Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

Functional Textile Pattern Generation Based on Visual Semantics and Aesthetics Using Diffusion Models

Yingying Xu, Minwei Zhao, Hao Nong

How to cite: Xu Y, Zhao M, Nong H. Functional Textile Pattern Generation Based on Visual Semantics and Aesthetics Using Diffusion Models. Textile & Leather Review. 2026; 9:1-18. https://doi.org/10.31881/TLR.2026.001

How to link: https://doi.org/10.31881/TLR.2026.001

Published: 16 January 2026

Functional Textile Pattern Generation Based on Visual Semantics and Aesthetics Using Diffusion Models

Yingying XU[1,2], Minwei ZHAO[1,3*], Hao NONG[4]

[1]School of Humanities, Shanghai Normal University, 100 Guilin Road, Xuhui District, Shanghai 200234, China

[2]College of Preparatory Education, Guangxi Minzu University, 188 Daxue East Road, Xixiangtang District, Nanning 530006, Guangxi, China

[3]School of Liberal Arts, Guangxi Minzu University, 188 Daxue East Road, Xixiangtang District, Nanning 530006, Guangxi, China

[4]Nanning Teacher Training Center, 4-1 Minle Road, Qingxiu District, Nanning 530012, Guangxi, China

*xyyangel14@126.com, the corresponding address is Affiliation 1.

Article

https://doi.org/10.31881/TLR.2026.001

Received 6 August 2025; Accepted 22 August 2025; Published 16 January 2026

ABSTRACT

This paper presents a novel method for generating functional textile patterns by introducing a dual-guidance diffusion model that disentangles functional control from stylistic description. We achieve this by translating abstract functional requirements into quantifiable semantic vectors, which are injected into the generation process via a dedicated encoder and cross-attention mechanism, enabling precise control over functional visual features. To this end, we first built a textile pattern dataset that links functional attributes such as "moisture-wicking" to corresponding visual elements such as channel-like textures. We then developed a Functional-Semantic Guided Diffusion Model (FS-DM). This model incorporates a specialized encoder into the LDM's U-Net, using a cross-attention mechanism to inject the desired functional semantics into the pattern generation process. This design enables fine-grained control over the final pattern's structure and appearance. The generated patterns were assessed using a multidimensional evaluation framework. Experimental results show that the method successfully creates patterns with clear visual guidance based on predefined functions. These patterns achieved a functional visual semantic relevance score 35.7% higher than baseline models and also scored exceptionally well on computational aesthetic metrics and in subjective user evaluations. This research offers a new, intelligent, and efficient paradigm for functional textile design with significant industrial application value.

KEYWORDS

diffusion model, functional textile, pattern design, visual semantics, deep learning

INTRODUCTION

Functional textiles represent a significant development direction in modern textile technology. By using special materials, structures, or post-treatment techniques, these approaches endow fabrics with specific functions such as moisture-wicking, thermal retention, antibacterial properties, and UV resistance, and are widely applied in sports and outdoor activities, healthcare, and protective applications [1]. Currently, the research and development focus of functional textiles mainly lies in breakthroughs in material science and weaving technology, while the design of their appearance patterns is often regarded as a secondary aspect or simply adopts traditional decorative patterns, resulting in a disconnection between the functionality and visual presentation of the products. As the most direct visual language of textiles, patterns that can ingeniously convey the inherent functions would substantially enhance the added value and market competitiveness of the products.

Traditional pattern design for functional textiles relies heavily on designers' personal experience and inspiration. The core challenge is the ambiguity of functional expression. Designers often struggle to translate abstract concepts (e.g., breathability) into concrete visual patterns, weakening the link between the design and functional requirements. Moreover, this design approach leads to inefficient innovation. It is not only time-consuming and labor-intensive with a long iteration cycle, but also tends to trap designers in conventional thinking patterns, making it hard to produce breakthrough works. Additionally, the entire design and evaluation process is highly subjective. There is a lack of objective and quantifiable standards when judging the aesthetic and functional implications of patterns, making it difficult to determine the superiority or inferiority of the schemes.

In recent years, deep generative models such as diffusion models [2] have made revolutionary progress in the field of image generation, achieving unprecedented levels in the quality, diversity, and controllability of the generated images. In the fashion and textile fields, researchers have attempted to use generative adversarial networks (GANs) [3] or diffusion models [4] to generate novel textile patterns. However, most of the existing research focuses solely on aesthetic creation or style transfer, and rarely addresses how to integrate the functional requirements of textiles into the generation process, that is, how to achieve purposeful design.

