancement to improve pattern resolution while removing generative artifacts and distortions. Furthermore, for highly nonstationary patterns, we propose a compositional generation framework to design complicated patterns interactively. The experimental results show that our framework can generate high-quality patterns that meet the strict standards of the printing and dyeing industry.

Index Terms—Pattern generation, latent diffusion model, attention, super-resolution, periodicity.

I. INTRODUCTION

Printing and dyeing refer to the process of applying desired patterns onto fabric, such as cloth or silk, and other

decorative surfaces, in which pattern design plays an essential role. In contrast to typical image designs, the patterns that are employed in the printing and dyeing industry face stricter requirements for **periodicity** and **high-resolution**. First, because dyeing and painting are typically performed on large sections of various types of large surfaces, patterns should consist of repetitive splicing of periodic units. This ensures that the overall design remains seamless and uniform throughout the entire surface. Second, the pattern must maintain high-resolution, because clear and sharp designs are crucial for achieving superior quality in the finished products.

Currently, most patterns are crafted by hand, using traditional methods such as painting or digital drawing tools. This process incurs substantial time and labor costs, which makes it difficult to meet the needs of digital printing and dyeing. The use of generative models in pattern design can significantly increase efficiency. In addition, these models provide customization and adaptability, meeting various design requirements. Latent Diffusion Models (LDMs) [1], [2], [3] have advanced notably in text-driven image generation and can facilitate the creation of high-quality images from given prompts [1], [4], [5], [6]. Moreover, existing customization technologies [7], [8], [9], [10], [11], [12], [13], [14], [15], [16] such as Low-Rank Adaptation (LoRA) [10] enable the fine-tuning of LDMs tailored to specific styles based on curated datasets. However, difficulties persist in using existing technologies to generate periodic and high-resolution patterns. Although fine-tuning an LDM on a dataset of periodic images and adding the term “periodic” to the prompt as classifier-free guidance [17] or imposing layout conditions [18], [19], [20], [21], [22], [23] may encourage the model to create somewhat periodic patterns, the intrinsic randomness of LDM generation poses challenges in achieving true periodicity. Existing image generation techniques struggle to produce the periodic pattern units required in the dyeing and printing process. As discussed earlier, this process imposes stringent requirements on periodicity, which makes direct splicing of generated patterns impractical. Consequently, a method that can directly generate periodic patterns is urgently needed to enable the effective application of current advanced image generation technologies in the printing and dyeing industry.

Moreover, the substantial computing power and memory that are needed for LDMs make directly creating ultrahigh-resolution patterns impractical (e.g., a US-Letter-sized pattern is approximately 3,960 pixels in height at 360 dpi). Directly generating such high-resolution images is impossible for most

Received 10 July 2025; revised 21 November 2025; accepted 24 November 2025. Date of publication 23 March 2026; date of current version 21 September 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62471009 and Grant 62332010, in part by the Key Laboratory of Science, Technology and Standard in Press Industry (Key Laboratory of Intelligent Press Media Technology), in part by The Fundamental Research Funds for the Central Universities, Peking University, and in part by Zhejiang Haiyin Digital Technology. The associate editor coordinating the review of this article and approving it for publication was Dr. Xuequan Lu. (Chenyu Niu and Shenghan Xu contributed equally to this work.) (Corresponding author: Shuai Yang.)

Xicheng Lan, Chenyu Niu, Shenghan Xu, Wenshuo Gao, Luyao Zhang, and Jiaying Liu are with the Wangxuan Institute of Computer Technology, Peking University, Beijing 100080, China (e-mail: lanxicheng@stu.pku.edu.cn; niuchenyu_0505@stu.pku.edu.cn; xsh2022@stu.pku.edu.cn; 2100017722@stu.pku.edu.cn; zhangluyao@pku.edu.cn; liujiaying@pku.edu.cn).

Shuai Yang is with the Wangxuan Institute of Computer Technology, Peking University, Beijing 100080, China, and also with the State Key Laboratory of Multimedia Information Processing, Peking University, Beijing 100080, China (e-mail: williamyang@pku.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TMM.2026.3676838>, provided by the authors.

Digital Object Identifier 10.1109/TMM.2026.3676838

hardware. An alternative approach is to use super-resolution technologies [24], [25], [26] as a postprocessing step. However, these methods typically emphasize faithfulness, namely, they assume that the low-resolution inputs are correct, which is seldom the case for us. As a result, the artifacts and distortions that are generated during the diffusion process remain irreparable and may even become more pronounced when patterns are enhanced to high-resolution by these technologies. Moreover, the application of existing super-resolution processing methods on periodic patterns can introduce inconsistencies at the splicing boundaries, which significantly hinder the practical deployment of image generation techniques in printing and dyeing workflows. Therefore, an effective technique should not only increase the image resolution but also repair imperfections in low-resolution images and ensure the periodicity of textural and structural details at higher resolutions.

In this paper, we propose a novel text-driven periodic and high-resolution pattern generation framework. We first design a Splice-and-Focus mechanism for LDMs that constrains latent features and modifies the self-attention mechanism into FocAttention to enable the improved LDM to generate natural and periodic patterns by either directly creating new designs or periodizing existing designs. Notably, this mechanism requires no training, thus offering an elegant yet highly effective solution for generating patterns that strictly adhere to scalable periodicity. In addition, we propose a new ReLife pipeline. It combines super-resolution and image detail enhancement technologies to simultaneously increase image resolution and eliminate artifacts and distortions while maintaining the required periodicity, thereby resulting in the generation of pattern details.

Compared with our previous work [27], we explore improving the image quality and expanding the scope of application. First, in terms of image quality, we transplant the basic LDM baseline from SD1.5 [1] to SDXL [28]. The generation process of SDXL results in higher basic resolution and enables control of the prompt for generating the image content. To adapt to the attention characteristics of SDXL, we introduce a new central attention mechanism into FocAttention to ensure image quality. Second, we propose Periodic Inpainting, which is a method for transforming nonperiodic patterns into periodic patterns. Third, we expand the scope of application to include the generation of complex patterns by introducing an interactive compositional generation framework. Furthermore, comprehensive experiments are conducted to analyze the pattern generation performance of the proposed DiffTile, including comparisons of visual results on more styles and complicated patterns, additional comparisons with diffusion-based super-resolution methods, and ablation studies on the improved attention mechanism and Periodic Inpainting design. In summary, our contributions are fivefold:

- We provide a new pattern generation framework, DiffTile, which is based on SDXL to meet the high-resolution and periodicity requirements of the printing and dyeing industry.
- We introduce a training-free Splice-and-Focus mechanism that constrains the LDM generation process with latent splicing and modified attention, which ensures strict periodicity.

- On the basis of the Splice-and-Focus mechanism, we design two generation modes to address different demands. Our method allows for new periodic pattern generation that is driven by prompts as well as the periodization of existing nonperiodic designs.
- We propose a new ReLife pipeline that progressively improves the resolution of the periodic pattern through a combination of super-resolution and detail enhancement and yields $\geq 4096 \times 4096$ high-quality patterns.
- We propose a compositional generation method to expand the scope of application to include complicated pattern design.

The remainder of the paper is organized as follows. In Section II, we review related works on diffusion-based image generation and super-resolution. Section III introduces the proposed Splice-and-Focus mechanism and ReLife pipeline for high-resolution periodic pattern generation. Section IV details the proposed compositional generation method for complicated patterns. Section V presents the experimental results, and Section VI provides concluding remarks.

