Culturally-Guided Style Transfer for Southeast Asian Text using Constrained Diffusion Models

Dongxue Cai[1], Li Chang[1]
[1]NanHai Academy of Art, Haikou University of Economics, Haikou 571127, Hainan, China
Corresponding author: Dongxue Cai (e-mail: HJYmsxy@163.com)

ABSTRACT Preserving culturally grounded semantics during multilingual literary style transfer remains challenging because existing text generation models often fail to maintain narrative structure, symbolic imagery, and rhetorical consistency across heterogeneous linguistic contexts. This study proposes a culturally guided style transfer framework based on constrained diffusion models that jointly models semantic hierarchies and stylistic representations in a unified latent space. By introducing style-modifying latent variables into the diffusion process and coupling them with hierarchical semantic embeddings, the proposed approach enables synchronized regulation of cultural semantics and literary style throughout forward diffusion and reverse denoising. Unlike conventional conditional generation methods, stylistic evolution emerges through semantic-aware latent perturbations, improving the preservation of cultural identity while maintaining semantic coherence. Experiments on five Southeast Asian literary corpora demonstrate stable cultural mapping and consistent style reconstruction, achieving high levels of cultural image consistency together with reliable narrative and rhetorical preservation across multilingual settings. Beyond literary generation, the proposed semantic-driven diffusion framework offers a transferable paradigm for intelligent information representation and cross-modal content encoding, providing potential insights for semantic communication, electromagnetic information transmission, and AI-assisted cultural computing where robust semantic preservation is essential.

INDEX TERMS diffusion models, semantic communication, literary style transfer, cultural semantics, cross-modal information representation

I. INTRODUCTION

SOUTHEAST Asian literature has developed a unique narrative structure and aesthetic paradigm within a multilingual and multicultural context, with its linguistic style carrying complex cultural metaphors and regional memories. This deep reservoir of regional aesthetics and cultural semantics is a crucial. However, existing AI generation models struggle to accurately capture cross-linguistic narrative logic and deep cultural semantics when processing such literary texts, leading to deviations in stylistic consistency and cultural representation in the generated texts [1,2]. This issue imposes limitations on the transfer and innovation of literary semantics in intelligent production and cultural design, and affects the potential for cross-cultural dissemination and application of Southeast Asian literary resources. It is urgent to investigate generation mechanisms that can simultaneously model semantic structure and literary style in order to build a text generation system that is adaptable to culture [3,4]. Even though research is currently developing a significant number of new models in semantic-style coupling, there are a lot of barriers to semantic-style coupling models due to the underlying differences in semantic hierarchy among multilingual literary corpora [5]. Models often become positionally prone to semantic drift and the associated narrative logic and rhetorical features become eroded. These models are not capable of accounting for case symbols or their symbolic expression of cultural imagery, which only signify an institutional or advanced mapping structure in latent space. This leads to the weakening of cultural symbols and a break in meaning in symbols during the style transfer, or the generation process [6,7]. In addition, most models utilize only a single layer attention mechanism in their generation models or a pathway where part of the text is a fixed word vector alignment. It is not an adequate mechanism for polysemous words, when a polysemy exists, and also metaphor usage in Southeast Asian language, which simply does not present the depth of cultural structure in the language [8,9]. As such, there are cases where the text generated by an AI model exhibits ideal levels for cultural consistency and semantic stability.

To tackle such issues, researchers have sought to improve the combination of semantics and style through multilingual alignment with Transformer architectures, style encoding via variational autoencoders, and conditional generation through GAN (Generative Adversarial Network) architectures [10,11]. Multilingual alignment techniques are capable of achieving cross-lingual representation by sharing word vectors, while lacking culture differentiation [12]. Variational autoencoders can have some benefits i n style extraction, but the latent variable distribution can collapse, meaning the style constraints cannot be guaranteed to be stable [13]. Conditional generation networks can control style output to some extent, but their generation path still relies on explicit labels, failing to form a dynamic coupling between semantics and style [14,15]. When dealing with multidimensional cultural corpora such as Southeast Asian literature, these methods still suffer from problems such as discontinuous semantic mapping and unstable style constraints, necessitating a generative mechanism that achieves dual synergy between semantics and style in the latent space. To address the problems of semantic hierarchical imbalance and unstable style transfer, this study proposes a Diffusion Models-driven joint modeling method for style and semantics. This method learns the dynamic distribution of literary styles by constraining cultural semantics during the forward diffusion stage, and introduces style latent variables to control the generation path during the reverse denoising stage, constructing a synchronous regulation mechanism for semantics and style. The implementation path includes the following: First, semantic hierarchical modeling is performed on multilingual corpora to form an embedded structure of narrative logic, rhetorical features, and cultural imagery; second, style regulation latent variables are introduced into the diffusion latent space to achieve the joint evolution of cultural semantics and literary style through conditional diffusion distribution; finally, latent variables guide semantic reconstruction in the reverse denoising generation stage to restore literary texts with stylistic consistency. This research establishes a dynamic mapping relationship between semantic and style latent spaces in the generation task of multilingual literature, providing a theoretical model and technical path for the controllable transfer of literary styles and the generation and reproduction of cultural semantics.

II. DIFFUSION-DRIVEN GENERATION MECHANISM FOR SOUTHEAST ASIAN LITERARY STYLE

A. SEMANTIC HIERARCHICAL MODELING OF LITERARY CORPUS

The semantic hierarchical modeling of Southeast Asian literary corpus aims at a unified semantic representation of multilingual texts, constructing a multi-layered embedding system from vocabulary, syntax to cultural semantics. Let the corpus set be $X = \{x_1, x_2,..., x_n\}$, where x_i represents the i-th literary text after word segmentation and language standardization. The model maps the input text to the semantic vector space through a multilingual semantic encoding function f, obtaining a semantic embedding matrix $H = [h_1, h_2,..., h_n]$, where $h_i \in \mathbb{R}^d$ is the representation of the text in the d-dimensional semantic space. The encoding function f adopts a cross-lingual Transformer structure, ensuring the consistency of multilingual semantics by sharing word vectors and aligning position encoding [16,17].