To bridge the gap between functional requirements and the visual expression of patterns, this paper proposes a novel method for generating functional textile pattern visual semantics and aesthetics by combining a diffusion model. The core contribution of this research is threefold: First, we formalize the ambiguous relationship between textile function and appearance by constructing a function–visual semantic library where abstract concepts (e.g., moisture-wicking) are decomposed into a quantifiable, vector-based representation of visual attributes (e.g., "channel", "streamline"). This structured approach moves beyond ambiguous natural language prompts. Second, we propose a novel Functional-Semantic Guided Diffusion Model (FS-DM). Our key architectural innovation is a dual-branch conditioning mechanism that processes general text descriptions and our specific functional vectors separately, allowing for disentangled control over aesthetics and function during the pattern generation process. Third, this research also establishes a scientific quantitative evaluation system, which integrates functional visual semantic relevance, computational aesthetic indicators, and user perception, providing a scientific basis for objectively and subjectively evaluating the comprehensive performance of the generated patterns. This research aims to explore a path for deeply integrating artificial intelligence technology into the design process of functional textiles, enabling designers to efficiently create innovative patterns that not only conform to engineering functional principles but also meet modern aesthetic demands, and promoting the intelligent transformation and upgrading of the functional textile industry.

RELEVANT WORK

Design Principles of Functional Textiles

The design of functional textiles is not only about the selection of materials; the fabric structure and surface patterns also play significant roles. For instance, in moisture-wicking textiles, by designing a groove structure with capillary effects or differential weaving between the inner and outer layers, the one-way transport of sweat can be accelerated [5]. Visually, this structure often presents regular channels or network-like textures. Similarly, in thermal insulation textiles, by forming a stationary air layer to prevent heat loss, their internal structure is usually fluffy and dense, which may manifest as visual features such as fluffiness or enclosure effect when expressed in patterns. Existing studies have shown that the visual patterns of textiles can significantly influence consumers' expectations of touch and emotional responses [6]. Therefore, extracting these functionally related structural features artistically and transforming them into the visual language of patterns is the core principle of functional pattern design.

Application of Artificial Intelligence in Textile Pattern Generation

In early work, researchers used rule-based or fractal algorithms to generate patterns with mathematical beauty [7,8]. Recently, deep learning, especially Generative Adversarial Networks (GANs), has been widely used in textile pattern generation. For instance, StyleGAN [9] is used to generate high-quality and diverse textile textures. However, the instability and mode collapse issues in GAN training limit their application.

The emergence of diffusion models provides a new solution for high-quality and highly controllable image generation. Diffusion models recover clear images from pure Gaussian noise through a gradual denoising process. By introducing conditional information (such as text descriptions) during the denoising process, the generated content can be precisely controlled. Existing research [4] uses text prompts to guide the diffusion model to generate textile patterns of specific styles, such as "Persian patterns" or "Baroque style". However, these methods rely on the model's understanding of existing artistic styles and cannot directly control the more fundamental visual elements related to functionality, such as the "streamline characteristics" or "porosity" of the patterns.

Semantic Control in Generative Models

To achieve fine control over the generated content, researchers have proposed various semantic injection methods. In diffusion models, the cross-attention mechanism is a key technique for implementing text-conditioned control [10]. By encoding the text prompt as a vector and performing attention calculations with image features at multiple levels of the U-Net, the model can align the text semantics with specific regions and features of the image. Additionally, there are studies that explore injecting custom concepts by modifying model weights and guiding the sampling process [11].

This study draws on the idea of cross-attention, but extends it from the general textual description to more specialized functional semantics, proposing a method of injecting structured functional vectors into the model to achieve more precise and more engineering-relevant guidance for pattern generation. However, relying solely on general textual descriptions, as is common in existing text-to-image models, presents a key limitation for functional design. Natural language can be ambiguous and fails to provide the granular, quantifiable control needed to enforce specific engineering-related visual structures (e.g., controlling the porosity level). Our work directly addresses this gap by moving from an unstructured text prompt to a structured functional vector, enabling a more precise and interpretable form of semantic guidance.

RESEARCH METHODS

The method flow proposed in this study is shown in Figure 1. It mainly consists of three core parts: the construction of a function–visual semantic dataset, the design of the Functional-Semantic Guided Diffusion Model (FS-DM), and the establishment of a multidimensional pattern evaluation system.