II. RELATED WORK

Latent Diffusion Models: Denoising diffusion probabilistic models (DDPMs), which were developed by Ho et al. [29], enable high-quality image synthesis via a learned reverse noising process. Rombach et al. introduced Latent Diffusion Models (LDMs) [1], which compress images into a lower-dimensional latent space before diffusion, which dramatically reduces computational cost while preserving fidelity. To accelerate sampling, Song et al. proposed Denoising Diffusion Implicit Models (DDIMs) [30], which are non-Markovian variants that require far fewer steps, and Nichol and Dhariwal refined noise schedules and variance estimation to reduce inference iterations by an order of magnitude [31]. Theoretical foundations from Sohl-Dickstein et al. underlie these methods [32]. Most recently, Stability AI's SDXL 1.0 has threefold-scale LDM backbones, adds dual text encoders and multiaspect conditioning, and includes a postrefiner for ultrahigh-fidelity output at resolutions up to 2048×2048 [28]. We build our model on the basis of SDXL to leverage its strong generative capacity and focus on adapting it to high-resolution and periodic pattern generation. Furthermore, our method differs from several structured generation approaches, including layout-control methods like ControlNet [18] and object-positioning methods like GLIGEN [19]. Whereas they focus more on fine-grained control of image details and layout, our method aims to generate strictly periodic patterns. Our method also has the potential to be integrated with ControlNet-based techniques to provide improved structural control without compromising the periodicity of the generated patterns.

Periodic Pattern Generation: Early procedural texture synthesis relied on reaction-diffusion systems: Witkin and Kass demonstrated that cyclic boundary conditions in a reaction-diffusion PDE model produce seamless, tileable patterns [33]. More recently, "Infinite Texture" was proposed, which fine-tunes an LDM on a single patch and aggregates score estimates to generate arbitrarily large, perfectly repeating textures from text prompts [34], whereas "Tiled Diffusion" embeds tiling

mechanics into the diffusion network to output multiple connected tiles in one forward pass [35]. Representation-space diffusion methods [36] enforce lattice symmetries in the latent denoising process to ensure strict periodicity. In the textile domain, Patel et al. [37] fine-tuned diffusion models on large fabric datasets and proposed quantitative tileability metrics for assessing pattern consistency. However, the approaches above share several limitations, including costly fine-tuning and poor adaptability to arbitrary inputs. Our Splice-and-Focus mechanism applies constraints at inference time without extra training. Moreover, our Periodic Inpainting enables fine-grained segmentation and inpainting of arbitrary input images to produce seamless, style-consistent periodic patterns.

Diffusion-Based Super-Resolution: Diffusion models have been adapted for super-resolution [38], [39], [40]: SR3 [39] uses a conditional U-Net to progressively remove noise to generate high-res output and achieves up to $8\times$ upscaling on faces and natural scenes. To mitigate quadratic attention costs at high resolutions, Inf-DiT [40] introduces unidirectional block attention, which reduces memory complexity and enables the achievement of 4096×4096 super-resolution on a single GPU. For blind super-resolution with unknown degradations, SR3+ [41] employs self-supervised degradations and noise-conditioning augmentations to outperform SR3 [39] and GAN-based methods [42] on out-of-distribution inputs. Dual-Diffusion (DDSR) [43] extends this to remote-sensing imagery by conditioning on both degradation kernels and low-resolution inputs within a dual-stage diffusion framework. StableSR [44] is a blind super-resolution framework that preserves pretrained diffusion priors using a time-aware encoder and controllable feature wrapping to enable high-quality, resolution-flexible restoration. SeeSR [45] is a semantics-aware real-world image super-resolution method that leverages both hard and soft prompts from a degradation-aware extractor to guide text-to-image diffusion models toward semantically faithful and detailed reconstructions. These methods presume the low-resolution input is “clean”, and they faithfully upscale and even amplify any artifacts or seams. By comparison, our ReLife bridges this gap by combining super-resolution with enhancement to eliminate the generative artifacts. We enforce consistency across tile boundaries by latent-space splicing to simultaneously achieve high-fidelity upscaling and strict periodicity preservation.

III. HIGH-RESOLUTION PERIODIC PATTERN GENERATION

A. Preliminaries

We build our framework upon the Latent Diffusion Model of SDXL. Latent diffusion models operate in the latent space with an autoencoder $\mathcal{D}(\mathcal{E}(\cdot))$, where $\mathcal{E}$ and $\mathcal{D}$ represent the encoder and decoder, respectively. Given an image I , LDM encodes it to a latent feature $x_0 = \mathcal{E}(I)$, and the diffusion forward process adds noise to the latent feature [29]:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad (1)$$

where $t = 1, \dots, T$ represents the time step, x_t represents the noisy latent at time step t , $\bar{\alpha}_t$ is a hyperparameter and ϵ is a randomly sampled Gaussian noise. In the diffusion backward

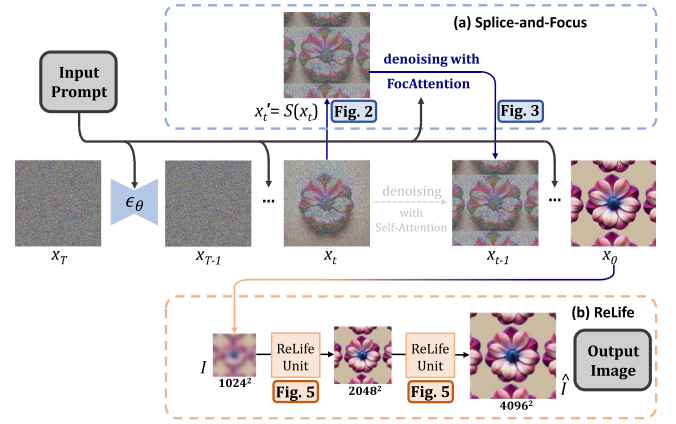


Fig. 1. Illustration of the proposed DiffTile framework for high-resolution periodic pattern generation.

process, a diffusion model ϵ_θ is trained to predict the noise of the latent feature to iteratively recover x_0 from x_T .

During inference, starting from a random Gaussian noise $x_T \sim \mathcal{N}(0, \mathbf{I})$ and a text prompt p as condition, we can sample x_{t-1} on the basis of x_t by DDIM sampling [30]:

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_{t \rightarrow 0} + \sqrt{1 - \bar{\alpha}_{t-1}}\epsilon_\theta(x_t, t, p), \quad (2)$$

where $\hat{x}_{t \rightarrow 0}$ represents the predicted x_0 at time step t ,

$$\hat{x}_{t \rightarrow 0} = (x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta(x_t, t, p))/\sqrt{\bar{\alpha}_t}, \quad (3)$$

and $\epsilon_\theta(x_t, t, p)$ represents the predicted noise of x_t . Finally, the generated image is obtained by decoding x_0 as $I = \mathcal{D}(x_0)$.

B. Overview of the DiffTile Framework

Fig. 1 presents an illustrated overview of our proposed DiffTile framework. Our framework consists primarily of two key components: Splice-and-Focus and ReLife. Given an input prompt p that describes the pattern content, we first sample Gaussian noise x_T and perform DDIM sampling. In the first stage, as shown in Fig. 1(a), during DDIM sampling, we enforce periodicity constraints through a plug-and-play splicing and pasting operation on the latent x_t at each time step t . Simultaneously, we replace the self-attention module in the denoising network ϵ_θ to focus only on valid regions, thereby ensuring high-quality periodic pattern generation. This yields a preliminary low-resolution image $I = \mathcal{D}(x_0)$. In the second stage, as shown in Fig. 1(b), we progressively increase the resolution of the periodic pattern I through an alternating process of super-resolution and diffusion-based detail enhancement. Throughout this phase, latent space splicing is applied to maintain strict periodicity, and a $4\times$ upsampled high-resolution output $\hat{I}$ is ultimately produced. To generate patterns of specific styles, we employ LoRA [10] to customize the model on curated pattern datasets.

