

PAPER

Mobile AIGC Image Generation Interactive Training Model for Design Education

Cheng Wang  Hubei Institute of Fine Arts,
Wuhan, Chinawangcheng@hifa.edu.cn**ABSTRACT**

In the field of design education, the current AI-generated content (AIGC) tools commonly face key issues such as poor adaptation to mobile platforms, disconnection between interaction logic and educational objectives, and the inability to support creative iteration with generated results. To address these challenges, this paper proposes a mobile AIGC image generation interactive training model tailored for design education scenarios. The research first constructs a three-layer architecture that includes multimodal interactive input, lightweight model generation, and educational feedback iteration, based on the core needs of creative inspiration, solution iteration, and professional feedback in design education. Then, through techniques such as knowledge distillation, quantization compression, and operator optimization, the Stable Diffusion model is improved and adapted for mobile platforms, achieving a balance between generation quality and resource consumption. Finally, an interactive mechanism is designed based on constructivist educational theory, supporting professional design inputs such as sketches, text, and color references, and incorporating design guidelines and creativity-expanding modules. To evaluate the model's performance, mobile devices of various specifications (low, medium, and high-end) are tested in three typical educational scenarios: 1) graphic design, 2) product design, and 3) user interface (UI) or user experience (UX) design. Comparison experiments are conducted with existing mobile AIGC models, such as mobile stable diffusion and lite generative adversarial network (GAN), using performance indicators such as Fréchet inception distance (FID), inception score (IS), educational adaptability score, and resource consumption metrics. The results show that the proposed model outperforms existing models in core performance metrics, with an average generation speed improvement of 42.3%, an educational objective match score of 8.7/10, and a 31.6% reduction in mobile memory usage. The findings provide effective technical support for creative generation, solution iteration, and personalized teaching in design education.

KEYWORDS

design education, AI-generated content (AIGC), image generation model, interactive training, lightweight model

Wang, C. (2026). Mobile AIGC Image Generation Interactive Training Model for Design Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(2), pp. 63–77. <https://doi.org/10.3991/ijim.v20i02.60137>

Article submitted 2025-10-11. Revision uploaded 2025-11-28. Final acceptance 2025-11-30.

© 2026 by the authors of this article. Published under CC-BY.

1 INTRODUCTION

In recent years, AI-generated content (AIGC) has made leapfrog developments in fields such as graphic design, product design, and user interface (UI) and user experience (UX) design [1, 2]. Core technologies such as diffusion models and GAN have become key supports for creative generation and solution optimization [3, 4]. However, their deployment generally relies on cloud computing power, making it difficult to meet the creative recording and iteration needs in design education scenarios, creating a gap between technological applications and educational contexts. Traditional design education has long faced practical issues such as high barriers to creative expression, low efficiency in solution iteration, and insufficient personalized guidance resources [5–7]. Although current mainstream AIGC tools possess basic generation capabilities, they lack targeted adaptation to the core processes of design education, with insufficient support for key teaching processes such as sketch conversion, style transfer, and design guidelines [2, 8]. The mobile versions of these tools often suffer from poor generation quality and complex interaction logic, making it difficult for them to be integrated into regular teaching processes.

Existing research in this field has formed three main strands but still has significant limitations. In the field of AIGC image generation technology, the technological evolution of diffusion models such as Stable Diffusion and DALL-E3, and GAN such as StyleGAN and ProGAN has driven breakthroughs in design applications. However, due to insufficient controllability of generation and difficulties in deploying models on mobile devices, these technologies cannot meet the educational requirements for the interpretability and real-time nature of generated results [9–11]. Research on AI applications in design education is mostly focused on single-function implementations such as AI-assisted sketch recognition and design style transfer, lacking systematic support for the entire educational process of creative inspiration, solution iteration, and professional feedback. There is insufficient synergy between technical tools and educational objectives [1, 12]. Research on mobile AI model optimization has formed a compression technology system involving pruning and quantization, and mobile inference frameworks such as TensorFlow Lite and PyTorch Mobile have gradually matured. However, existing lightweight models generally focus on speed improvement and resource consumption reduction, neglecting the adaptability of generation results to educational scenarios. There are significant gaps in feature retention and matching with educational needs in the design field [13, 14]. In summary, existing research has not constructed a collaborative framework for technology miniaturization, interactive education, and application scenarios, which cannot meet the core needs of mobile AIGC tools in design education. This research gap provides the core entry point for this study.