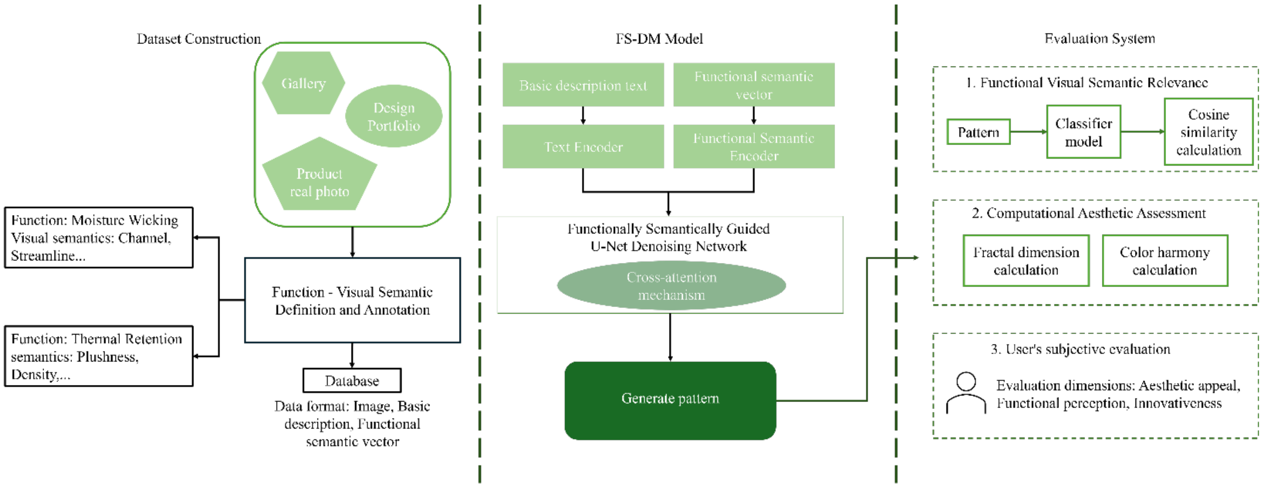

Figure 1. Framework for the generation and evaluation of functional textile patterns

Function–Visual Semantic Dataset Construction

To enable the model to understand and learn the correlation between functionality and visual features, we constructed a dataset named FuncTex-Pattern. The construction process of the dataset is as follows:

1. Data collection: Approximately 20,000 high-quality textile pattern images were collected from multiple sources, including commercial textile image libraries, design portfolio websites, and photographs of existing functional garments/products.

2. Functional semantic definition: We defined three core functional categories and decomposed them into a set of quantifiable visual semantic keywords. The selection of these keywords was based on textile engineering principles and consumer perception research [5,6].

- **Moisture-wicking:** Aimed at quickly transferring sweat from the skin surface to the outer layer of the fabric. Its visual semantic keywords include channel, streamline, gradient, network, directionality.
- **Thermal retention:** Aimed at reducing heat loss. Its visual semantic keywords include: fluffy, dense, enveloping, layered, soft.
- **Breathability:** Aimed at allowing air and water vapor to pass through. The selection of five keywords for each functional category was based on a combination of domain expert consultation and a pilot study.

Initially, a larger set of potential keywords was compiled from textile engineering literature and consumer reports. Through iterative discussions with our expert panel, this list was refined to the five most distinct and visually representable terms for each function, ensuring minimal conceptual overlap. This five-dimensional structure was deemed optimal as it provided sufficient granularity to describe the core visual features without introducing excessive complexity or redundancy into the annotation and modeling process.

3. Data annotation: We recruited 10 experts with backgrounds in textile design and materials science to annotate each image in the dataset. The annotation information includes:

- Base description: Objective description text of the pattern style, elements, and colors (e.g., "a blue abstract curvilinear pattern").
- Functional semantic vector: A five-dimensional floating-point vector corresponding to five visual semantic keywords under a certain functional category. The values are in the range [0, 1], representing the degree of conformity of the pattern to the corresponding visual semantic. For example, a typical moisture-wicking pattern may have higher scores in the Channel and Streamline dimensions, while the score in the Fluffy dimension may be close to 0. To anchor the annotation process, we established a clear rubric for the 0-to-1 scale: a score of 0 indicates the complete absence of the visual semantic; a score around 0.5 signifies that the feature is present but not dominant or is only partially expressed; and a score of 1.0 represents a clear, strong, and dominant manifestation of the visual feature throughout the pattern.

Ultimately, each piece of data takes the form of (image, base description, functional semantic vector), providing structured input for the subsequent training of the model. To ensure the consistency and reliability of the annotations, we implemented a rigorous multi-step process. Initially, all 10 experts participated in a calibration session where they collaboratively annotated a pilot batch of 100 images to establish a shared understanding of the rating criteria for each semantic keyword. Following this, the main dataset was divided and annotated independently by at least three experts per image.

We then calculated the inter-rater reliability (IRR) using the intraclass correlation coefficient (ICC), which is well-suited for continuous data rated by multiple raters. Our analysis yielded an ICC score of 0.82, indicating a high level of agreement among the experts. For images where there were significant rating discrepancies (e.g., a standard deviation greater than 0.2 for any single semantic value), the involved experts held a consensus meeting to discuss their reasoning and arrive at a final, agreed-upon vector. This meticulous process ensures that the functional semantic vector in our FuncTex-Pattern dataset is a robust and reliable representation of the visual semantics.