C. Periodic Pattern Generation Via Splice-and-Focus

To meet the key requirements for scalable and periodic patterns in the design, printing and dyeing industries, this section focuses on generating a primary periodic pattern. Our goal is for

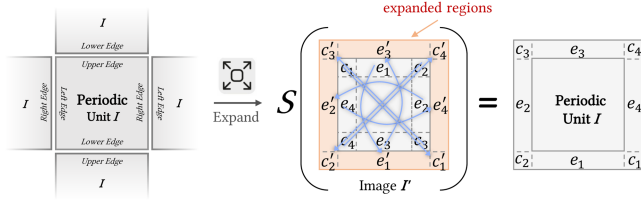


Fig. 2. The extensible periodic image maintains continuity after extension. The central valid area I remains unchanged, while the outer expanded regions are constrained to achieve continuity.

the generated primary pattern to be repeated seamlessly over a large piece of canvas or fabric without any visible splicing. Although diffusion models have shown powerful generative abilities, they fail to synthesize strictly periodic patterns even after fine-tuning on real periodic images and using specific keywords in the prompt, such as “periodic”.

To address this challenge, we introduce a novel plug-and-play Splice-and-Focus mechanism, which comprises **Latent Splice** and **FocAttention**, into the LDMs to ensure precise repetition. As shown in Fig. 2, the key strategy is to expand the periodic image unit I outward to a larger image I' . The seamless alignment of the left and right edges, as well as the top and bottom edges of I , can naturally be understood as ensuring the consistency between the expanded regions and their corresponding regions within I . In other words, if these corresponding regions of I' are the same (e.g., $e_1 = e'_1$ and $c_1 = c'_1$), then I should be strictly periodic. To achieve this goal, we propose **Latent Splice** and **FocAttention**, from the aspects of latent features and attention, respectively.

Building upon the spatial constraints above, we propose two pipelines for periodic pattern generation. One pipeline treats the task as a fully generative process and imposes spatial constraints during generation to directly obtain periodic patterns. Thus, when users want to create periodic patterns from scratch, this pipeline is helpful for directly customizing the patterns on the basis of prompts. The other pipeline, which is named **Periodic Inpainting**, transforms a generated nonperiodic pattern into a periodic pattern by leveraging spatial constraints together with foreground segmentations. Periodic Inpainting can be applied to arbitrary input images, whether generated by AI or designed by humans. Thus, when users need to generate a new periodic pattern that is essentially the same as an existing favorite pattern, the second pipeline provides a quick solution in one minute.

1) *Latent Splice*: LDMs project images to latent features [46] and generate target images in the latent space. Generating periodic images corresponds to generating periodic latent features, which is equivalent to ensuring uniformity across different sections in the latent space after our region expansion. Specifically, we begin by sampling a random latent feature $x_T \sim \mathcal{N}(0, \mathbf{I})$, whose resolution corresponds to I' after it is projected back to the image space. Afterward, we define a substitution function $\mathcal{S}$, which operates in the latent space of the LDMs:

$$x'_t = \mathcal{S}(x_t), \quad (4)$$

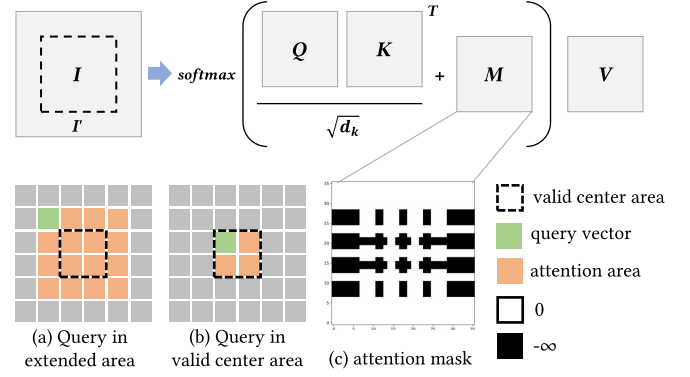


Fig. 3. Illustration of our FocAttention. For ease of understanding, we omit the process of reshaping tensors into vectors.

where x_t represents the noisy latent feature of the LDMs at time step t . As illustrated in Fig. 2, $\mathcal{S}$ splices x_t and directly copies the valid central part of x_t to the corresponding overlapping outer regions to ensure strict consistency. The resulting x'_t is subsequently used in the next denoising step to obtain x_{t-1} , and this process is continued until $t = 0$, at which point the fully denoised latent feature is generated [29].

2) *Focattention*: FocAttention enforces soft consistency by constraining the valid attention regions to ensure a smoother stitching effect in the generation process. Our experiments indicate that hard latent substitutions may result in unnatural seams. Because the valid portion of the final image is the central region, our objective is to improve the standard self-attention mechanism of LDMs:

$$\text{Self-Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V. \quad (5)$$

In our previous work [27], we added a hard constraint to address only the periodic unit area. Specifically, we modified the vectors K and V to retain only the valid center area, which resulted in K' and V' , respectively. Then the self-attention became the following:

$$\text{FocAttention}(Q, K, V) = \text{softmax}\left(\frac{QK'^T}{\sqrt{d_k}}\right) V'. \quad (6)$$

However, our experiments show that such a hard constraint sometimes leads to quality degradation, such as spatially inconsistent pattern styles and texture fragmentation. To ensure strict periodicity and visual harmony, we propose a soft constraint to redistribute attention using an attention mask M , which disables attention access to inner regions adaptively to query locations, as shown in Fig. 3(c). Specifically, we create an attention mask M so that each query in the expanded region can attend only to regions closer to the center than itself. In such regions, the value of the attention mask is set to 0, whereas in other regions, it is set to $-\infty$, as shown in Fig. 3(a). Formally, suppose the latent size is $(2H, 2W)$ and that the coordinates of the top-left token are $(0, 0)$. Then the attention mask M is of size $(4HW, 4HW)$. We define a function $\text{dist}(i, j) = \max(|H - 0.5 - i|, |W - 0.5 - j|)$ to map the token at (i, j) to its distance from itself to the center

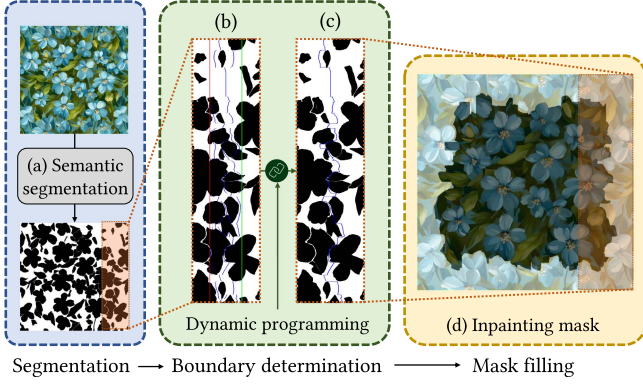


Fig. 4. Obtaining the ideal mask of the center reserved area. (a) Foreground segmentation by SAM2. (b) Boundary constraining and candidate boundary selection. (c) Mask boundary determined by dynamic programming. (d) Mask determined by filling the boundaries.

point $(H - 0.5, W - 0.5)$. Then, the attention mask M is derived as follows:

$$M_{2Wi+j, 2Wm+n} = \begin{cases} 0, & (m, n) \text{ in valid center area or} \\ & \text{dist}(i, j) \leq \text{dist}(m, n), \\ -\infty, & \text{else.} \end{cases} \quad (7)$$

Queries in the valid center area remain the same as in our previous design [27] to attend only to the center area, as illustrated in Fig. 3(b). Our improved FocAttention takes the form

$$\text{FocAttention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} + M \right) V, \quad (8)$$

FocAttention efficiently modifies the attention mechanism, but enables the model to focus on the valid region in the center. In Section V, we demonstrate in Fig. 15 that by incorporating (8) into the denoising process, we can generate natural periodic images.

3) *Periodic Inpainting*: In addition to the purely generative pipeline, we introduce another approach, which can convert arbitrary nonperiodic images into seamless periodic outputs. The pipeline can be divided into two stages: foreground object segmentation and periodic inpainting.

