

Application of Diffusion Models in Complex Visual Communication Design and Their Enhancement of Design Efficiency

Jingwen Zhou, Weijing Yuan

School of Art and Design, Heilongjiang Institute of Technology, Harbin 150000, Heilongjiang, China

Corresponding author: Weijing Yuan (e-mail: Ywj202604@163.com).

ABSTRACT Diffusion models provide controllable and high-fidelity generative tools for complex visual communication design through progressive denoising and flexible conditional guidance. This study investigates their application from four perspectives: technical adaptability, workflow reconstruction, efficiency quantification, and large-scale deployment. By analyzing brand visual identity, multimodal interfaces, and cross-media narrative tasks, the study shows that diffusion models can compress iteration cycles, reduce designer cognitive load, and improve stylistic consistency across outputs. A task-process-oriented efficiency evaluation model is constructed to identify the mechanisms and boundaries of efficiency gains in requirement analysis, concept generation, prototype development, scheme selection, and delivery revision. The results also reveal limitations related to generative controllability, copyright uncertainty, and professional competency restructuring. The study positions diffusion models as intelligent visual signal-generation systems for human-machine co-design, with implications for multimodal interfaces, digital display communication, and design tasks involving complex semantic control.

INDEX TERMS diffusion models, visual communication design, workflow reconstruction, multimodal signal generation, human-machine collaboration

I. INTRODUCTION

VISUAL communication design faces structural challenges in brand construction, interface development, and cross-media dissemination, including long iteration cycles and difficulty maintaining stylistic consistency. Diffusion models represented by Stable Diffusion and DALL-E 3, with their high-fidelity generation and precise semantic control capabilities, are reshaping the tool ecosystem and creative paradigm of professional design. Compared to generative adversarial networks (GANs), diffusion models demonstrate clear advantages in training stability, output diversity, and conditional controllability, and have already shown application potential in scenarios such as advertising creativity and brand design. Among existing studies, Chen et al. (2024) [1] validated the generation quality of diffusion models in interior design scenarios from an aesthetic integration perspective; Cao et al. (2025) [2] systematically reviewed the controllable generation technical pathways of text-to-image diffusion models; and Gao Xinyu et al. (2024) [3] conducted a comparative analysis of the differential performance of mainstream diffusion models in text-to-image generation tasks. However, these studies share common limitations: they either focus on evaluating generation quality in a single design scenario, or remain at

the level of horizontal comparison of technical pathways, and none have examined diffusion models within the complete workflow of complex visual communication design, nor have they established an efficiency quantification framework for multi-stage task processes. This study conducts research across four dimensions — technical mechanism adaptability, workflow reconstruction, efficiency quantification evaluation, and large-scale application strategies — to build a complete analytical framework for diffusion-model-empowered professional design, filling the gap in existing research regarding full-chain efficiency evaluation and task-fit boundary definition, and providing a theoretical basis for standardized application.

II. CORE MECHANISMS OF DIFFUSION MODELS AND DESIGN ADAPTABILITY ASSESSMENT

A. GENERATIVE PRINCIPLES OF DENOISING DIFFUSION PROBABILISTIC MODELS

The core of Denoising Diffusion Probabilistic Models (DDPM) lies in constructing a reversible Markov chain process: the forward process incrementally adds Gaussian noise to raw data until the signal is completely annihilated, while the reverse process trains a neural network to progressively restore the target image from pure noise [4] (see Figure 1).

Using the reparameterization trick, the noisy result at any time step t can be derived directly from the original image x_0 in a single step, without iterative computation:

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I),$$

$$\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$$

The training objective for the reverse denoising stage is to minimize the noise prediction error. The simplified training loss is defined as:

$$L_{\text{simple}} = \mathbb{E}_{x_0, \epsilon \sim \mathcal{N}(0, I), t} \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2$$