Functional-Semantic Guided Diffusion Model (FS-DM)

We improved upon a pre-trained latent diffusion model (Latent Diffusion Model, LDM) [10] and proposed the FS-DM. The core idea is to add an additional functional semantic guidance branch on top of the standard text guidance, as shown in Figure 2. While a standard LDM can be guided by text prompts, we found that simply including functional keywords (e.g., "a moisture-wicking pattern") in the prompt yields inconsistent and generalized results, as the model lacks a deep, structured understanding of the underlying visual principles. To solve this, our FS-DM introduces a parallel guidance branch specifically for functional semantics, ensuring that these critical attributes are explicitly represented and controlled.

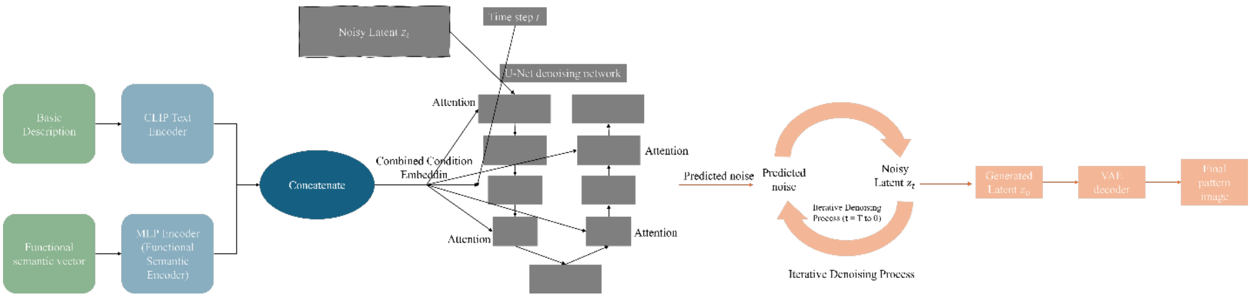

Figure 2. Schematic of the proposed FS-DM architecture. The model uses a dual-branch conditioning mechanism where text and functional semantic embeddings are concatenated to guide the U-Net. The U-Net operates in a latent space, iteratively denoising a noisy latent (zt) over multiple timesteps to produce a final denoised latent (z_0), which is then decoded by a VAE into the final pattern image.

Model structure:

Image encoder/decoder: Utilizing a pre-trained variational autoencoder (VAE), it is responsible for compressing the high-dimensional pixel space of the image into a low-dimensional latent space, and reconstructing the latent representation into a pixel image after the generation process is completed. This significantly reduces the computational complexity of the diffusion process.

- Text encoder: Utilizes a pre-trained CLIP text encoder to convert the basic descriptive text into text embedding vectors c_{text}.
- Functional semantic encoder: This is the key module we have designed. It is a small multi-layer

perceptron (MLP) that takes a five-dimensional functional semantic vector v_{func} as input and maps it to an embedding c_{text} with the same dimension as the text embedding c_{func}.

$$c_{func} = MLP(v_{func}) \tag{1}$$

We acknowledge that expanding a five-dimensional vector (v_{func}) into a high-dimensional embedding (c_{func}) warrants clarification. The rationale behind this design is to project the sparse, human-defined functional semantics into the same rich latent space occupied by the CLIP text embeddings (c_{text}). This allows the functional guidance to interact with the model's pre-existing, nuanced understanding of visual concepts within the U-Net's cross-attention layers.

The multi-layer perceptron (MLP) is not merely upscaling the dimension; it is a learned, non-linear transformation. During fine-tuning, the MLP learns to map specific combinations of the 5 input values to specific, complex activation patterns in the high-dimensional space. For example, a high value in the Channel dimension and Directionality dimension of v_{func} will be mapped by the trained MLP to a unique high-dimensional vector in c_{func}. This vector, when processed by the cross-attention mechanism, activates the corresponding visual features (e.g., parallel lines, flow) that the base model already understands from its extensive pre-training. This process effectively translates the low-dimensional functional instructions into a high-dimensional 'language' that the U-Net can interpret to guide the denoising process, ensuring that the functional control is both powerful and synergistic with the text-based style guidance.