In semantic hierarchical modeling, the embedding vector h_i is decomposed into three parts: the narrative logic vector n_i, the rhetorical feature vector r_i, and the cultural image vector c_i, namely:

$$h_i = W_n n_i + W_r r_i + W_c c_i \tag{1}$$

In formula (1), W_n, W_r, and W_c refer to each mapping matrix for semantic subspace. The narrative logic vector n_i is calculated based on syntactic dependenc trees and the attention weights of discourse structure, the rhetorical feature vector r_i is generated based on implicit metaphor recognition and analysis of phonological patterns, and the cultural image vector c_i is obtained from the cross-linguistic cultural entity alignment model to ensure the distinguishability of the semantic space in the culture dimension. To facilitate a more likewise semantic layer in the model, this paper introduces a regularization semantic objective function.

$$\begin{aligned} L_{\text{sem}} = \|f_\theta(x_i) - (W_n n_i + W_r r_i \\ + W_c c_i)\|_2^2 + \lambda_1 \|n_i - r_i\|_1 \\ + \lambda_2 \|r_i - c_i\|_1 \end{aligned} \tag{2}$$

In formula (2), $\|\cdot\|_2$ refers to the Euclidean distance metric, $\|\cdot\|_1$ indicates the L_1 norm constraint, and λ_1 and λ_2 are the balance coefficients to control the hierarchical correlation among the semantic subspaces. This loss term is developed so that the model can discern the structural differences between narrative, rhetoric and cultural features without discarding the overall semantic meaning.

After training, the semantic embedding matrix H constitutes the semantic basis distribution $P_0(h)$ in the latent space of the diffusion model. This distribution describes the initial state of the literary corpus under the multidimensional semantic hierarchy, providing a controllable semantic prior for subsequent diffusion modeling of style latent variables [18,19]. Through this hierarchical representation, the model achieves the co-embedding of semantic content, rhetorical structure, and cultural imagery, enabling Southeast Asian literary corpus to present a stable hierarchical distribution in the semantic space.

In cultural design, narrative logic corresponds to the overall compositional relationship of patterns, rhetorical features can be mapped to pattern rhythm or color rhyme, and cultural imagery is embodied in visual elements with regional symbolism [20,21]. This semantic hierarchical modeling method establishes a transferable em-bedding foundation between "literary semantics and visual patterns," enabling the model to transform literary semantics into the cultural expression logic of design in subsequent generation.

B. DIFFUSION MODELING OF STYLE SEMANTIC LATENT VARIABLES

Unlike conventional conditional diffusion frameworks in text generation, where control signals are injected during the reverse denoising phase to bias the sampling trajectory toward predefined attributes, the proposed mechanism integrates cultural semantic regulation into the stochastic forward diffusion process itself. In this formulation, style-modifying latent variables do not operate as post hoc constraints on decoded tokens, nor as classifier-based gradient corrections. Instead, they function as semantic-aligned modifiers that reshape the noise accumulation dynamics, ensuring that cultural imagery, rhetorical cadence, and narrative organization co-evolve with semantic degradation throughout the diffusion trajectory. This design reframes cultural style not as an auxiliary condition but as a structural dimension of semantic uncertainty, which is particularly suited to modeling culturally grounded literary expression where stylistic form and semantic content are inseparable.

The notion of cultural semantic constraints in the forward diffusion stage does not imply the direct encoding of subjective cultural rules into Gaussian noise. Instead, the semantic embedding space itself serves as the carrier of cultural structure, and the forward diffusion process operates over this structured space rather than over raw linguistic tokens. Noise perturbation therefore interacts with culturally aligned semantic dimensions learned during hierarchical encoding, causing culturally salient features to decay more slowly along the diffusion trajectory. This interpretation remains consistent with DDPM theory while extending its applicability to semantically structured latent spaces.

Diffusion modeling of style semantic latent variables is based on the semantic embedding matrix H in the previous section. Literary style regulation latent variables $z \in \mathbb{R}^d$ are introduced into the latent space of the diffusion model to achieve joint modeling of cultural semantics and style features [22,23]. Let the initial semantic distribution be $P_0(h)$. Gaussian noise is gradually added during the forward diffusion process, causing the semantic vector to evolve from the observable semantic state h_0 to the completely noisy state h_t. This process is defined as:

$$q(h_t \mid h_{t-1}) = \mathcal{N}\left(h_t; \sqrt{1-\beta_t}\, h_{t-1}, \beta_t I\right) \tag{3}$$

In formula (3), $t \in \{1,,..., T\}$ represents the diffusion step number, β_t is the diffusion rate control parameter, and I is the unit covariance matrix. This process establishes a continuous mapping of semantic embedding from deterministic states to noisy states, forming a semantic diffusion trajectory $\{h_0, h_1, \dots, h_t\}$. To inject literary style information into the diffusion process, a style regulation latent variable z is introduced, and conditional constraints are imposed on each semantic state h_t to construct a conditional diffusion distribution:

$$q(h_t \mid h_{t-1}, z) = \mathcal{N}\left(h_t; \sqrt{1-\beta_t}\,(h_{t-1} + \gamma_t z), \beta_t I\right) \tag{4}$$

In formula (4), γ_t is the style modulation coefficient, which controls the intensity of the influence o f latent variable z on the current semantic diffusion. In contrast to generic stylistic attributes such as tone or register, the latent variable z is constrained by the semantic hierarchy defined i n S ection " Semantic H ierarchical M odeling of Literary Corpus", which anchors stylistic modulation to culturally specific n arrative p atterns a nd s ymbolic imagery. As a result, the modulation term $\gamma_t z$ does not uniformly bias the diffusion process but selectively amplifies o r attenuates noise along semantic dimensions associated with cultural representation, thereby preserving culturally meaningful structures that would otherwise collapse under isotropic diffusion. The latent variable z is generated by the style encoding function $g_\phi(\cdot)$, and is obtained based on feature extraction from the style corpus $S = \{s_1,..., s_m\}$:

$$z = \mathrm{g}\varphi(S) \tag{5}$$

This latent variable represents the continuous distribution of literary style in the latent space and is used as a learnable parameter during training. It is optimized together with the semantic embedding h_t, thereby achieving dynamic coupling between cultural semantics and stylistic semantics. The model training objective is defined as minimizing the variational lower bound loss of the diffusion process:

$$L_{diff} = E_{(q(\mathrm{h_t} \mid \mathrm{h_0}, z))} \left[\|\varepsilon - \varepsilon\theta(\mathrm{h_t}, t, z)\|^2\right] \tag{6}$$