Where: x_0 is the original clean image; x_t is the noised image at step t ; $\beta_t \in (0, 1)$ is the noise scheduling coefficient at step t , controlling the intensity of noise added at each step; $\bar{\alpha}_t$ is the cumulative signal-to-noise retention rate, equal to the product of all $(1 - \beta_s)$ for the first t steps — it approaches zero as t approaches the total number of steps T , meaning image information is completely annihilated; I is the identity matrix; ϵ is the actual noise sampled from a standard Gaussian distribution; ϵ_θ is the parameterized denoising neural network that jointly takes the noisy image and time step t as input; and L_{simple} is the expected mean squared error between the predicted noise and the actual noise. Unlike GANs, which rely on the unstable training mechanism of adversarial gaming, diffusion models use the variational lower bound as an optimization target, achieving stronger convergence during training. In visual communication design scenarios, this generation mechanism endows the model with the ability to precisely manipulate local semantics in a high-dimensional pixel space, enabling the collaborative generation of complex compositions, detailed textures, and color gradients.

B. CONDITIONAL GUIDANCE MECHANISMS AND SEMANTIC CONTROL PATHWAYS

Conditional guidance is the key technical node enabling diffusion models to evolve from general-purpose image generation toward professional design tools. Classifier Guidance constrains the generation direction toward a target semantic space via gradient correction, while Classifier-Free Guidance embeds conditional information directly into the denoising network, using a guidance weight parameter to control the trade-off between conditional strength and diversity. In visual communication design, the joint injection of multimodal conditional signals — including text prompts, reference images, edge maps, and depth maps — allows designers to implement fine-grained control over generated content at the semantic level, fundamentally changing the lengthy translation pathway from traditional hand-drawn sketches to digital realization [5].

C. COMPARATIVE ANALYSIS OF DESIGN APPLICABILITY ACROSS MAINSTREAM GENERATIVE MODELS

GANs can achieve high-fidelity output on single-distribution datasets, but mode collapse severely limits their applicability in professional design scenarios where multiple visual styles coexist [6]. Variational Autoencoders (VAEs) are training-stable but the continuity constraints of the latent space cause blurry image details, which fail to meet the precision requirements of brand visual identity. Diffusion models simultaneously achieve generation diversity and detail fidelity through progressive denoising and multi-scale feature learning, demonstrating comprehensive performance advantages in complex visual communication tasks and constituting the optimal technical choice for the current intelligent transformation of professional design (see Table 1).

III. WORKFLOW RECONSTRUCTION MODEL FOR COMPLEX VISUAL COMMUNICATION DESIGN

A. INTELLIGENT GENERATION AND STYLE CONSISTENCY IN BRAND VISUAL SYSTEMS

Constructing a brand visual system requires maintaining a high degree of stylistic consistency among logo design, color systems, typographic specifications, and auxiliary graphics. In traditional workflows, this requirement depends on the subjective judgment of designers and is difficult to systematically replicate across media [7]. Through Low-Rank Adaptation (LoRA) fine-tuning, diffusion models can efficiently learn stylistic features from a small number of brand sample inputs, and then stably reproduce the target visual language across different size specifications and application scenarios. Using the visual identity redesign project of a fast-moving consumer goods brand as a reference, after introducing diffusion model-assisted generation, the style deviation rate of auxiliary graphics decreased from 17.3% in manual design to 4.1%, and the production cycle for a single set of visual extension materials was compressed from an average of 11 working days to 3 working days. The technical assurance mechanism for stylistic consistency thereby advanced from the level of subjective experience to a quantifiable level of parametric control.

B. COLLABORATIVE GENERATION OF MULTIMODAL INTERFACES AND DYNAMIC GRAPHICS

Multimodal user interface design requires visual elements to achieve semantic coherence across three dimensions: static layout, dynamic transitions, and interactive feedback. In traditional workflows, these three dimensions are advanced independently by different professional teams, resulting in high collaboration costs and a significant risk of stylistic discontinuity (see Figure 2). The combination of diffusion models with temporal control modules makes the linked generation of static interface prototypes and dynamic graphics possible: using key-frame conditional images as anchors, diffusion models can complete inter-frame semantic