- U-Net denoising network: This is the core of the diffusion model, responsible for predicting the noise to be added to the latent representation z_t at each time step *t*. We made modifications to each cross-attention module in U-Net. The original cross-attention only received the text embedding c_{text} as the condition. In FS-DM, we concatenated (combined) the text embedding c_{text} and the functional semantic embedding c_{func} to form a combined conditional embedding c_{concat}.

$$c_{concat} = Concat(c_{text}, c_{func}) \tag{2}$$

To ensure stable training and effective fusion of the two guidance signals, several steps are taken. First, the functional-semantic encoder (MLP) is specifically designed to output a functional semantic embedding (c_{func}) that has the exact same dimension as the text embedding (c_{text}) from the CLIP encoder, thus resolving any dimensional mismatch prior to concatenation. Second, both embedding vectors are subjected to L2 normalization before concatenation to ensure they contribute on a similar scale, preventing either the stylistic or functional guidance from overpowering the other. No explicit weight balancing is applied to the concatenated vector itself; instead, the model learns to balance the influences of the two information streams implicitly through the fine-tuning of the cross-attention layers on our FuncTex-Pattern dataset. Then, the c_{concat} is input into the cross-attention layer, enabling it to simultaneously focus on the style description and functional visual features of the pattern during the denoising process.

$$Attention(Q, L, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right) \ where\ K, V\ are\ derived\ from\ c_{concat} \tag{3}$$

Training and inference: During the training phase, we fine-tune the pre-trained LDM using the FuncTex-Pattern dataset. The loss function is the same as that of the standard LDM, which is the mean squared error between the predicted noise and the actual added noise. In this way, the model learns to associate specific functional semantic vectors with the corresponding pattern structure features in the latent space.

In the inference (generation) phase, users can input a basic description (such as "black and white geometric patterns") and a functional semantic vector (such as the vector representing "strong moisture-wicking" [0.9, 0.8, 0.3, 0.6, 0.7]). FS-DM will generate a textile pattern that not only conforms to the text description but also embodies the specified functional visual features based on these two inputs.

Multidimensional Pattern Evaluation System

To conduct a comprehensive and objective assessment of the quality of the generated patterns, we designed an evaluation system consisting of three parts.

- Functional visual semantic relevance (FVSR): This metric is designed to quantify the degree of matching between the generated pattern and the target functional semantics. We trained a multi-label image classifier (based on ResNet-50) that can take a pattern as input and output its corresponding functional semantic vector. To ensure the objectivity of this evaluation and mitigate the risk of data leakage between the generator and the evaluator, we implemented a strict dataset partitioning strategy. The FuncTex-Pattern dataset was split into a training set (80%) and a held-out test set (20%). The FS-DM generator and all baseline models were trained or fine-tuned *exclusively* on the 80% training set. The FVSR classifier, in turn, was trained on this same 80% training set but was validated on a separate 10% validation split taken from the training data. Crucially, all final FVSR scores reported in this paper were calculated using patterns evaluated by the classifier after it had been finalized, and this evaluation was performed on generated images conditioned by prompts and vectors that the classifier had not been trained on. While we acknowledge that both models learn from the same data distribution, this separation of the test set for the generator's evaluation provides a degree of independence. The high correlation between these objective FVSR scores and the subjective user evaluations further validates this metric as a meaningful indicator of functional-semantic alignment. The FVSR score is obtained by calculating the cosine similarity between the semantic vector predicted by this classifier for the generated pattern and the input target semantic vector. The higher the score, the more the visual features of the pattern conform to the expected functional orientation.

$$FVSR = \cos\left(v_{target}, v_{predicted}\right) = \frac{v_{target} \cdot v_{predicted}}{\|v_{target}\| \cdot \|v_{predicted}\|} \tag{4}$$

- Computational aesthetic assessment (CAA): To quantitatively evaluate the aesthetic quality of the generated patterns, we selected two well-established computational metrics that capture fundamental aspects of visual appeal in textile design: fractal dimension and color harmony.
- Fractal dimension: We chose this metric because of its proven correlation with a pattern's perceived visual complexity, a key factor that contributes to its aesthetic richness and detail. Calculated using the box-counting method, a higher fractal dimension often corresponds to a greater level of detail and intricacy, which are key indicators of aesthetic interest in textile surfaces.

Color harmony: As color is a primary component of aesthetic experience, we selected this metric to evaluate the pleasantness of the visual palette. By analyzing the color composition against established color theory models, we can objectively score the coherence of the color scheme. Specifically, our algorithm is based on a computational model of color harmony theory. For each generated pattern, the top five dominant colors are first extracted using k-means clustering. These colors are then converted from the standard sRGB color gamut to the perceptually uniform CIELAB color space, assuming a D65 standard illuminant as the white point. The harmony score is calculated based on the geometric relationships (differences in hue angle, chroma, and lightness) between these dominant colors. The final score is a weighted sum that rewards combinations conforming to classical harmony templates such as analogous, complementary, or triadic schemes, adapted from the color theory models of Moon and Spencer. This method provides a quantifiable assessment of the palette's aesthetic coherence. While other metrics such as symmetry, visual balance, or specific complexity measures could also contribute to a comprehensive aesthetic analysis, we prioritized fractal dimension and color harmony as they represent two of the most critical and distinct axes of textile aesthetics: structural complexity and chromatic appeal. This focused approach provides a robust and interpretable, though not exhaustive, quantitative assessment of the patterns' aesthetic qualities.