In formula (6), ϵ is the real noise vector, ϵ_θ is the model predicted noise, and θ is the diffusion network parameter. The model learns in back-propagation the noise removal rules and style latent variable regulation at each diffusion step by updating parameters θ and ϕ. While the style latent variable z is still operating on the semantic diffusion path, the latent space distribution transitions from $P_0(h)$ to the style regulation distribution $P_t(h|z)$, which is continuously differentiable in both the semantic and style dimensions. Through the modeling process, the diffusion model learns a dynamic distribution trajectory in the latent space constrained by style semantics. The semantic embedding evolves co-evolves under noise perturbation and the regulation of the latent variable [24,25], completing the basic modeling of style transfer in Southeast Asian literary corpus, and providing controlled conditional signals for text generation in the back-denoising stage [26].

From a cultural design perspective, the style latent variable z can be understood not only as a literary style control signal but also as a "pattern style encoder." It imposes constraints on cultural semantics in the latent space, enabling the model to learn the potential distribution patterns of different regional cultures in color, structure, and rhythm [27,28]. Through the dynamic regulation of the style latent variable, cross-modal transfer from "literary narrative semantics" to "visual style" can be achieved, providing a parameterized style control mechanism for intelligent design.

III. TEXT RECONSTRUCTION GENERATION BASED ON REVERSE DENOISING

Text reconstruction generation based on inverse denoising starts with the style semantic distribution learned by the diffusion model in the latent space, and gradually restores the initial semantic state h_0 from the final noise vector h_t to achieve the generation and style reproduction of literary texts [29,30]. The goal of the inverse process is to gradually remove noise according to the conditional distribution $p(h_{t-1} \mid h_t, z)$ to recover the semantic content under the constraints of style latent variables. This process is described by the following generative equation:

$$h_{t-1} = (1 / \sqrt{(1 - \beta_t)}) \cdot (h_t - \beta_t \cdot \varepsilon\theta(h_t, t, z)) + \sigma_t \cdot \xi \tag{7}$$

In formula (7), $\epsilon_\theta(h_t, t, z)$ is the model's predicted noise function, σ_t is the noise standard deviation, and ξ is the random perturbation sampled from the standard normal distribution. This formula ensures that semantic reconstruction follows the noise removal rules in each diffusion step while preserving the constraint effect of the style latent variable z on the semantic trajectory, so that the generation path smoothly converges to the stylized target text distribution in the semantic space [31,32].

In each denoising generation step, the model calculates the conditional probability density:

$$p(h_{t-1} \mid h_t, z) = \mathcal{N}(h_{t-1}; \mu_\theta(h_t, t, z), \Sigma_t) \tag{8}$$

In formula (8), $\mu_\theta(h_t, t, z)$ is the mean predicted by network parameter θ. Σ_t is the covariance matrix associated with timestep t. This conditional distribution is obtained by learning a form of ϵ_θ that approximates the distribution, which allows the model to recover semantic features without changing the stability of the gradient as trained backpropagation.

As the semantic embedding denoised to h_0 is processed by the model, the model uses the decoding function $f^{-1}\theta\ (\cdot)$ to map the semantic vector back to the natural language sequence y, producing generated text with literary style characteristics, i.e.:

$$y = f^{-1}\theta(h_0) \tag{9}$$

In equation (9), $f^{-1}\theta\ (\cdot)$ is the diffusion decoder, which is used to produce the inverse mapping from semantic vectors to linguistic symbols. During the decoding process, the style latent variable z continues to play the role of an input control signal that preserves the consistency of cultural imagery, rhetorical cadence, and narrative structure.

The optimization goal as part of the generation stage remains the same as that in the diffusion process: to enable both reliable semantic recovery and stable style control by minimizing the error between the predicted noise and the actual noise. During training, the model constrains both semantic consistency and style preservation at the same time, allowing for continuity and controllability in the expression of cultural semantics in the latent space at the reverse denoising stage.

Following the training, the de-noising process establishes a reversibly mapping from noise to semantics, from latent space to language, so that published Southeast Asian literary corpora can generate text that is stylistically consistent, semantically coherent, and culturally imagistic, which represents a deep integration of the transfer of literary text style and new literary text generation [33,34].

In the design generation task, this reverse denoising process can be understood as an equivalent process of "semantic-pattern reconstruction." The denoising trajectory in the semantic space is equivalent to the gradual manifestation of cultural imagery in the design space, and the generated text semantics can be further transformed into text input prompts for pattern generation [35]. Therefore, the model can not only generate stylized literary texts but also provide stable and interpretable semantic control signals for the pattern generation system, achieving a unified mapping from literary style to design style.

A. MULTILINGUAL SOUTHEAST ASIAN LITERARY CORPUS CONSTRUCTION

The multilingual Southeast Asian literary corpus used in this study was constructed to support hierarchical semantic modeling and culturally grounded style transfer rather than surface-level lexical variation. The corpus consists of approximately 1.2 million sentences drawn from canonical and contemporary literary texts written in Filipino, Vietnamese, Thai, Indonesian, and Malay, with each language contributing a balanced proportion of narrative prose and poetic discourse. Texts were selected based on their sustained use of culturally embedded symbolism, narrative conventions, and metaphorical language that reflect regional literary traditions. Corpus curation followed a semantic density filtering process, where texts exhibiting shallow descriptive language or formulaic narrative structures were excluded to ensure that multi-layered cultural semantics were present and structurally extractable. All texts underwent language normalization and segmentation while preserving culturally salient expressions, ensuring that semantic hierarchies related to narrative logic, rhetorical form, and cultural imagery remained intact for subsequent modeling.