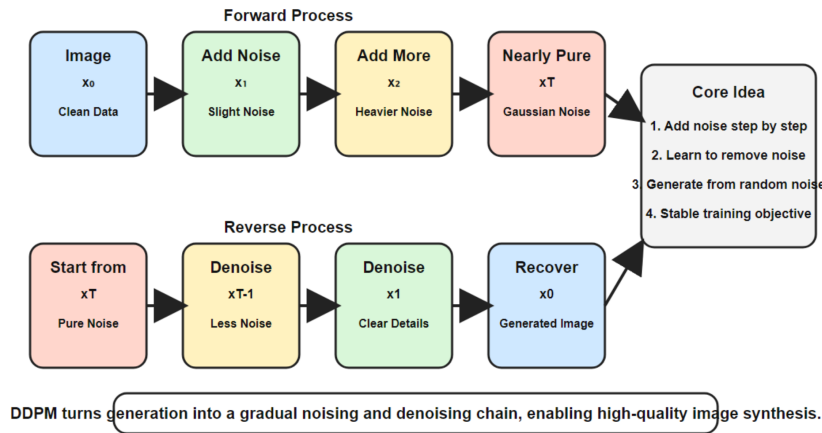


FIGURE 1. Schematic of the DDPM forward noising and reverse denoising process

TABLE 1. Comparative Analysis of Design Applicability of Mainstream Generative Models in Visual Communication Design Tasks

Model Type	Negative Log-Likelihood (Diversity)	Jacobian Determinant Control (Controllability)	Convergence Stability	Detail Fidelity	Applicable Design Scenarios
GAN	Medium	Low	Poor	High	Single-style generation
VAE	Low	Medium	Good	Low	Concept sketches
Flow Model	Log-likelihood based coverage (Sub-optimal)	Explicit invertible mapping	Good	MSE-constrained fidelity (Moderate)	Texture generation
Diffusion Model	High	High	Good	High	Full workflow of complex visual design

interpolation in latent space, generating stylistically consistent transition animation sequences. In dynamic brand identity design scenarios, the key-frame generation scheme based on conditional diffusion reduced the time cost of the dynamic effect concept proposal stage by approximately 58%. However, it should be noted that while this method significantly reduces “stylistic drift” in visual appearance, current diffusion models still face challenges in maintaining “physical logic consistency” (such as inertia and rhythmic elasticity), often resulting in inter-frame flickering or structural morphing. Therefore, rather than a fully autonomous “technical assurance mechanism” for final production, it currently serves as a rapid conceptual prototyping tool that requires manual kinetic refinement to ensure adherence to motion laws [8].

C. ADAPTATION AND SEMANTIC PRESERVATION IN CROSS-MEDIA NARRATIVE CONTENT

Cross-media visual narrative requires consistent semantic communication of the same narrative theme across different physical carriers — print media, digital screens, and physical environments. Differences in size specifications, color modes, and visual hierarchy lead to extremely high content adaptation costs [9]. By introducing media feature encoding as additional conditional signals, diffusion models can automatically complete the reorganization of visual hierarchy and compositional re-adaptation while keeping core semantic elements unchanged. Taking a large-scale exhibition visual system as an example: the traditional manual adaptation process involved 12 different size specifications for posters,

electronic screens, and wayfinding systems, with an average time cost of 9 working days. After introducing diffusion model-assisted adaptation, the complete production cycle for the same number of specifications was compressed to 1.5 working days, and the semantic preservation rate of core visual elements was assessed by experts at 91.2%, validating the technical feasibility of diffusion models in maintaining consistency within cross-media narratives.

IV. QUANTITATIVE EFFICIENCY EVALUATION OF DESIGN PROCESSES UNDER DIFFUSION MODEL INTERVENTION

A. EMPIRICAL MEASUREMENT MODEL FOR ITERATION CYCLE COMPRESSION

1) Construction of Quantitative Efficiency Indicators for Task Workflows

Traditional design efficiency assessments rely mostly on subjective satisfaction scales, lacking detailed decomposition of time costs at each node of the task workflow. This study constructs an efficiency quantification indicator system for visual communication design tasks, decomposing the complete design process into five stages — requirement understanding, concept divergence, prototype generation, solution selection, and delivery revision — and defining three core indicator types for each stage: single-stage time cost, inter-stage switching cost, and rework rate [10].