User study: We conducted a user study with 30 participants, divided into two groups: 15 ordinary consumers and 15 textile professionals (including designers, material engineers, and product managers).

- Participant recruitment and demographics: Participants were recruited through university mailing lists and professional networking platforms. The consumer group had a balanced gender distribution (8 female, 7 male) and an age range of 22–45 years (average 29.6). The professional group also had a balanced gender distribution (7 female, 8 male), an age range of 26–51 years (average 35.2), and an average of 8.4 years of industry experience. All participants provided informed consent before the study.

Evaluation protocol: To mitigate potential biases, the study was conducted in a double-blind manner. The generated patterns from all models were presented in a randomized order, and participants were not informed which model generated which pattern. For each pattern, participants were first asked to rate its aesthetic appeal and novelty on a five-point Likert scale. Subsequently, they were shown the target function (e.g., moisture-wicking) and asked to rate the pattern's functional perception on the same scale. To ensure consistency and eliminate variability, this target function was presented as a standardized, concise text label (e.g., "moisture-wicking", "thermal retention", or "breathability") without any additional explanatory text or imagery. This two-step process prevented the functional description from influencing their initial aesthetic judgment.

EXPERIMENTS AND RESULTS

Experimental Setup

We compare the FS-DM with the following baseline models:

Text-to-Image LDM (T2I-LDM): A standard LDM that is guided solely by text prompts. To simulate our approach as closely as possible, we directly incorporate functional semantic keywords into the text prompts (for example, "a black-and-white geometric pattern with a sense of channel and streamline").

StyleGAN2-ADA: A mainstream GAN model, trained using the FuncTex-Pattern dataset and capable of unconditional generation.

All models were trained or fine-tuned on the same FuncTex-Pattern dataset. For the sake of a focused and clear comparative analysis, this paper presents the detailed experimental results for the moisture-wicking and breathability categories. These two categories were selected as they represent distinct and highly characteristic visual-semantic challenges, sufficiently demonstrating the core capabilities and advantages of our proposed FS-DM. The resolution of the generated patterns is 512 × 512 px.

Qualitative Result Analysis

Figure 3 shows the comparison of the patterns generated by FS-DM and T2I-LDM under the same text prompt "abstract blue motion-style pattern" and different functional semantic vectors.

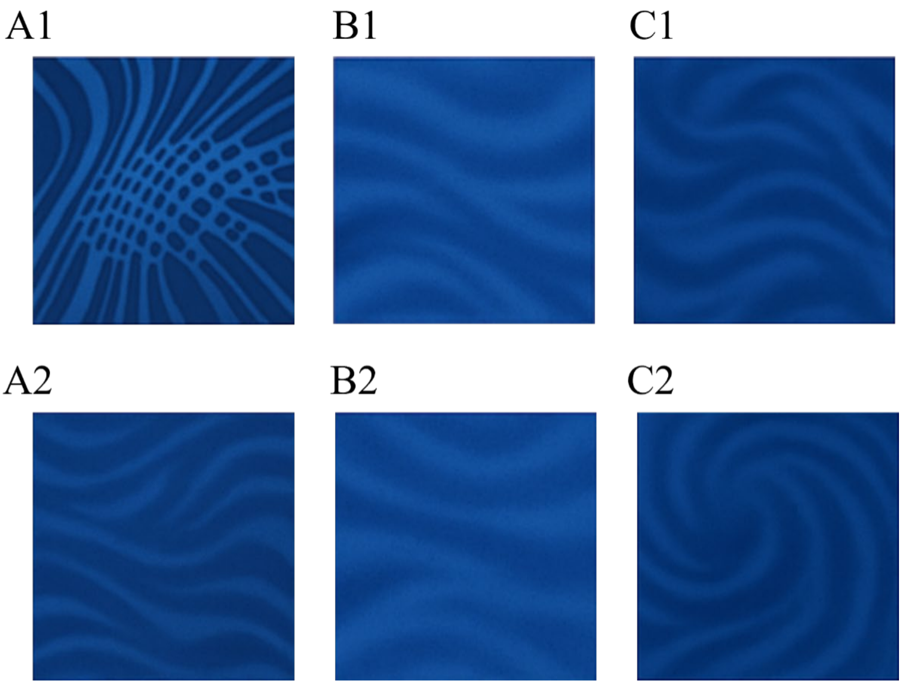

Figure 3. Comparison of patterns generated by different models under functional guidance. For a fair comparison between models under the same prompt, all images were generated using an identical initial random seed. (A1) FS-DM: moisture-wicking; (A2) FS-DM: breathability; (B1) T2I-LDM: moisture wicking; (B2) T2I-LDM: breathability; (C1 and C2) StyleGAN2-ADA: unconditional.