In the first stage, we aim to accurately identify the central content to be preserved during the inpainting process. We experimentally determine that a regular rectangular mask may result in stitching artifacts. An ideal mask should avoid covering as few foreground areas as possible. Specifically, as illustrated in Fig. 4, we perform the following steps:

- *Segmentation*: We segment the foreground region that has high saliency using SAM2 [47].
- *Boundary constraints*: As shown in Fig. 4(b), for each side, we define a green outer boundary and a red inner boundary. Pixels inside the inner boundary and outside the outer boundary are preserved and inpainted, respectively. We determine the mask boundary between the inner boundary and the outer boundary.
- *Candidate boundary selection*: Within the region between the inner and outer boundaries, we identify candidate mask boundaries (e.g., the blue lines in Fig. 4(b)) by selecting

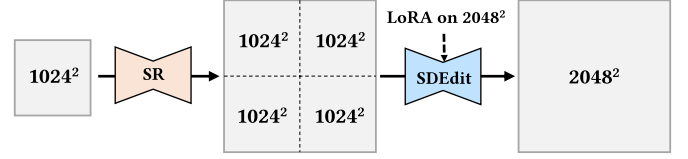


Fig. 5. One single unit of our ReLife pipeline.

the midpoints of consecutive white pixels in each row (or column). Such candidates separate different objects while avoiding cutting through the same object.

- *Boundary determination*: Dynamic programming is applied to select an optimal subset of the candidate boundaries such that the line (e.g., blue lines in Fig. 4(c)) that connects them crosses the fewest foreground pixels. This line forms a continuous mask boundary line on one side.
- *Mask filling*: The above steps are repeated independently for all four sides. Finally, the closed area that is defined by the above four boundaries forms the inpainting mask m , as shown in Fig. 4(d).

The second stage is a standard inpainting process [48], [49] with two previously introduced splice-and-focus constraints:

$$x_{t-1}^{\text{unknown}} = \sqrt{\bar{\alpha}_{t-1}} \hat{x}_{t \rightarrow 0} + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_{\theta}(\mathcal{S}(x_t), t, p), \quad (9)$$

$$x_{t-1}^{\text{known}} = \sqrt{\bar{\alpha}_{t-1}} x_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon, \quad (10)$$

$$x_{t-1} = (1 - m) \odot x_{t-1}^{\text{known}} + m \odot x_{t-1}^{\text{unknown}}. \quad (11)$$

In particular, our method does not require additional training. Thus, it offers an elegant yet highly effective solution for constraining the generative process to enable the creation of scalable periodic images.

D. ReLife for Super-Resolution and Enhancement

In this section, the primary periodic pattern that is generated in Section III-C is resolved and refined to meet the resolution standards for printing and dyeing. Current super-resolution methods excel at restoring images [24], [25], [26] or enriching texture details [39], [44], [50], but they focus mainly on fidelity rather than generation. They tend to preserve the artifacts and defects of the input structures or textures. Thus, these methods cannot automatically repair artifacts or regenerate clear structures, such as replacing broken flower branches with realistic branches.

To address this unique challenge, we design a specialized ReLife pipeline, as shown in Fig. 1(b), whose units are illustrated in Fig. 5, to simultaneously improve the image resolution and details to meet the stringent quality requirements of printing and dyeing. Our pipeline is composed of two key stages: one dedicated to super-resolution, and the other to generation for defect correction. These stages are alternated to ensure optimal results.

1) *Super-Resolution Stage*: We apply a pretrained image super-resolution model (e.g., Real-ESRGAN [42], [51]) to obtain a high-resolution image. Afterward, we segment the high-resolution image into overlapping subimages to ensure that each subimage remains within the processing capacity of the LDMs in the next stage.

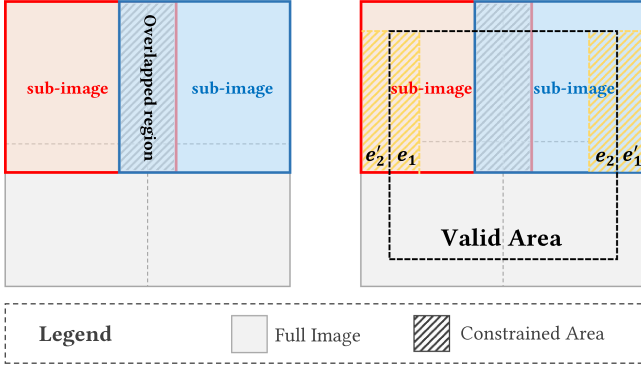


Fig. 6. Constrain corresponding parts of sub-images for seamless stitching.

2) *Generation Stage*: We use SDEdit [52] for image-guided image synthesis. Specifically, SDEdit introduces noise into the super-resolved subimage and then denoises and enhances its content under the guidance of text prompts. In addition, we employ LoRA [10] to fine-tune the model with resolution-specific datasets to ensure that the model can generate structures and details at the target resolution.

Finally, we stitch the segmented subimages. Taking Fig. 6 as an example, to ensure seamless and consistent stitching, we constrain the corresponding areas in the subsequent blue subimage based on the previously generated red subimage. In each time step of SDEdit, we impose constraints on seamlessness and periodicity. In terms of seamlessness, we segment the overlapping subimages. To enforce consistency within each overlapped region, we set the overlapped region of the red subimage as the known part and inpaint the remaining region of the blue subimage following [48], [49]. In terms of periodicity, we employ a modified latent splicing strategy, *i.e.*, copying the latent space content in e_1 and e'_2 from the red subimage to the corresponding regions e'_1 and e_2 , respectively, of the blue subimage.

By alternating these steps with the above two constraints, we are able to enhance periodic patterns while preserving a consistent style and improving the image resolution.

IV. COMPOSITIONAL PATTERN GENERATION

In practice, many exquisite designs are composed of multiple elements to form complex nonstationary patterns. We consider the Dunhuang caisson patterns in Fig. 7(a) as an example. They exhibit prototypical complexity, where each motif comprises multiple concentric layers of strip-like bands that surround a centrally symmetric circular medallion, thus producing rich content, pronounced repetition, and radial symmetry. Directly learning such highly structural multilayered data distributions with diffusion models often results in failures to preserve seamless periodicity and in texture mixing and missing details.

This section focuses on complicated nonstationary pattern design. We propose a compositional pattern generation pipeline that is based on our Splice-and-Focus mechanism and instantiate our pipeline on the caisson pattern design. Specifically, our extended pipeline decomposes the complicated pattern into two

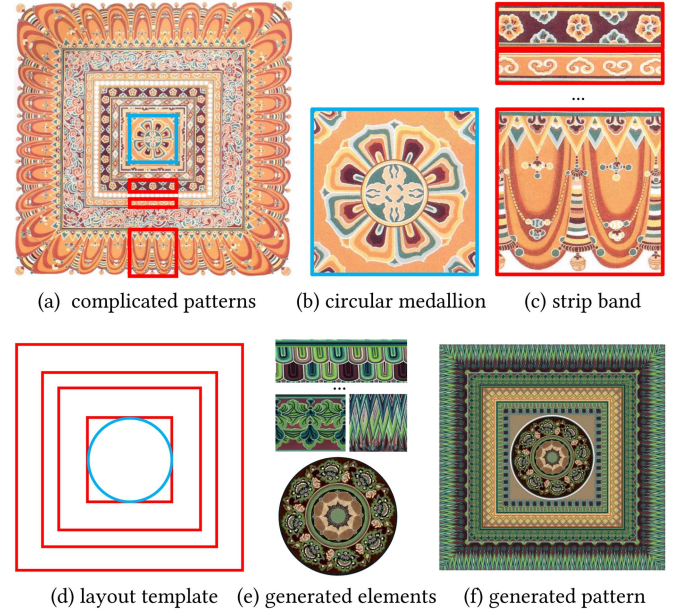


Fig. 7. Example of complex nonstationary patterns, comprising multiple outer strip bands encircling an inner concentric circular medallion. Our method decomposes the complicated pattern into two fundamental elements, generate and combine them according to their layout to yield a faithful pattern design.

fundamental components: outer strip bands and an inner circular medallion.