IV. EXPERIMENTAL VERIFICATION OF SOUTHEAST ASIAN LITERARY STYLE TRANSFER GENERATION

To evaluate the regional generality of the proposed diffusion-based semantic-style modeling framework, experimental validation was conducted across five Southeast Asian cultural contexts representing distinct linguistic families and literary traditions. Rather than treating cultural sensitivity as a country-specific phenomenon, the experimental design examines whether the latent semantic-style coupling mechanism maintains structural stability and stylistic coherence under heterogeneous cultural distributions.

A. SEMANTIC EMBEDDING DISTRIBUTION ANALYSIS

After completing the semantic hierarchical modeling, this section conducts a visual analysis of the semantic embeddings of multilingual corpora of Southeast Asian literature to examine the ability of the diffusion model to distinguish cultural semantic features in the latent space. Based on five subsets of literary corpora from Vietnam, Thailand, Indonesia, the Philippines, and Malaysia, the model maps the high-dimensional semantic representation to a two-dimensional plane using t-SNE (t-distributed Stochastic Neighbor Embedding). The clustering trends and differentiation characteristics of the embedding space are examined from two perspectives: semantic aggregation structure and cultural distribution boundaries. Figure 1 shows the distribution patterns of different cultural corpora in the latent space, reflecting the spatial organization of semantic embeddings across cultural dimensions.

In Figure 1, the Filipino corpus is located in the lower left region, where the points are densely clustered, indicating that the model maintains high consistency and stability in semantic modeling. The Thai corpus is located in the upper left region, where it also exhibits a clustered distribution. However, the local diffusion of this corpus is relatively small, suggesting a relatively balanced relationship between narrative structure and cultural expression in this latent space. The Indonesian corpus is located in the lower right region, where the points are evenly distributed. This indicates that rhetorical features and cultural imagery are co-embedded in this region. The Vietnamese corpus is located in the upper right region, where the points are relatively loosely clustered. This suggests that there are many differences in semantic expression styles within this corpus. The Malay corpus is located in the upper right region, where the point clustering is moderate. This indicates that this corpus presents an intermediate feature in the overall semantic space. These results demonstrate that the proposed model can obtain culturally separable and structurally stable embedding patterns in the latent space, with clear cluster boundaries, and semantic features from different cultures are consistently embedded into the proposed model.

B. STYLE LATENT VARIABLE DIFFUSION CHARACTERISTICS

In the latent variable diffusion phase of style, the objective of the experiment is to uncover the dynamic law of style intensity changes with the increasing number of diffusion steps in latent space across different Southeast Asian literary corpora, in order to ascertain the regulatory stability and evolutionary properties of the diffusion model in multicultural style modeling. In this phase, five literary corpora were selected from Vietnam, Thailand, Indonesia, the Philippines, and Malaysia. Through the quantification of the intensity changes in style latent variables over the course of 100 diffusion steps, the project identified the steady amplification process of style signals in the diffusion phase, as a result of semantic noise perturbation. Figure 2 illustrates the dynamic evolution curves of style latent variables across the diffusion process, which demonstrates the convergence trend and difference in style distribution across the latent space of different cultural corpora.

The curves in Figure 2 exhibit traditional nonlinear growth characteristics. The style intensity of the Vietnamese corpus increases to 0.58 at step 40 and reaches a stable value of 0.91 at step 100, indicating that the model achieves stronger co-convergence of narrative and rhetorical signals in the semantic mapping of the corpus. The intensity of the Filipino corpus increases significantly in the mid-stage, increasing by 0.22 from step 30 to step 50, suggesting that the cultural symbol components in this corpus may be more sensitive to the moderating effect of latent variables. The Thai and Indonesian corpora reach intensity values of 0.70 and 0.67 respectively at step 60, indicating that the style semantics in these corpora show a high degree of consistency in the diffusion trajectory. The final value of the Malaysian corpus is 0.84, reflecting a relatively neutral degree of cultural embedding in its semantic expression. The overall trend indicates that the style latent variables show a stable strengthening trend during the continuous diffusion process, and the convergence differences of different cultural corpora in the latent space reflect the hierarchical adaptability of the diffusion model in the representation of multilingual literary styles.

C. PERFORMANCE ANALYSIS OF TEXT GENERATION QUALITY AND RECONSTRUCTION STABILITY

In the experiment of text generation and reconstruction, the study evaluates both the intrinsic quality of newly generated content and the stability of semantic and stylistic recovery along the reverse denoising trajectory. In addition to style-related reconstruction indicators, text generation quality was assessed using intrinsic linguistic measures that reflect the validity and originality of generated content. Fluency was evaluated through normalized language model perplexity over generated sequences, coherence was measured via sentence-level semantic continuity across generated passages, and lexical diversity was quantified using distributional variation in vocabulary usage across outputs. These measures characterize whether the diffusion process produces linguistically well-formed and semantically coherent texts that extend beyond structural imitation of the source corpus. By continuously tracking the output quality of the model in different denoising steps, the stability of semantic information reconstruction and the coherence recovery of language structure during the noise backpropagation process are analyzed, revealing the convergence characteristics of the model in the semantic-style latent space and the evolutionary law of the generation stage. Figure 3 shows the results of this process.