As shown in Figure 3, the FS-DM demonstrates outstanding precision control. Under the moisture-wicking condition, the pattern (A1) clearly presents directional streamlines and channel structures, visually suggesting the process of rapid liquid transport; and under the breathability semantic is given, its pattern (A2) presents a similar mesh-like and porous structure, vividly suggesting the circulation of air. In sharp contrast, T2I-LDM, although also attempting to incorporate functional keywords, its results (B1, B2) appear relatively vague and generalized. For instance, the moisture-wicking pattern it generates has lines, but lacks a clear sense of channels and directionality. This indicates that simply adding functional words to the text prompt makes it difficult for the model to capture the underlying structured visual requirements. As for the unconditional generation of StyleGAN2-ADA (C1 and C2), although its patterns have good aesthetic quality, the content is completely random and cannot be controlled according to any specific requirements. Beyond the examples shown in Figure 3, we conducted extensive qualitative evaluations across a broader range of prompts. For instance, when combining a "monochrome geometric" text prompt with high breathability functional semantics, the FS-DM consistently generated patterns with clear porous and mesh-like structures integrated within the geometric forms. In contrast, the T2I-LDM baseline produced generic geometric designs that lacked these specific functional features. This robust performance across diverse styles further validates the effectiveness of our dual-guidance approach, a conclusion supported by the quantitative FVSR scores in Table 1. It is also important to acknowledge the model's limitations by examining a failure case, where FS-DM can produce less coherent results when the stylistic text prompt and the functional semantic vector are in strong semantic conflict. For instance, when attempting to generate a pattern for a prompt like "a delicate, soft floral lace" while simultaneously providing a strong moisture-wicking vector that favors sharp, channel-like structures, the model struggles to harmonize these dissonant concepts. The resulting pattern often appears as a disjointed mix of floral elements and harsh lines, failing to achieve a unified aesthetic. This indicates that while the dual-guidance mechanism is effective, further research is needed to improve its ability to negotiate and blend highly contradictory inputs.

Quantitative Result Analysis

Tables 1 and 2 respectively show the scores of different models in objective evaluation indicators and user subjective evaluations.

Table 1. Comparison of objective evaluation indicators for generated patterns

Model	FVSR (Moisture-wicking)	FVSR (Breathability)	Fractal dimension	Color harmony
T2I-LDM	0.56	0.51	1.68	0.75
StyleGAN2-ADA	0.23	0.25	1.75	0.72
FS-DM	0.76	0.72	1.72	0.79

Functional visual semantic relevance (FVSR): The FVSR scores of FS-DM in both functional categories were significantly higher than those of the baseline model. Particularly in the moisture-wicking function, the score reached 0.76, which was 35.7% higher than that of T2I-LDM. This strongly demonstrates the effectiveness of the proposed functional semantic guidance mechanism, and the model indeed generated visual structures highly relevant to the target function.

Computational aesthetics assessment (CAA): In terms of fractal dimension, the scores of all models were similar, indicating that the complexity and detail richness of the generated patterns reached a relatively high level. In terms of color harmony, FS-DM was slightly higher than the other models, which might be attributed to the more ordered structure brought about by the functional semantic guidance, thereby also promoting the harmony of color organization.

Table 2. Average subjective evaluation scores for generated patterns

Model	Evaluation dimensions	Ordinary consumers	Textile professionals
T2I-LDM	Aesthetic appeal	3.8	3.7
	Functional perception	2.9	2.5
	Innovativeness	3.5	3.6
StyleGAN2-ADA	Aesthetic appeal	3.9	3.8
	Functional perception	1.8	1.5
	Innovativeness	4.2	4.1
FS-DM	Aesthetic appeal	4.3	4.4
	Functional perception	4.1	4.2
	Innovativeness	4.0	4.3

Aesthetics and functional perception: FS-DM achieved the highest scores in both the aesthetic appeal and functional perception dimensions, and was unanimously recognized by both ordinary consumers and professionals. This indicates that this method successfully achieved a balance between functional expression and aesthetic value. Particularly, the score for functional perception was far higher than that of other models, verifying the effectiveness of its visual communication.

Innovativeness: In terms of innovativeness, FS-DM scored similarly to the unconditional StyleGAN2-ADA, and was higher than T2I-LDM. This suggests that imposing functional constraints not only did not stifle creativity, but instead, by guiding the model to explore new structural combinations, stimulated purposeful and valuable innovation.