In response to this gap, this study aims to construct a mobile AIGC image generation interactive training model suitable for design education scenarios, addressing the three key problems of low-generation quality on the mobile end, disconnection between interaction logic and educational goals, and difficulty supporting creative iteration. On the theoretical level, the research enriches the theoretical system of mobile AIGC technology in the field of education, clarifies the core coupling mechanism between technology miniaturization and educational scenario adaptation, and improves the interdisciplinary integration framework of technology-education-design, providing theoretical references for cross-field technology implementation. On the practical level, the research results can provide lightweight, highly adaptable design education AIGC tools, enhancing teaching efficiency and the quality of student

creative output by optimizing creative generation and solution iteration processes, and providing concrete technical support for the digital transformation of design education. Accordingly, the innovative aspects of the research are reflected in three areas: on the architecture level, a three-layer structure is proposed, consisting of a multimodal education interaction layer, lightweight generation model layer, and feedback iteration optimization layer, which achieves the deep coupling of design education processes with mobile AIGC technology for the first time and clarifies the core functions and collaborative logic of each layer; on the interaction level, a multimodal fusion input mode for sketches, text, and color is designed, along with a dual-track feedback mechanism of design guidelines and creativity-expansion guidance, precisely matching the core needs of professional design education; on the optimization level, a design style feature distillation strategy is proposed, which selectively retains key features in the design field during model size compression, significantly improving the educational adaptability of the generated results.

To clearly present the research context, the subsequent chapters of this paper are arranged as follows: Section 2 explains the model design principles, overall architecture, and key interaction mechanisms, clarifying the technical implementation path; Section 3 details the dataset construction, training strategy design, and specific implementation plan for mobile-end optimization; Section 4 verifies the model performance and module effectiveness through comparative experiments and ablation experiments with multiple devices and scenarios; Section 5 discusses the theoretical contributions of the research, differences and limitations compared to existing studies, summarizes the research findings, and suggests directions for future research.

2 MODEL DESIGN AND ARCHITECTURE

2.1 Design principles

The model design is centered on educational scene adaptation, while balancing the constraints of edge-side performance and optimizing the interactive experience, forming a collaborative design principal system. Educational adaptability is the core principle, requiring the model generation logic to deeply align with the complete process of creative inspiration, solution iteration, and professional feedback in design education, ensuring that the technological implementation always serves the teaching objectives and avoids deviation due to technical optimization. The edge-side performance principle specifies quantitative metrics to ensure practical value. For mainstream Android and iOS mobile devices, core metrics are set, including a generation time of no more than five seconds for a 512×512 resolution image and peak memory usage of less than 2GB, balancing generation quality and edge-side resource consumption. The interaction usability principle focuses on design professionals' habits, simplifying operation links while supporting natural input methods such as sketching, text descriptions, and color picking, which align with design creation intuition and reduce the tool learning cost.

2.2 Overall architecture design

Based on the characteristics of design education processes and mobile technology constraints, this study proposes a “three-layer two-loop” overall architecture.

Through the hierarchical collaboration of multimodal interaction, lightweight generation, and feedback iteration, combined with technology and education-driven dual feedback loops, it achieves the unification of technological performance and educational value. The architecture takes lightweight generation as the core technology, multimodal interaction as the educational entry point, and feedback iteration as the optimization path, with the three layers interconnected: The multimodal education interaction layer is responsible for accurately capturing input requirements and feedback information in the design education scene. The lightweight AIGC generation model layer undertakes the core task of efficient generation on the edge side, while the feedback iteration optimization layer realizes dynamic adaptation between the model and the educational scene.

The multimodal education interaction layer contains two core modules: the input module integrates three types of design professional input forms—sketches, text, and color references—through an attention-based feature alignment network, converting heterogeneous modal information into feature vectors with a unified dimension, ensuring effective fusion of multi-source information. The feedback module constructs a “machine-human” dual-source feedback system. The machine feedback provides real-time normative suggestions based on a design knowledge base, while the human feedback collects style adjustments, element optimizations, and other requirements annotated by teachers and students, jointly forming an education-oriented feedback dataset. The lightweight AIGC generation model layer is based on the Stable Diffusion 1.5 framework, with the core U-Net module being reconstructed and replaced with a lightweight U-Net using MobileNetV2 as the backbone network. At the same time, a design style feature distiller and an edge-side inference engine are embedded. The former extracts design domain style features from the cloud-based teacher model to guide student model training, while the latter integrates TensorFlow Lite optimization operators, dynamically selecting GPU, NPU, or other acceleration backends based on device hardware characteristics. The feedback iteration optimization layer achieves dynamic optimization through small-sample fine-tuning and personalized adaptation. Based on the feedback data accumulated in the interaction layer, a small-sample fine-tuning strategy is used to update model parameters, enhancing educational adaptability. By mining style preferences and domain features from users’ historical interaction data, a personalized generation strategy is constructed to realize scene adaptation for “thousands of people, thousands of models.”