- **Strip bands**: Our DiffTile framework can naturally generate strip band periodic units, as shown in Fig. 7(c).
- **Circular medallion**: Directly constraining the circular periodicity is difficult. We employ polar coordinate transformations to transform the circular periodicity into horizontal periodicity, which can then be effectively handled by our DiffTile framework.

Finally, the two kinds of outputs are composited according to their layer order to produce a faithful design of the complex caisson pattern, as illustrated in Fig. 7(d)–(f).

A. Interactive Compositional Generation Framework

1) *Model Customization*: To construct a dataset for LoRA, we begin by manually cropping the periodic units of each outer strip-band and the central circular medallion from the caisson patterns. After that, we apply the polar coordinate transform to the circular units, as shown in Fig. 8(a), (b), which helps to transform the circular periodicity into horizontal periodicity. We then customize SDXL on the strip units and the unwrapped circular units.

2) *Element Generation*: After model customization, we invoke our Splice-and-Focus Mechanism and FocAttention to generate horizontally periodic units. For the strip bands, we horizontally/vertically tile the periodic units to reconstruct each concentric ring layer of the caisson. For the central medallions, we apply the inverse polar coordinate transform and assemble the periodic units to restore the perfectly symmetric circular medallion, as shown in Fig. 8(c), and (d).

3) *Element Composition*: Once all the elements are generated, we composite them by stacking the tiled strip-band layers

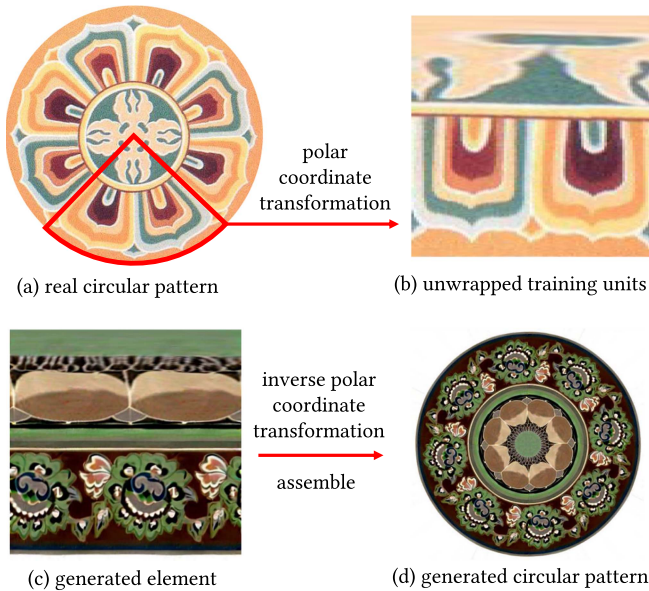


Fig. 8. Preprocess real circular patterns using polar coordinate transformation for model customization. During inference, we can generate the unwrapped elements, and apply inverse polar coordinate transformation to restore the full circular pattern.

from the outermost layer to the innermost layer and overlaying the reconstructed medallion at the center, following the layout templates from the training dataset. Users can interactively select the layout, choose the section of the layout that they wish to populate with a generated pattern, and adjust the pattern scale to flexibly personalize their design.

V. EXPERIMENTAL RESULTS

Implementation Details: We incorporate our Splice-and-Focus Mechanism and ReLife Pipeline into the Stable Diffusion XL base and refiner model, and customize the model with LoRA on a curated dataset that contains 60-80 samples. Training is performed on a single NVIDIA RTX 4090 GPU with a batch size of 2 for 10,000 steps. The LoRA scaling factor is set to 0.8, and we employ DDIM [30] sampling for 100 steps for image generation and evaluation.

For Periodic Inpainting, the outer boundary is set to $\frac{1}{16}$ of the image size, and the inner boundary is set to $\frac{1}{4}$ of the image size. For inpainting, we use SDXL-Refiner with 70 steps and a noise strength of 0.6.

For ReLife, we use Real-ESRGAN [42] as the super-resolution backbone. The high-resolution image is segmented into 4 parts for the 2048×2048 stage and 16 parts for the 4096×4096 stage. LoRA is trained on data under the corresponding resolution. The LoRA scaling factor is set to 1.0, and sampling is performed for 50 steps with a noise strength of 0.5 for SDEdit.

A. High-Resolution Periodic Pattern Generation

Fig. 9 illustrates the stitching effect of the periodic pattern generated by our method. The results qualitatively demonstrate that our approach can produce seamless periodic patterns while

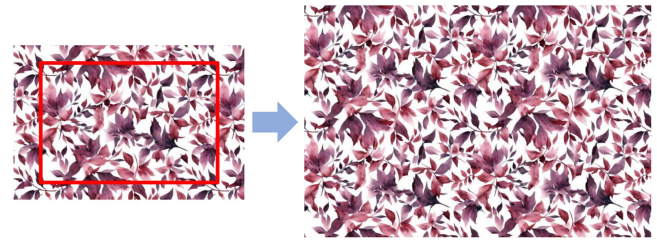


Fig. 9. In Splice-and-Focus mechanism, the periodic regions are spliced and extended. The central image is a seamless periodic pattern.

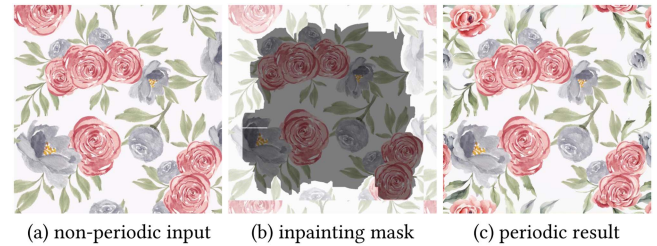


Fig. 10. Our Periodic Inpainting pipeline applied to nonperiodic images. Resulting patterns achieve strict periodicity while preserving original style.

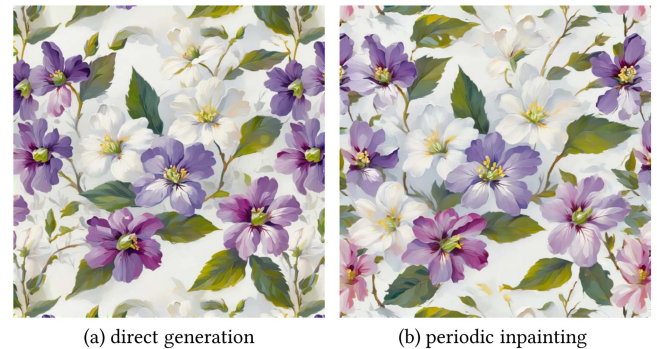


Fig. 11. Comparison between our direct generation pipeline and the Periodic Inpainting pipeline, with the same seed.

maintaining the intended style. Notably, our scheme not only supports the generation of square periodic patterns, but also can be applied to rectangles of arbitrary sizes.