The three indicators demonstrate a generally non-linear increasing trend,Across all evaluated corpora, perplexity values remained within a stable range, while coherence and diversity measures exhibited convergence patterns aligned

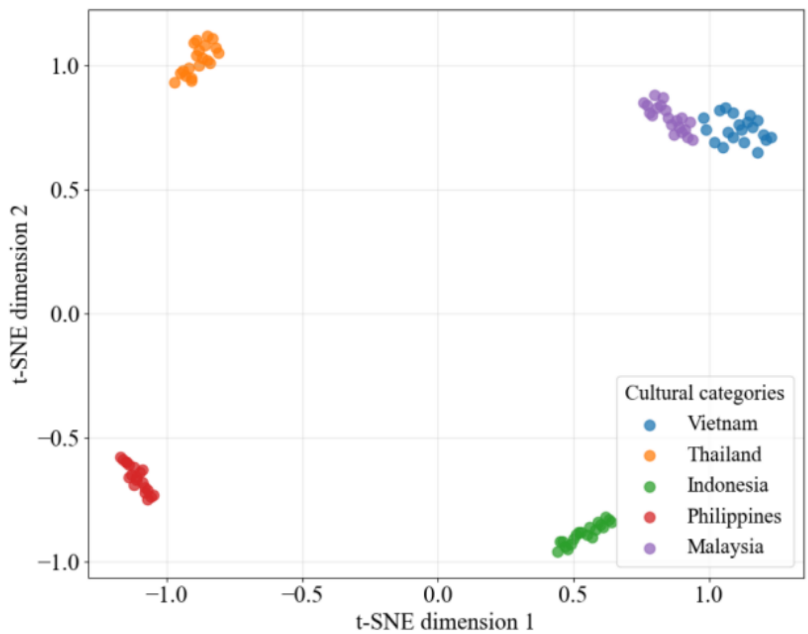

FIGURE 1. Semantic Embedding Distribution of Southeast Asian Literary Corpora

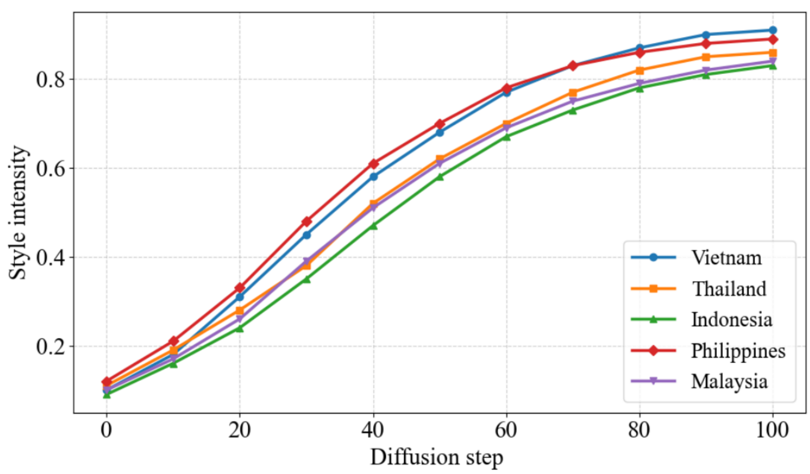

FIGURE 2. Dynamic Evolution of Style Latent Variable Diffusion

with denoising progression, indicating that the generation process supports the creation of structurally novel and linguistically coherent texts rather than template-based reproduction, which reflects the progressive stabilization of the model at the semantic and narrative levels in Figure 3. The semantic consistency increased to a value of 0.77 after 40 denoising steps, and stabilized at 0.92 after 100 steps. This indicates, given that semantic information was coalescing around existing, strong structures in the mid-stages, that these early steps were effectively allowing semantic stabilization. The narrative coherence increased, but began to plateau after about 40 steps at 0.89. The relative change in narrative coherence was slower than that of semantic consistency, which suggests that restoring narrative logic relies on the stable formation of the semantic structure. The smoothness of the generation improved with relative steadiness, increasing to 0.91, in the mid-to-late stages of denoising, suggesting that the language rhythm in terms of corpus generation remained smooth and consistent under latent style constraints. These overall performance patterns suggested that the model continued to build the collaborative effect of the strengthened context of semantic structure and rhythm of narrative language in continuous denoising, until it stabilized its generation state in the semantic restoration of style-harmonized text.

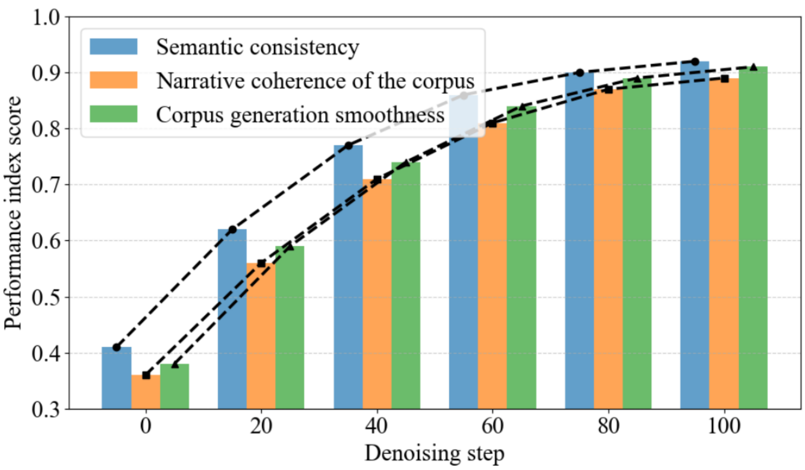

FIGURE 3. Performance Changes in Text Denoising and Reconstruction Stages

D. EVALUATION OF LITERARY STYLE PRESERVATION

Cultural image consistency is operationalized as the alignment stability between cultural imagery embeddings extracted from original texts and those derived from generated outputs. This metric is computed by projecting both source and generated texts into the cultural image subspace defined in Section "Semantic Hierarchical Modeling of Literary Corpus" and measuring the cosine similarity between their normalized cultural image vectors across aligned narrative segments. The reported score represents the mean similarity across the evaluation corpus, with higher values indicating stronger preservation of culturally salient symbolic structures under generative transformation. Validation is achieved through cross-corpus consistency analysis, ensuring that the metric reflects structural alignment rather than lexical overlap.