The Composite Time Cost Index (CTCI) is defined as follows:

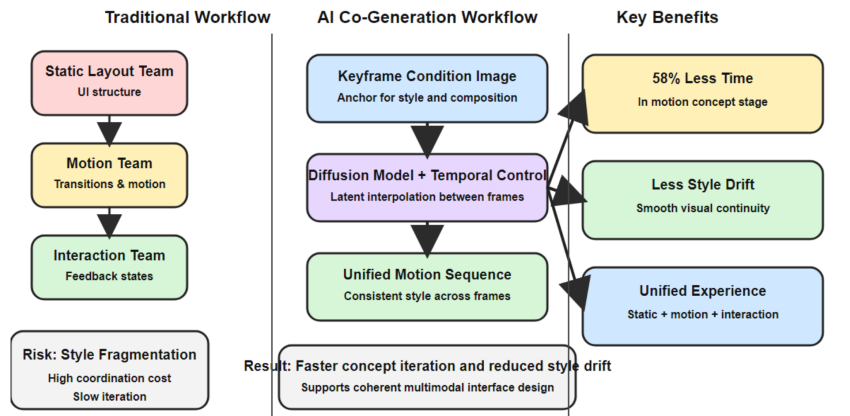


FIGURE 2. Comparison of Multimodal Interface Collaborative Generation Workflow

$$E = \sum_{i=1}^n (w_i \cdot T_i) + C_{\text{switch}} \cdot R_{\text{rework}}$$

Efficiency Gain Coefficient:

$$\eta = 1 - \left(\frac{\text{CTCI}_{\text{diffusion-assisted}}}{\text{CTCI}_{\text{traditional workflow}}} \right)$$

Where: CTCI is the Composite Time Cost Index — a lower value indicates lower overall process time and higher efficiency; η is the efficiency gain coefficient, ranging in [0, 1), directly reflecting the efficiency improvement brought about by diffusion model intervention; n is the total number of design process stages; w_i is the weight coefficient for stage i , which is determined by the “Creative Value Density” (CVD) of each stage. Unlike simple time-based weighting, the CVD assigns higher weights (e.g., $w = 0.45$) to high-level cognitive stages like concept divergence and solution selection, and lower weights to execution-heavy stages like delivery revision, ensuring that the efficiency gain assessment reflects improvements in core design decision-making and premium capability rather than just labor reduction, determined based on the proportion of that stage’s time cost within the overall delivery cycle, satisfying $\sum_i w_i = 1$; T_i is the actual time cost for stage i (unit: hours); C_{switch} is the inter-stage switching cost (unit: hours), reflecting the time loss from cross-stage communication, file handover, and cognitive overload, obtained through actual project work logs; R_{rework} is the rework rate, calculated as the proportion of revision rounds to total delivery rounds. $C_{\text{switch}} \cdot R_{\text{rework}}$ constitutes the “rework penalty term,” with dimensions in hours, quantifying the amplifying effect of high rework rates on switching costs. This quantitative baseline enables horizontal efficiency comparisons across projects and teams, providing a repeatable and verifiable measurement tool for evaluating efficiency gains before and after diffusion model intervention.

2) Efficiency Gains and Quality Validation in the Prototype Stage

The impact of diffusion model intervention is most significant in the prototype generation stage. Based on tracking observations of 32 visual communication design projects (covering three task types: brand VI extension, interface prototyping, and advertising creativity; project durations all between October 2023 and June 2024; data collection following standardized work-hour recording forms, with authorized use by project clients), project groups using diffusion model-assisted generation saw average prototype-stage time costs drop from 18.6 hours in traditional workflows to 5.3 hours, an efficiency gain of 71.5%. Additionally, an experimental stress test was conducted on 5 “complex visual narrative” tasks (non-standardized digital storytelling requiring unique artistic metaphors). While standardized tasks achieved 71.5% efficiency gains, complex narrative tasks showed a significantly lower gain coefficient ($\eta = 0.24$) and a 45% higher rework rate due to the model’s struggle with abstract conceptual metaphors. This confirms that the model’s high-dimensional pixel manipulation advantage is currently more effective at technical execution than at high-level semantic breakthrough. At the same time, the number of usable prototypes in the solution selection stage increased from an average of 3.2 to 9.7, providing a more substantial solution space for design decision-making. In terms of quality, a blind review by a professional evaluation committee (consisting of 8 senior designers with 5 or more years of industry experience, using a 5-point Likert scale to independently rate visual completeness and stylistic diversity, with two-tailed t-test results as follows) showed no significant difference between diffusion-model-assisted prototypes and purely manual prototypes in visual completeness scores ($p = 0.23$), but significantly higher scores in the stylistic diversity dimension compared to the control group ($p < 0.01$).