Fig. 10 presents the intermediate and final results of our proposed Periodic Inpainting pipeline. Our method intelligently identifies the regions that require completion while strategically avoiding key pattern elements such as flowers and leaves. By integrating our Splice-and-Focus mechanism, the pipeline successfully inpaints the masked regions with periodic content that maintains both structural continuity and stylistic consistency with the unmasked areas. Fig. 11 presents a comparative analysis of our two modes of periodic pattern generation. Two modes achieve comparable results. We performed extended experiments on our periodic inpainting pipeline over watercolor, oil painting and fabric style. Each style contains 50 aperiodic floral patterns. On a single NVIDIA RTX 4090, the running time averaged over 150 patterns is 29 ± 0.4 s, exhibiting nearly consistent running time. This means that our solution enables

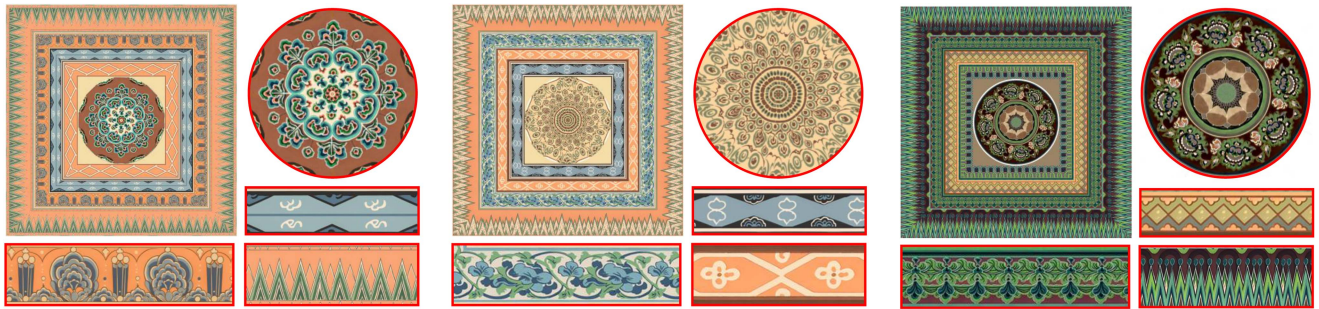


Fig. 12. Results of complex composite pattern generation. Local patterns are enlarged in red regions for better visualization.

the stable generation of a corresponding periodic pattern from a single image within half a minute.

B. Complex Compositional Pattern Generation

Fig. 12 shows the results of our pipeline on complex compositional pattern generation. The results demonstrate that our method effectively handles multilayer nonstationary composite pattern generation tasks. As shown, all outputs employ the classic Dunhuang caisson layout, with four nested square frames surrounding a central circular medallion motif, but exhibit different stylistic characteristics.

C. Comparison With State-of-The-Art Methods

To demonstrate the advantages of our approach in generating high-quality periodic patterns, we compare our model with four state-of-the-art text-driven image diffusion models, SDXL [28], FLUX.1,¹ DALL-E 3² and CrossFlow [6]. To assess the periodicity, we introduce the word “periodic” into the prompts for all the models. For each set of prompts, we generate 100 images per model for comparison.

To assess the generation of high-resolution patterns, we compare our model with five state-of-the-art super-resolution models, Real-ESRGAN [42], SwinIR [53], StableSR [44], SeeSR [45] and Tiled-SDXL [18]. All methods use the same 1024×1024 low-resolution patterns that are generated by our model as inputs to create 4096×4096 high-resolution patterns.

1) *Qualitative Analysis*: Fig. 13 presents qualitative results that illustrate the performance of each model under six prompt conditions. While all the models generate patterns that are aligned with the text prompts, only our model consistently produces patterns that strictly adhere to periodicity, thus exhibiting superior quality and artistic value. Additionally, our model excels in terms of the finer details of pattern generation and better style consistency.

Fig. 14 presents a visual comparison of high-resolution pattern generation between our method and other super-resolution baselines. The input low-resolution image in Fig. 14(a) has rigid edges that are inconsistent with the embroidery characteristics. The baselines focus mainly on stamen fidelity, thus their outputs still maintain rigid edges. Moreover, Tiled-SDXL [18] tends to

TABLE I
QUANTITATIVE ANALYSIS

Method	FID↓	TexTile↑	Quality↑	Periodicity↑
SDXL	213.62	0.73	0.016	0.038
FLUX.1	199.25	0.85	0.081	0.026
DALL-E 3	163.82	0.75	0.133	0.050
CrossFlow	201.37	0.30	0.065	0.000
Ours	99.27	0.98	0.705	0.886

introduce minor differences in the stitching areas, thereby resulting in obvious color discontinuities in the center of the stamen. Our method introduces realistic thread-like details and refines the stamen morphology without color discontinuities.

2) *Quantitative Analysis*: For quantitative evaluation, we use the Fréchet Inception Distance (FID) [54] to measure the quality and naturalness of the generated images. We applied 10 groups of prompt words, each group of prompt words contained 100 patterns, and used the average value of each group of FID calculation results as the experimental result. We also conduct a user study to evaluate the quality and periodicity of the results. Specifically, we asked users to select the result with the highest visual quality and the one with the best periodicity. We report the percentage of user selections in Table I over 39 users in 10 cases. We report TexTile score [55] in Table I to quantitatively evaluate the periodicity of the results. The high TexTile score shows that our method significantly outperforms the baseline models in generating images with strict periodicity. Our model consistently designs higher-quality and strictly periodic images.

D. Ablation Study

We conduct ablation studies on different components of our method to evaluate their impact on the generated results.

- *Latent Splice*: Fig. 15(a) presents the images generated using identical prompt words and noise seeds, with the omission of Latent Splice. Without this component, the resulting images fail to maintain the desired periodicity.
- *FocAttention*: As shown in Fig. 15(b)–(d), removing FocAttention leads to a significant drop in image quality, leading to mixed textures and biased colors. Although hard FocAttention performs well on SD 1.5 in our previous work [27], it fails on SDXL and generates messy over-repetitive textures. The proposed soft FocAttention

¹<https://github.com/black-forest-labs/flux>

²<https://cdn.openai.com/papers/dall-e-3.pdf>



Fig. 13. Qualitative comparisons with (a) SDXL [28], (b) FLUX.1, (c) DALL-E 3, (d) CrossFlow [6] and (e) the proposed method.

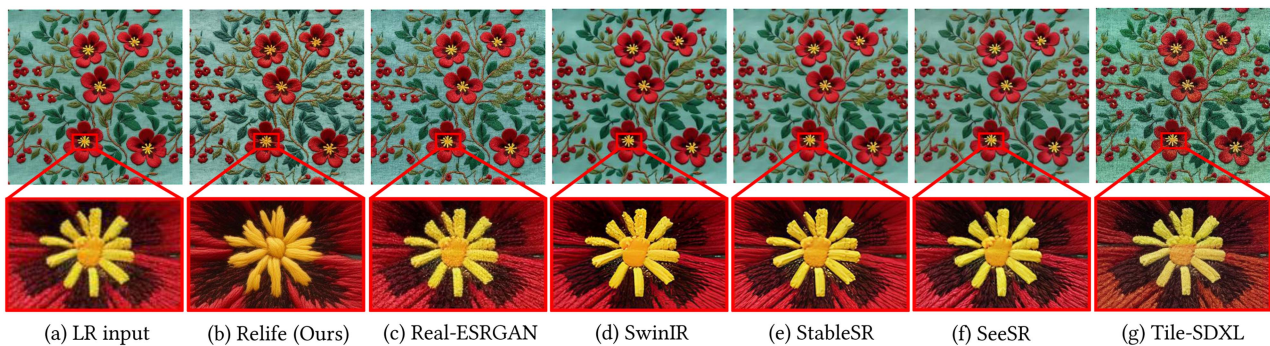


Fig. 14. Comparison with image super-resolution methods. Prompts: "embroidered red flowers, embroidered green leaves, linen weave cyan background."

performs well on SDXL, generating strictly periodic patterns that are the most consistent with the original pattern in Fig. 15(a). We also performed a quantitative analysis under various experimental settings to study the effect of different modules in our Splice-and-Focus mechanism. The results, summarized in Table II, indicate that Latent Splice is the key to achieve high periodicity. However, it alone or with hard FocAttention will degrade the image quality.

By combining Latent Splice and our newly proposed soft FocAttention, the image quality could be well preserved, striking a good balance between the periodicity and quality. In addition, we report the running time to examine the computational efficiency of our method (50 diffusion steps on an NVIDIA RTX 4090) in Table II. It can be seen that the Latent Splice and FocAttention are efficient, adding minimal computational overhead to the base model.