Multilingual literary texts from Southeast Asian Cultural contexts are going to have varying styles trans-ferred in the diffusion model generative system. In order to reveal the model's preservation patterns in dimensions such as cultural imagery, narrative story structure, rhetor-ical style, and language fluency, this paper performs a systematic assessment of model style preservation using the multilingual corpora trained previously in this research section. The experiment can posit comparison of the original corpus semantic mappings with the generated text in the latent space, extracting the differences in cross-cultural consistency of semantic intention mappings in addition to linguistic style expression - exploring the overall stability and adaptability of the diffusion model in variation of the semantic-style dual dimensions reviewed. Figure 4 portrays the distribution of literary corpora across countries in the four evaluative indicators, reflecting the structure of the generative result model across style generative performance of varying multicultural styles dimension contexts.

Figure 4 shows that the corpora had generally high scores in cultural imagery consistency and language fluency, which discussed that the model had a relatively stable cultural symbol map as latent variables diffused. The Filipino corpus achieved a cultural imagery consistency score of 0.934 along with a narrative structure preservation score of 0.906, and a rhetorical style reproducibility score of 0.892. The high distribution values of the Filipino corpus suggest that the model established a more consolidated style convergence trajectory in linguistic contexts with strong cultural symbolism. The Vietnamese corpus scored all four indicators over 0.862, and there was again relatively concentrated semantic space distribution with a reasonable balance of narrative through structure and rhetorical rhythm. In contrast, the Indonesian and Malaysian corpora had scores below 0.86 for narrative structure preservation and below 0.86 for rhetorical style reproducibility reflecting a strong semantic diffusion and style dilution effect on these corpora. The general trend seems to demonstrate that the diffusion model achieves more stable style preservation and control with support for enough semantic alignment with dense cultural imagery in the previously described corpora and overall validity of literary generation, as a form of semantic control, across multilingual contexts.

E. COMPREHENSIVE PERFORMANCE ANALYSIS OF GENERATED TEXTS

Analyzing the performance is discussed in this section, assessing the multilingual literary texts created by the diffusion model in relationship to four dimensions: cultural innovation, style fusion, depth of semantic generation, and emotional expressiveness. This is to assess how the model demonstrates a comprehensive adaptability for cross-cultural literary creation within the semantic-style latent space. Based on the aforementioned corpus system, the experiment statistically analyzed the multidimensional performance results of generated texts in different languages, constructed a multidimensional numerical matrix, and visualized it to reflect the overall generation distribution and internal differences of the model under multicultural styles.

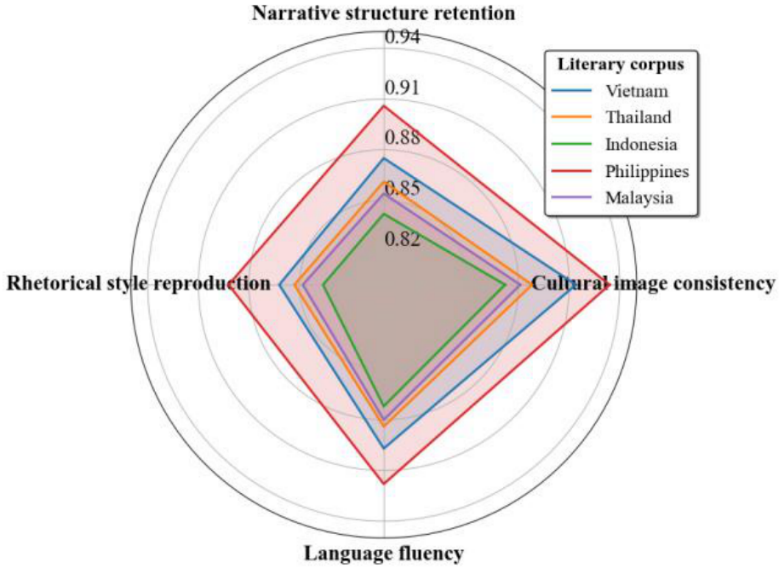

FIGURE 4. Distribution of Literary Style Preservation Indicators

Figure 5 shows the performance pattern of literary corpora from different countries in terms of comprehensive feature dimensions, revealing the overall balance of the model in the process of cultural semantic generation.

Figure 5 shows that the Filipino corpus exhibits the highest values in cultural innovation and style integration, reaching 0.91 and 0.92 respectively, indicating that the generation of cultural imagery and style integration are more natural, and the model has formed a relatively stable convergence trajectory in the latent space of this corpus. The Vietnamese corpus is in a high range in both semantic generation depth and emotional expressiveness, with an emotional expressiveness of 0.89, reflecting the subtlety of the text's narrative emotion and the richness of its semantic layers. The Thai corpus has a relatively stable overall distribution, with style integration remaining at 0.87, indicating that the model has learned the potential patterns of the language style quite well, but the innovation is limited. The Indonesian and Malaysian corpora have relatively similar indicators, with cultural innovation values of 0.82 and 0.84 respectively. These relatively low numbers indicate the model's low sensitivity to capturing local cultural symbols during the semantic reconstruction process. Overall, there is a trend showing that the generated results from different corpora demonstrate a hierarchical distribution in the cultural and stylistic dimensions. The model shows stable generation abilities and consistency in cultural quality during multilingual literary style transfer.

V. CONCLUSION

This research centers on multicultural semantic modeling and style generation within Southeast Asian literature. Crucially, the developed methodology provides a direct pathway for the industry to translate these complex cultural semantics into unique, market-ready design aesthetics, thereby enriching product innovation with authentic regional narratives. A linking and coupling channel between semantic structure and literary style was developed through the mechanisms of diffusion and denoising enabled through Diffusion Models. It was demonstrated in experimental results that video model construction exhibits strong semantic stability and strong style consistency during the reverse generation stage. Semantic consistency reached 0.92 at 100 time steps, narrative consistency converged to 0.89 and smoothness converged to 0.91. The style latent variables in the Vietnamese corpus stabilized at a style intensity of 0.91 with continuous diffusion. Overall, the results show a dynamic stabilization between semantics and style. The diffusion model exhibited a joint and mutual distribution of narrative logic, rhetorical rhythm, and cultural imagery within latent space that allows the generated text to possess continuity and cultural depth in both language rhythm and cultural representation of the text. This shows there is computational potential for the diffusion model to reconstruct literary style at complex levels of semantic meaning. The research still encounters limitations such as the perhaps blurred boundaries of style and uneven mapping of cultural imagery at low density. Future research can combine multi-layer semantic regularization and cross-modal cultural alignment strategies to enhance the model's sensitivity to micro-semantic structures and its ability to generate cross-cultural styles, promoting its in-depth application in Southeast Asian literary innovation writing and cultural design generation. These advanced strategies are vital for the industry's global design efforts, ensuring that