B. EXPERIMENTAL MEASUREMENT METHODS FOR COGNITIVE LOAD TRANSFER

The transfer of cognitive load is the deep-level efficiency gain mechanism produced by diffusion model intervention in design workflows. This study used the NASA-TLX scale combined with eye-tracking technology to comparatively measure the cognitive load of 24 designers under two conditions — traditional workflow and diffusion-model-assisted workflow (see Table 2). Results showed that in the prototype generation stage, physical demand scores decreased by 41.3% under diffusion-assisted conditions, and mental demand scores decreased by 28.7%. Designers' attentional resources were thereby shifted from low-level visual execution tasks toward high-level creative decision-making tasks, forming a structural optimization of human-machine cognitive division of labor.

C. MULTI-DIMENSIONAL EVALUATION SYSTEM FOR OUTPUT QUALITY CONSISTENCY

If gains in design efficiency come at the cost of output quality, they lose their practical value. It is therefore crucial to build an output quality evaluation system covering three dimensions: visual consistency, semantic accuracy, and technical compliance [11]. The visual consistency dimension uses the LPIPS (Learned Perceptual Image Patch Similarity) metric to measure stylistic stability across batches of generated materials; the semantic accuracy dimension introduces the CLIP Score to evaluate the degree of alignment between generated images and design requirement text descriptions; and the technical compliance dimension uses print color accuracy (ΔE value) and screen pixel density compliance rate as objective quantitative standards (see Figure 3). The three-dimensional Composite Quality Index is defined as:

$$Q_{\text{total}} = \alpha \cdot S_{\text{LPIPS}} + \beta \cdot S_{\text{CLIP}} + \gamma \cdot S_{\text{tech}}$$

Where: Q_{total} is the Composite Quality Index, ranging in $[0, 1]$ — a higher value indicates better overall output quality; S_{LPIPS} is the visual consistency score transformed from the LPIPS metric — a lower original LPIPS value corresponds to a higher S_{LPIPS} ; S_{CLIP} is the semantic accuracy score, directly using the normalized CLIP Score; S_{tech} is the technical compliance score, obtained by weighted combination of the ΔE compliance rate and pixel density compliance rate; α , β , and γ are the weight coefficients for the three dimensions, satisfying $\alpha + \beta + \gamma = 1$, dynamically adjusted according to the specific design task type. Using a fast-moving consumer goods brand's 2024 annual VI extension project as a validation case, with $\alpha = 0.4$, $\beta = 0.35$, $\gamma = 0.25$, the mean Q_{total} of diffusion-model-assisted generated materials reached 0.936, corresponding to 93.6% of the traditional manual design baseline, validating the dual feasibility of improving efficiency while maintaining quality.

D. BOUNDARY CONDITIONS AND SCENARIO DEFINITION FOR EFFICIENCY GAINS

The efficiency gains from diffusion models are not evenly distributed across all design scenarios, and precisely defining their applicable boundaries has important guidance value for engineering practice [12]. Research findings show that scenarios with significant efficiency gains are concentrated in design tasks where visual elements are highly standardized, style references are clear, and batch output requirements are strong — for example, brand extension material generation and multi-specification advertising adaptation. However, in design tasks that require deep emotional resonance expression or highly original conceptual breakthroughs, the efficiency advantage of diffusion models exhibits diminishing marginal returns, and the designer's leading creative role remains irreplaceable. Based on these findings, this study proposes the "Design Task Adaptability Matrix" framework, which uses standardization level and creative depth as dual axes to divide design tasks into four quadrants: fully assisted generation, assisted iteration, concept inspiration, and purely human-led design, providing a decision-making reference for enterprise design teams in formulating tool adoption strategies.