Fig. 15. Ablation study of Latent Splice and FocAttention.

TABLE II
QUANTITATIVE ANALYSIS OF ABLATION STUDY

Settings	TexTile $\uparrow$	Quality $\uparrow$	Time(s)
SD XL (Baseline)	0.73	0.480	26.16
SD XL + Latent Splice	0.96	0.023	29.33
SD XL + Latent Splice + hard FocAttention	0.82	0.000	29.41
SD XL + Latent Splice + soft FocAttention	0.98	0.497	31.20

- *ReLife*: ReLife is composed of super-resolution and generation stages. If removing the generation stages, ReLife degrades to Real-ESRGAN. In Fig. 14(b) and (c), we compare the super-resolution results of ReLife and Real-ESRGAN against the original low-resolution image in Fig. 14(a), to study the effect of the generation stages. It can be seen that both methods successfully generate images with higher resolution. However, compared to Fig. 14(c), the generation stage of ReLife effectively refines the stamen morphology in Fig. 14(b). Moreover, by imposing seamlessness and periodicity constraints in the generation stage, ReLife not only improves overall image quality but also preserves the image periodicity.
- *Compositional Generation*: In Figs. 12 and 16, we compare our compositional generation pipeline and a naive pipeline. If simply fine-tuning the SDXL base model using the whole complex patterns rather than decomposing the elements, the result shown in Fig. 16 demonstrates significant quality degradation. Specifically, the results exhibit blurred details, lack of periodicity, and low overall diversity. In contrast, our compositional generation effectively simplifies the generation, and our Splice-and-Focus ensures strict periodicity.

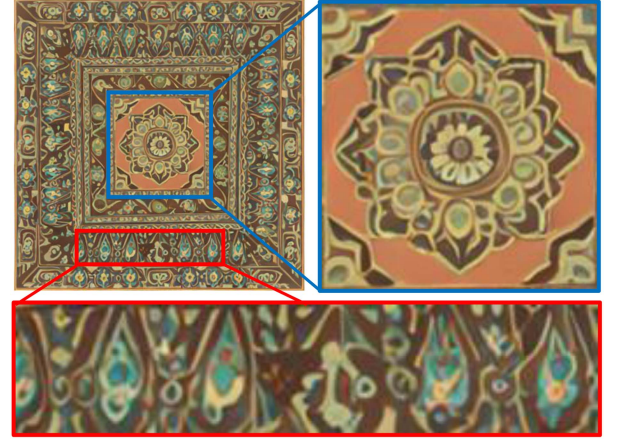


Fig. 16. The generated results obtained through simple fine-tuning using the entire complicated pattern lacks stable periodicity and inter-layer diversity, with blurred and chaotic details.

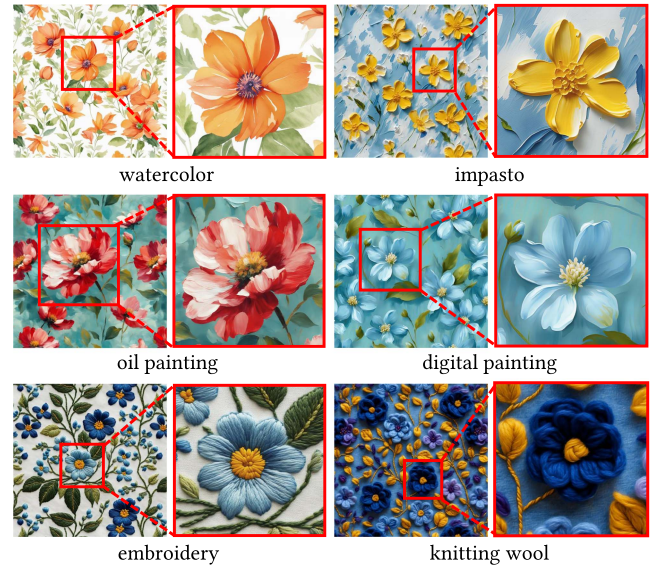


Fig. 17. Customized pattern generation.

E. More Results

Fig. 17 demonstrates the 4096×4096 patterns produced by our customized models for six distinct styles. Each generated pattern exhibits unique material properties and textural characteristics while maintaining strict periodicity and high-resolution quality. The visualization highlights our method's capability to adapt to diverse artistic styles.

Fig. 18 demonstrates the rendered application of our generated patterns in garment designs, showcasing our DiffTile's potential to generate production-ready patterns suitable for diverse fashion applications.

VI. CONCLUSION

In this paper, a novel DiffTile framework for generating high-resolution periodic patterns tailored to the printing and



Fig. 18. Rendered application of our generated patterns on garment designs.

dyeing industry is presented. We propose a novel training-free splice-and-focus mechanism to ensure strict periodicity and seamless pattern generation when integrated with LDMs. We incorporate a foreground segmentation module that enables the transformation of nonperiodic patterns into patterns with strict periodicity. Additionally, a ReLife pipeline is proposed to increase image resolution while effectively correcting artifacts. Finally, we design a compositional pattern generation pipeline for complicated nonstationary pattern design. The experimental results demonstrate that our method outperforms existing techniques and shows potential for generating production-ready patterns that are suitable for diverse fashion applications.