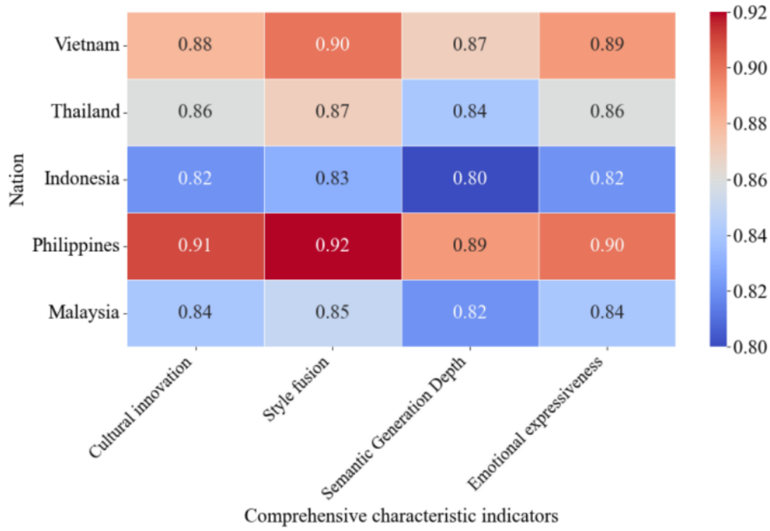

FIGURE 5. Heatmap of Comprehensive Features of Multi-National Literary Generated Texts

AI-generated apparel and fabric aesthetics accurately reflect subtle cultural nuances, thereby facilitating highly resonant, cross-cultural product releases in international markets.

AUTHOR CONTRIBUTIONS

Conceptualization –Dongxue Cai and Li Chang; methodology – Dongxue Cai and Li Chang; investigation – Dongxue Cai and Li Chang; writing-original draft preparation – Dongxue Cai and Li Chang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Q. Yi, X. Chen, C. Zhang, Z. Zhou, L. Zhu, and X. Kong, "Diffusion models in text generation: a survey," PeerJ Computer Science, vol. 10, no. 1, pp. e1905, 2024, doi: 10.7717/peerj-cs.1905.

[2] S. Colella, "The Language of the Digital Air": AI-Generated Literature and the Performance of Authorship," Humanities, vol. 14, no. 8, pp. 164, 2025, doi: 10.3390/h14080164.

[3] S. Pang, X. Chen, Y. Xie, H. Zhan, B. Yin, and Y. Lu, "Diff-TST: Diffusion model for one-shot text-image style transfer," Expert Systems with Applications, vol. 263, no. 1, pp. 125747, 2025, doi: 10.1016/j.eswa.2024.125747.

[4] Y. Wu, H. Zhao, W. Chen, Y. Yang, and J. Bu, "TextStyler: A CLIP-based approach to text-guided style transfer," Computers & Graphics, vol. 119, no. 1, pp. 103887, 2024, doi: 10.1016/j.cag.2024.103887.

[5] E. Troiano, A. Velutharambath, and R. Klinger, "From theories on styles to their transfer in text: Bridging the gap with a hierarchical survey," Natural Language Engineering, vol. 29, no. 4, pp. 849-908, 2023, doi: 10.1017/S1351324922000407.

[6] J. Peng, Y. Zhou, X. Sun, L. Cao, Y. Wu, and F. Huang, "Knowledge-driven generative adversarial network for text-to-image synthesis," IEEE Transactions on Multimedia, vol. 24, no. 1, pp. 4356-4366, 2021, doi: 10.1109/TMM.2021.3116416.

[7] Q. Wang, S. Li, Z. Wang, X. Zhang, and G. Feng, "Multi-source style transfer via style disentanglement network," IEEE Transactions on Multimedia, vol. 26, no. 1, pp. 1373-1383, 2023, doi: 10.1109/TMM.2023.3281087.

[8] N. Ichien, D. Stamenković, and J. Holyoak K, "Large language model displays emergent ability to interpret novel literary metaphors," Metaphor and Symbol, vol. 39, no. 4, pp. 296-309, 2024, doi: 10.1080/10926488.2024.2380348.

[9] T. Chay-intr, H. Kamigaito, and M. Okumura, "Character-based Thai word segmentation with multiple attentions," Journal of Natural Language Processing, vol. 30, no. 2, pp. 372-400, 2023, doi: 10.5715/jnlp.30.372.

[10] A. Scalercio and A. Paes, "Masked transformer through knowledge distillation for unsupervised text style transfer," Natural Language Engineering, vol. 30, no. 5, pp. 973-1008, 2024, doi: 10.1017/S1351324923000323.

[11] Y. Shi, S. Zhang, C. Zhou, X. Liang, X. Yang, and L. Lin, "Gtae: graph transformer–based auto-encoders for linguisticconstrained text style transfer," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 12, no. 3, pp. 1-16, 2021, doi: 10.1145/3448733.

[12] D. Wu and C. Monz, "Beyond Shared Vocabulary: Increasing Representational Word Similarities across Languages for Multilingual Machine Translation," arXiv e-prints, vol. 2305, no. 1, pp. 14189, 2023, doi: 10.18653/v1/2023.emnlp-main.605.

[13] S. Cao, J. Li, P. Nelson K, and A. Kon M, "Coupled VAE: Improved accuracy and robustness of a variational autoencoder," Entropy, vol. 24, no. 3, pp. 423, 2022, doi: 10.3390/e24030423.

[14] H. Zhang, M. Wang, L. Zhang, Y. Wu, and Y. Luo, "Image Generation with Global Photographic Aesthetic Based on Disentangled Generative Adversarial Network," Applied Sciences, vol. 13, no. 23, pp. 12871, 2023, doi: 10.3390/app132312871.