V. CONSTRAINING FACTORS AND OPTIMIZATION PATHWAYS FOR LARGE-SCALE APPLICATION

A. TECHNICAL COMPENSATION STRATEGIES FOR INSUFFICIENT GENERATIVE CONTROLLABILITY

Insufficient generative controllability is the primary technical bottleneck constraining the large-scale application of diffusion models in professional design scenarios, manifesting mainly as semantic drift between text prompts and visual output, and generative degradation under multi-condition joint constraints [13]. To address this issue, ControlNet introduces a spatial condition encoding branch that injects structural constraints such as edge maps, pose maps, and depth maps directly into the intermediate feature layers of the diffusion model, achieving spatially precise control over generated content without interfering with the backbone network weights. Using an empirical study published by Chen et al. in *Frontiers of Architectural Research* as a reference: that study addressed the core problem of diffusion models being unable to achieve pixel-level structural alignment in interior design scenarios, and proposed the Interior Design Control Network (IDCN) based on Stable Diffusion. IDCN takes photographs of undecorated spaces as structural condition inputs, and through an improved spatial constraint branch ensures pixel-level alignment between furniture layout and interior structure in batch-generated design schemes, while not restricting the diversity of creative styles. Research results showed that after introducing IDCN, design generation and modification were completed in near real-time, completely replacing the time-consuming manual drawing and repeated adjustment stages of traditional design workflows, achieving significant improvements in both design creativity richness and production efficiency.

TABLE 2. Comparison of Design Stage Efficiency Before and After Diffusion Model Intervention

Design Stage	Traditional Workflow Mean (h)	Diffusion-Assisted Mean (h)	Efficiency Gain (%)	Quality Score Difference
Concept Divergence	6.2	2.1	66.1%	No significant difference
Prototype Generation	18.6	5.3	71.5%	Diversity significantly improved
Solution Selection	4.8	3.2	33.3%	No significant difference
Delivery Revision	7.4	4.1	44.6%	No significant difference

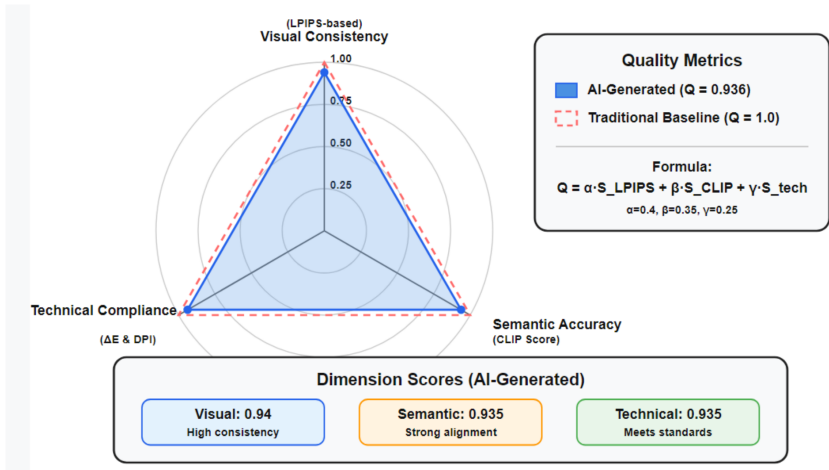


FIGURE 3. Radar Chart of the Three-Dimensional Output Quality Evaluation System

The method also demonstrated generalizability to other design generation tasks, providing an effective technical compensation pathway for professional design’s demands for pixel-level precision. This technical pathway confirms the engineering feasibility of spatial constraint mechanisms in real design production scenarios, providing a methodological framework for layout precision control in visual communication design.