REFERENCES

- [1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10674–10685.
- [2] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proc. Int. Conf. Comput. Vis.*, 2023, pp. 4195–4205.
- [3] P. Esser et al., "Scaling rectified flow transformers for high-resolution image synthesis," in *Proc. IEEE Int. Conf. Mach. Learn.*, 2024, pp. 1–12.
- [4] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," 2022, *arXiv:2204.06125*.
- [5] C. Saharia et al., "Photorealistic text-to-image diffusion models with deep language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 36479–36494.
- [6] Q. Liu, X. Yin, A. Yuille, A. Brown, and M. Singh, "Flowing from words to pixels: A noise-free framework for cross-modality evolution," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 2755–2765.
- [7] N. Ruiz et al., "DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 22500–22510.
- [8] N. Kumari, B. Zhang, R. Zhang, E. Shechtman, and J.-Y. Zhu, "Multi-concept customization of text-to-image diffusion," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 1931–1941.
- [9] L. Han et al., "SVDiff: Compact parameter space for diffusion fine-tuning," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7289–7300.
- [10] E. J. Hu et al., "LoRA: Low-rank adaptation of large language models," in *Proc. Int. Conf. Learn. Representations*, 2022, pp. 1–13.
- [11] M. N. Everaert, M. Bocchio, S. Arpa, S. Süsstrunk, and R. Achanta, "Diffusion in style," in *Proc. Int. Conf. Comput. Vis.*, 2023, pp. 2251–2261.
- [12] Y. Xu, X. Xu, H. Gao, and F. Xiao, "SGDM: An adaptive style-guided diffusion model for personalized text to image generation," *IEEE Trans. Multimedia*, vol. 26, pp. 9804–9813, 2024.
- [13] Y. Jiang, Q. Liu, D. Chen, L. Yuan, and Y. Fu, "AnimeDiff: Customized image generation of anime characters using diffusion model," *IEEE Trans. Multimedia*, vol. 26, pp. 10559–10572, 2024.
- [14] H. Zhang, T. Wu, and Y. Wei, "Multi-view user preference modeling for personalized text-to-image generation," *IEEE Trans. Multimedia*, vol. 27, pp. 3082–3091, 2025.
- [15] M. Liu et al., "TFCustom: Customized image generation with time-aware frequency feature guidance," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 2714–2723.
- [16] J. Jiang et al., "MC²: Multi-concept guidance for customized multi-concept generation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 2802–2812.
- [17] J. Ho and T. Salimans, "Classifier-free diffusion guidance," in *Proc. NeurIPS Workshop Deep Generative Models Downstream Appl.*, 2021, pp. 1–8.
- [18] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proc. Int. Conf. Comput. Vis.*, 2023, pp. 3836–3847.
- [19] Y. Li et al., "GLIGEN: Open-set grounded text-to-image generation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 22511–22521.
- [20] Q. Phung, S. Ge, and J.-B. Huang, "Grounded text-to-image synthesis with attention refocusing," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 7932–7942.
- [21] C. Mou et al., "T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, no. 5, pp. 4296–4304.
- [22] S. J. Liu, N. Kumari, A. Shamir, and J.-Y. Zhu, "Generative photomontage," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 7931–7941.
- [23] S. Xiao et al., "OmniGen: Unified image generation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 13294–13304.
- [24] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [25] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.
- [26] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 23593–23606.
- [27] X. Lan, W. Gao, L. Zhang, J. Liu, and S. Yang, "Splice, focus and relife: High-resolution periodic pattern generation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2025, pp. 1–4.
- [28] D. Podell et al., "SDXL: Improving latent diffusion models for high-resolution image synthesis," in *Proc. Int. Conf. Learn. Representations*, 2024, pp. 1–13.
- [29] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 6840–6851.
- [30] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–12.
- [31] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. IEEE Int. Conf. Mach. Learn.*, 2021, pp. 8162–8171.
- [32] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. IEEE Int. Conf. Mach. Learn.*, 2015, pp. 2256–2265.
- [33] A. Witkin and M. Kass, "Reaction-diffusion textures," in *Proc. Conf. Comput. Graph. Interactive Techn.*, 1991, pp. 299–308.
- [34] Y. Wang, A. Holynski, B. L. Curless, and S. M. Seitz, "Infinite texture: Text-guided high resolution diffusion texture synthesis," 2024, *arXiv:2405.08210*.
- [35] O. Madar and O. Fried, "Tiled diffusion," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 7795–7804.
- [36] A. Sinha, S. Jia, and V. Fung, "Representation-space diffusion models for generating periodic materials," 2024, *arXiv:2408.07213*.
- [37] H. F. Karagoz, G. Baykal, I. A. Eksi, and G. Unal, "Textile pattern generation using diffusion models," 2023, *arXiv:2304.00520*.
- [38] H. Li et al., "SRDiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.
- [39] C. Saharia et al., "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023.
- [40] Z. Yang et al., "Inf-Dit: Upsampling any-resolution image with memory-efficient diffusion transformer," in *Proc. Eur. Conf. Comput. Vis.*, 2024, pp. 141–156.
- [41] H. Sahak, D. Watson, C. Saharia, and D. Fleet, "Denoising diffusion probabilistic models for robust image super-resolution in the wild," 2023, *arXiv:2302.07864*.

- [42] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2021, pp. 1905–1914.
- [43] M. Xu, J. Ma, and Y. Zhu, "Dual-diffusion: Dual conditional denoising diffusion probabilistic models for blind super-resolution reconstruction in RSIs," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6008505.
- [44] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy, "Exploiting diffusion prior for real-world image super-resolution," *Int. J. Comput. Vis.*, vol. 132, no. 12, pp. 5929–5949, 2024.
- [45] R. Wu et al., "SeeSR towards semantics-aware real-world image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 25456–25467.
- [46] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 12873–12883.
- [47] N. Ravi et al., "SAM 2: Segment anything in images and videos," in *Proc. Int. Conf. Learn. Representations*, 2025, pp. 1–17.
- [48] A. Lugmayr et al., "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11461–11471.
- [49] O. Avrahami, O. Fried, and D. Lischinski, "Blended latent diffusion," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 1–11, 2023.
- [50] B. Xia et al., "DiffIR: Efficient diffusion model for image restoration," in *Proc. Proc. Int. Conf. Comput. Vis.*, 2023, pp. 13095–13105.
- [51] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–16.
- [52] C. Meng et al., "SDEdit: Guided image synthesis and editing with stochastic differential equations," in *Proc. Int. Conf. Learn. Representations*, 2022, pp. 1–14.
- [53] J. Liang et al., "SwinIR: Image restoration using swin transformer," in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 1833–1844.
- [54] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–12.
- [55] C. Rodriguez-Pardo, D. Casas, E. Garces, and J. Lopez-Moreno, "Textile: A differentiable metric for texture tileability," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 4439–4449.



Xicheng Lan received the B.S. degree in computer science from Peking University, Beijing, China, in 2025. He is currently working toward the Ph.D. degree with the Wangxuan Institute of Computer Technology, Peking University. His research interests include image generation and stylization.



Chenyu Niu is currently working toward the B.S. degree in computer science and technology from Yuanpei College, Peking University, Beijing, China. He is a Research Intern with the Wangxuan Institute of Computer Technology, Peking University. His research interests include image restoration and generation.



Shenghan Xu is currently working toward the B.S. degree in artificial intelligence from Yuanpei College, Peking University, Beijing, China. He is a Research Intern with the Wangxuan Institute of Computer Technology, Peking University. His research interests include image generation and image quality assessment.



Wenshuo Gao received the B.S. degree in computer science from Peking University, Beijing, China, in 2025. He is currently working toward the Ph.D. degree with the Wangxuan Institute of Computer Technology, Peking University. His research interests include image/video generation and editing.



Luyao Zhang received the B.S. and M.S. degrees in integrated microelectronics from Peking University, Beijing, China, in 2015 and 2018, respectively. She is currently working toward the Ph.D. degree with the Wangxuan Institute of Computer Technology, Peking University. Her research interests include image generation and image restoration.



Jiaying Liu (Fellow, IEEE) received the Ph.D. degree (Hons.) in computer science from Peking University, Beijing, China, 2010. She was a Visiting Scholar with the University of Southern California, Los Angeles, CA, USA, from 2007 to 2008. She was a Visiting Researcher with Microsoft Research Asia, in 2015, supported by the Star Track Young Faculties Award. She is currently a Professor and Boya Young Fellow with the Wangxuan Institute of Computer Technology, Peking University. She has authored more than 100 technical articles in refereed journals and proceedings, and holds 70 granted patents. Her research interests include multimedia signal processing, compression, and computer vision. She is a senior member of CSIG and a distinguished member of CCF. Dr. Liu was a member of Multimedia Systems and Applications Technical Committee, and Visual Signal Processing and Communications Technical Committee in IEEE Circuits and Systems Society. She was the recipient of IEEE ICME 2020 Best Paper Award and IEEE MMSP 2015 Top10% Paper Award. She was an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Journal of Visual Communication and Image Representation*, the Technical Program Chair of ACM MM Asia-2023/IEEE ICME-2021/ACM ICMR-2021/IEEE VCIP-2019, the Area Chair of CVPR-2021/ECCV-2020/ICCV-2019, ACM ICMR Steering Committee member, and the CAS Representative at the ICME Steering Committee. She was the APSIPA Distinguished Lecturer from 2016 to 2017.



Shuai Yang (Member, IEEE) received the B.S. and Ph.D. degrees (Hons.) in computer science from Peking University, Beijing, China, in 2015 and 2020, respectively. He was a Visiting Student with the National Institute of Informatics, Tokyo, Japan, from March 2017 to August 2017. He was a Visiting Scholar with Texas A&M University, College Station, TX, USA, from 2018 to 2019. He was a Postdoctoral Research Fellow with Nanyang Technological University, from 2020 to 2023. He was a Research Assistant Professor with S-Lab, Nanyang Technological University, Singapore, from 2023 to 2024. He is currently an Assistant Professor with the Wangxuan Institute of Computer Technology, Peking University. His research interests include image stylization, image translation, and image editing. He was the recipient of IEEE ICME 2020 Best Paper Awards and IEEE MMSP 2015 Top 10% Paper Awards. He was the Area Chair of NeurIPS-2025, ACM MM-2024/25, and BMVC-2023/24/25.