[15] X. Han, Y. Wu, and R. Wan, "A method for style transfer from artistic images based on depth extraction generative adversarial network," Applied Sciences, vol. 13, no. 2, pp. 867, 2023, doi: 10.3390/app13020867.

[16] W. Lin J, W. Su T, and C. Chang C, "Chinese Story Generation Based on Style Control of Transformer Model and Content Evaluation Method," Algorithms, vol. 18, no. 3, pp. 168, 2025, doi: 10.3390/a18030168.

[17] R. Zandie and H. Mahoor M, "Topical language generation using transformers," Natural Language Engineering, vol. 29, no. 2, pp. 337-359, 2023, doi: 10.1017/S1351324922000031.

[18] H. Li, F. Xu, and Z. Lin, "ET-DM: Text to image via diffusion model with efficient Transformer," Displays, vol. 80, no. 1, pp. 102568-102577, 2023, doi: 10.1016/j.displa.2023.102568.

[19] C. Li, L. Zhang, and Q. Zheng, "Utilizing latent diffusion model to ac-

celerate sampling speed and enhance text generation quality," Electronics, vol. 13, no. 6, pp. 1093, 2024, doi: 10.3390/electronics13061093.
[20] A. Wahida and M. H. Himawan, "Contemporary textile design creation sourced from visual aesthetics of Kawung motif classical batik," Mudra Jurnal Seni Budaya, vol. 39, no. 4, pp. 494-507, 2024, doi: 10.31091/mudra.v39i4.1807.
[21] H. An M and A. R. Jang, "Development of textile pattern design by MC Escher's tessellation technique using chaekgeori icons," Fashion and Textiles, vol. 10, no. 1, pp. 15, 2023, doi: 10.1186/s40691-023-00336-w.
[22] W. Ke, Y. Guo, Q. Liu, W. Chen, P. Wang, H. Luo, et al., "MDM: Meta diffusion model for hard-constrained text generation," Knowledge-Based Systems, vol. 283, no. 1, pp. 111147, 2024, doi: 10.1016/j.knosys.2023.111147.
[23] C. Liu, W. Yin, Y. Xu, Q. Zhan, D. Zhang, and R. Ahmad, "DDM4TST: Diffusion Model for Fine-grained Text Style Transfer by Disentangled Representation," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 24, no. 10, pp. 1-17, 2025, doi: 10.1145/3749195.
[24] Y. Wang, B. Zhang, W. Liu, J. Cai, and H. Zhang, "STMAP: A novel semantic text matching model augmented with embedding perturbations," Information Processing & Management, vol. 61, no. 1, pp. 103576, 2024, doi: 10.1016/j.ipm.2023.103576.
[25] J. Huang, D. Zheng, J. Li, and Patrick Cheong-Iao Pang, "EmbDiffounder: Semantic-Transfer Enhanced Sequence Diffusion Model for Text Generation," IEEE Access, vol. 13, no. 1, pp. 205521-205531, 2025, doi: 10.1109/ACCESS.2025.3638810.
[26] A. Asperti, D. Evangelista, S. Marro, and F. Merizzi, "Image embedding for denoising generative models," Artificial Intelligence Review, vol. 56, no. 12, pp. 14511-14533, 2023, doi: 10.1007/s10462-023-10504-5.
[27] Z. Yu, J. Jin, J. Zhao, Z. Fu, and J. Yang, "TtfDiffusion: Training-free and text-free image editing in diffusion models with structural and semantic disentanglement," Neurocomputing, vol. 619, no. 1, pp. 129159, 2025, doi: 10.1016/j.neucom.2024.129159.
[28] S. Chen, H. Zhang, M. Guo, Y. Lu, P. Wang, and Q. Qu, "Exploring low-dimensional subspace in diffusion models for controllable image editing," Advances in neural information processing systems, vol. 37, no. 1, pp. 27340-27371, 2024, doi: 10.52202/079017-0859.
[29] N. Gao, Y. Lu, P. Chen, G. Sun, R. Liang, and Y. Zhang, "TWDT: Training-Free Word-Level Controllable Diffusion Model for Text Generation," Knowledge-Based Systems, vol. 330, no. 1, pp. 114437, 2025, doi: 10.1016/j.knosys.2025.114437.
[30] Y. Xu, M. Gong, S. Xie, W. Wei, M. Grundmann, and K. Batmanghelich, "Semi-implicit denoising diffusion models (siddms)," Advances in neural information processing systems, vol. 36, no. 1, pp. 17383, 2024, doi: 10.48550/arxiv.2306.12511.
[31] W. Han, H. Cho, D. Kim, and J. Y. Kim, "SAL-PIM: A Subarray-level processing-in-memory architecture with LUT-based linear interpolation for transformer-based text generation," IEEE Transactions on Computers, vol. 74, no. 9, pp. 2909-2922, 2025, doi: 10.1109/TC.2025.3576935.
[32] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, and R. Berg, "Structured denoising diffusion models in discrete state-spaces," Advances in neural information processing systems, vol. 34, no. 1, pp. 17981-17993, 2021, doi: 10.48550/arXiv.2107.03006.
[33] R. Jiang, C. Zheng G, T. Li, R. Yang T, D. Wang J, and X. Li, "A survey of multimodal controllable diffusion models," Journal of Computer Science and Technology, vol. 39, no. 3, pp. 509-541, 2024, doi: 10.1007/s11390-024-3814-0.
[34] C. Scribano, D. Pezzi, G. Franchini, and M. Prato, "Denoising diffusion models on model-based latent space," Algorithms, vol. 16, no. 11, pp. 501, 2023, doi: 10.3390/a16110501.
[35] T. Choi H, K. Nakamura, and W. Hong B, "Decoupled Latent Diffusion Model for Enhancing Image Generation," IEEE Access, vol. 13, no. 1, pp. 130505-130516, 2025, doi: 10.1109/ACCESS.2025.3592163.