CONCLUSION

We propose a diffusion-based approach to generate functional textile patterns that unify visual semantics and aesthetics. By constructing a function–visual semantic dataset and designing the Functional-Semantic Guided Diffusion Model (FS-DM), the method transforms abstract functional requirements into controllable, aesthetically compelling visual patterns. Experimental results demonstrate the effectiveness of this method from both qualitative and quantitative perspectives: compared to the baseline models, the patterns generated by FS-DM show significant advantages in terms of the accuracy of functional visual expression, aesthetic quality, and user perception evaluation.

The significance of this research lies in introducing a data-driven and intelligent design workflow to the traditional textile design field. Designers can be liberated from the cumbersome and repetitive conception work and instead play a more creative role: defining high-level design goals and functional semantics, and using AI tools to quickly generate, iterate, and optimize design proposals. We acknowledge that while our work establishes a novel paradigm, the current study is focused on three primary functional categories. The generalization of this framework to a broader spectrum of textile functions, therefore, represents a primary direction for future work.

Future research can be further developed in multiple aspects: First, the functional semantic library can be expanded to incorporate more functional types (such as UV resistance, antibacterial, flame retardancy, etc.) and explore more diverse visual expression methods; This study does not explore the model's behavior when presented with mixed or conflicting functional semantic vectors, such as a request for a pattern that is both highly breathable (visually porous) and high in thermal retention (visually dense). Preliminary tests suggest that under such conflicting guidance, the model tends to generate blended or compromised patterns that may not be physically coherent. Future work should investigate methods to manage these functional conflicts, perhaps by training the model on a dataset of multifunctional textiles or by developing a constraint-based generation process that can prioritize or negotiate between competing semantic inputs; Second, the generation of two-dimensional patterns can be combined with the simulation of three-dimensional fabric structures, so that the generated patterns are not only visually but also directly physically related to the functions; Third, an interactive design system can be developed, allowing designers to guide the pattern generation by adjusting the functional semantic vectors in real time or directly sketching on the canvas, to achieve a higher degree of personalized customization. In summary, this research demonstrates the great potential of deep generative models in empowering the development of high-end and high-value-added textiles, providing beneficial exploration for the intelligent and digital transformation of the textile industry.

Availability of Data and Materials

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

Author Contributions

Minwei Zhao and Yingying Xu designed the study; all authors conducted the study; Minwei Zhao and Hao Nong collected and analyzed the data. Yingying Xu and Hao Nong 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.

Ethics Approval and Consent to Participate

This survey was conducted in compliance with Ethics Committee of Shanghai Normal University. 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.

Acknowledgments

Not applicable.

Funding

Not applicable.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

[1] Almeida LD. Functionalisation of textiles: Future perspectives. 2006. Available from: https://api.semanticscholar.org/CorpusID:54589960

[2] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems. 2020; 33:6840-6851.

[3] Karras T, Laine S, Aila T, editors. A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 15-20 June 2019; Long Beach, CA, USA. New York City, NY, USA: IEEE; 2019.

[4] Karagoz HF, Baykal G, Eksi IA, Unal G. Textile pattern generation using diffusion models. arXiv preprint arXiv:230400520. 2023.

[5] Zhang P, He W, Wu G, Liu Y, Zhang X, Chen W, et al. Hierarchically structured dual-mode sensor with directional moisture-wicking capability for weak physiological signal monitoring and long-range motion tracking. Chemical Engineering Journal. 2025; 165585:27. Available at SSRN: https://ssrn.com/abstract=5227434 or http://dx.doi.org/10.2139/ssrn.5227434

[6] Mo X, Yang X, Hu B. The interaction of clothing design factors: how to attract consumers' visual attention and enhance emotional experience. Journal of Fashion Marketing and Management: An International Journal. 2023; 27(2):220-240.

[7] Militký J, Bajzík V. Surface roughness and fractal dimension. Journal of the Textile Institute. 2001; 92(3):91-113.

[8] Lam ADK-T, Hsu Y-L, editors. Fractal dimension of the surface shape of textile patterns. Proceedings of the 2010 Second International Conference on Multimedia and Information Technology; 24-25 April 2010; Kaifeng, China. New York City, NY, USA: IEEE; 2010.

[9] Wu X, Li L. An application of generative AI for knitted textile design in fashion. The Design Journal. 2024; 27(2):270-290.

[10] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B, editors. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition; 18-24 June 2022; New Orleans, LA, USA. New York City, NY, USA: IEEE; 2022.

[11] Gal R, Alaluf Y, Atzmon Y, Patashnik O, Bermano AH, Chechik G, et al. An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:220801618. 2022.