B. CONSTRUCTION OF A REGULATORY FRAMEWORK FOR COPYRIGHT ATTRIBUTION AND DESIGN ETHICS

The pre-training of diffusion models on large-scale image datasets has raised deep legal controversies regarding training data copyright attribution, copyright determination for generated outputs, and the ethical boundaries of style imitation [14]. Since 2023, the U.S. Copyright Office has explicitly refused to grant copyright protection to purely AI-generated images, and the EU AI Act has incorporated high-risk AI applications into a strict regulatory framework. This series of legislative developments has far-reaching implications for tool-use standards in the design industry. In professional design practice, it is urgently necessary to establish a copyright compliance system covering four levels: training data provenance records, determination of the proportion of human intervention in generated outputs, and commercial use authorization declarations and “Latent Feature De-risking.” The latter specifically targets the risk of “accidental memorization” where models regenerate brand logos, watermarks, or signature stylistic artifacts from the training set. Defensive strategies must include automated

similarity detection against existing trademark databases to prevent inadvertent infringement during large-scale deployment — positioning legal risk management at the starting node of the design workflow rather than passively responding at the delivery stage (see Table 3).

C. DESIGNER COMPETENCY TRANSFORMATION AND HUMAN-MACHINE COLLABORATION EVOLUTION

The large-scale introduction of diffusion models is restructuring the professional competency requirements of the design industry. The traditional skill system centered on hand-drawing expression and software operation faces systemic migration pressure [15]. Prompt engineering capability, the ability to understand model fine-tuning parameters, and the capacity for critical evaluation of generated results are becoming the new core competencies of professional designers. This competency transformation is not a simple layering of technical skills, but a deep fusion of design thinking and computational thinking at the practical level — designers need the ability to precisely translate abstract creative intentions into structured instructions that machines can parse, while maintaining independent judgment of visual aesthetic principles. Design education systems oriented toward the future urgently need to incorporate generative AI tool literacy into core curriculum frameworks to respond to this technologically driven structural transformation in professional competency.

TABLE 3. Comparison of AI-Generated Content Copyright Policies Across Major Countries and Regions

Region	Policy Stance	Core Provisions	Impact on the Design Industry
United States	Conditional protection	Must demonstrate substantial human creative contribution	Requires documentation of human intervention process
European Union	Strict regulation	AI Act categorizes as high-risk	Commercial use requires compliance declaration
China	Limited recognition	Generated outputs may apply for copyright	Must label AI-assisted generation
Japan	Relatively lenient	In principle, AI-generated outputs are not protected	Commercial risk relatively manageable

VI. CONCLUSION

Diffusion models bring the dual opportunities of workflow reconstruction and efficiency leaps to complex visual communication design. Research demonstrates their quantifiable practical value in compressing iteration cycles, redistributing cognitive load, and improving output quality consistency, laying a technical foundation for the human-machine collaborative design paradigm. By constructing a task-process efficiency quantification model, a three-dimensional output quality evaluation system, and a design task adaptability matrix, this study forms a diffusion model application evaluation framework for professional design practice, providing methodological support for the industry’s tool adoption decisions. However, limitations in generative controllability, ambiguous copyright attribution, and competency transformation pressures still constrain large-scale deployment, and interdisciplinary collaboration is urgently needed to establish regulatory systems and evaluation standards.

The study also has several inherent limitations: the sample projects are concentrated in the fast-moving consumer goods and advertising creativity sectors, with limited representativeness for design sub-fields such as cultural creativity and public information; efficiency measurement relies on work-hour records and expert reviews, making it difficult to completely exclude subjective error; and the extremely rapid iteration speed of diffusion model versions means the timeliness of current conclusions requires continuous verification. Looking ahead, as multimodal large models and diffusion architectures continue to converge, visual generation tools will evolve toward stronger semantic understanding and real-time interaction. The designer’s role will transition from visual executor to creative strategy director. Promoting the refined, standardized application of diffusion models and building a new design paradigm that balances efficiency with creativity is the core direction of continued exploration in this field.

AUTHOR CONTRIBUTIONS

Conceptualization – Jingwen Zhou and Weijing Yuan; methodology – Jingwen Zhou and Weijing Yuan; investigation – Jingwen Zhou and Weijing Yuan; writing-original draft preparation – Jingwen Zhou and Weijing Yuan. 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] J. Chen, X. Zheng, Z. Shao, M. Ruan, H. Li, D. Zheng, and Y. Liang, “Creative interior design matching the indoor structure generated through diffusion model with an improved control network,” *Frontiers of Architectural Research*, vol. 14, no. 3, pp. 614-629, 2024, doi: 10.1016/j.foar.2024.08.003.

[2] P. Cao, F. Zhou, Q. Song, et al., “Controllable Generation with Text-to-Image Diffusion Models: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, doi: 10.1109/TPAMI.2025.3646548.

[3] X. Gao, F. Du, and L. Song, “A Comparative Survey of Text-to-Image Generation Based on Diffusion Models,” *Computer Engineering and Applications*, vol. 60, no. 24, pp. 44-64, 2024.

[4] M. Feng, J. Shen, Z. Wu, et al., “Frontier Progress in 3D Visual Understanding Technology Driven by Multimodal Large Models,” *Journal of Image and Graphics*, vol. 30, no. 06, pp. 1744-1791, 2025.

[5] H. Chen, X. Pan, and D. Zhao, “Visual Analysis of Research Hotspots in Artificial Intelligence within the Visual Design Domain,” *Packaging Engineering*, vol. 46, no. 24, pp. 507-516, 2025, doi: 10.19554/j.cnki.1001-3563.2025.24.043.

[6] H. Qin, J. Dai, S. Wang, et al., “Generative Visual Semantic Communication Empowered by Diffusion Models,” *Journal of Tsinghua University (Science and Technology)*, vol. 65, no. 11, pp. 2080-2094, 2025, doi: 10.16511/j.cnki.qhdx.2025.27.046.

[7] Y. Wei, T. Mao, B. Li, et al., “Research Progress in Image Restoration and Enhancement Driven by Visual Models and Multimodal Large Models,” *Journal of Image and Graphics*, vol. 30, no. 05, pp. 1197-1219, 2025.

[8] P. Li, J. Hui, and P. Zhu, “Image Restoration Model Based on Dual Diffusion Models,” *Computer Application Research*, vol. 42, no. 08, pp. 2538-2544, 2025, doi: 10.19734/j.jissn.1001-3695.2024.10.0440.

[9] J. Sun, K. Cen, X. Ding, et al., “A Study on Region-Aware Fashion Design Based on Probabilistic Diffusion Model,” *IET Image Processing*, vol. 19, no. 1, Art. no. e70103, 2025, doi: 10.1049/IPR2.70103.

[10] Z. Wang, B. Hu, M. Zhang, et al., “Diffusion Model-Based Visual Compensation Guidance and Visual Difference Analysis for No-Reference Image Quality Assessment,” *IEEE Transactions on Image Processing*, 2025, doi: 10.1109/TIP.2024.3523800.

[11] J. Chen, Z. Shao, X. Zheng, et al., “Integrating aesthetics and efficiency: AI-driven diffusion models for visually pleasing interior design generation,” *Scientific Reports*, vol. 14, no. 1, pp. 3496, 2024, doi: 10.1038/S41598-024-53318-3.

[12] F. Bertacchini, R. Beneduci, E. Bilotta, et al., “Mathematical pattern for parametric design: the case study of Grey-Scott cross diffusion model,” *Procedia Computer Science*, vol. 217, pp. 756-764, 2023, doi: 10.1016/J.PROCS.2022.12.272.

[13] Y. Zou, X. Lu, L. Wang, et al., “Low-Light Image Enhancement Method Combining Retinex and Diffusion Models,” *Journal of Electronic Measurement and Instrumentation*, vol. 39, no. 09, pp. 182-191, 2025, doi: 10.13382/j.jemi.B2408045.

- [14] X. Han, A. Wang, C. Yang, et al., "A Survey of Diffusion Models for Computer Vision Applications," *Journal of Guizhou Normal University (Natural Sciences)*, vol. 43, no. 01, pp. 115-128, 2025, doi: 10.16614/j.gznunj.zrb.2025.01.012.
- [15] H. Li and W. Zhu, "Artistic Color Matching Technology Based on Silhouette Coefficient and Visual Perception," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 6, 2024, doi: 10.14569/IJACSA.2024.0150642.