The three-layer architecture forms a continuous optimization mechanism through technology, and education-driven dual feedback loops. The technology feedback loop, following the “input-generation-feedback-fine-tuning” chain, converts feedback data from the interaction layer into real-time model parameter update signals, achieving incremental optimization of model performance on the edge. The educational feedback loop, with the core of “student input-model generation-teacher feedback-student iteration,” converts model generation results into creative support and normative guidance in the teaching process, promoting personalized design teaching. The collaborative operation of the dual feedback loops allows the architecture to possess both self-optimization capabilities and deep integration into the dynamic process of design education. Figure 1 presents a schematic diagram of the designed architecture.

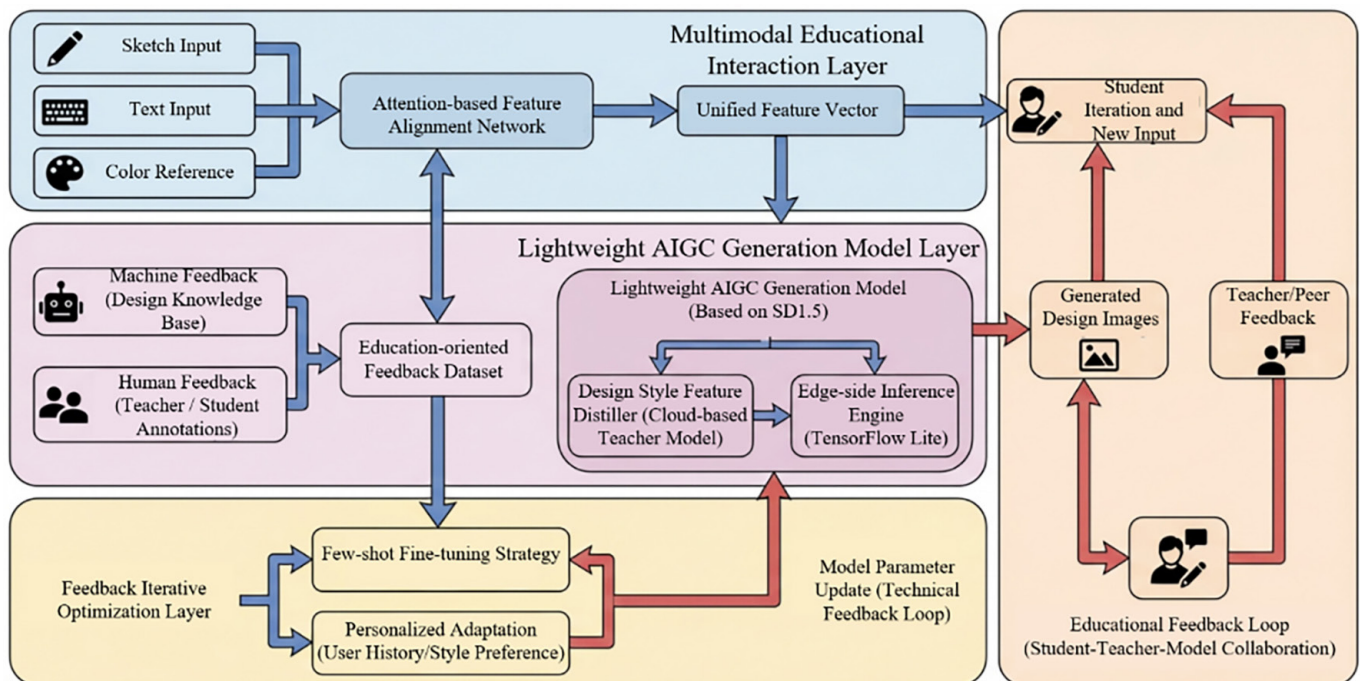


Fig. 1. Schematic diagram of the “Three-Layer Two-Loop” architecture for mobile AIGC in design education

2.3 Key interaction mechanism design

The progressive multimodal input mechanism is constructed based on the natural progression of design creation “from draft to refinement,” breaking the limitations of traditional single-step inputs and achieving progressive collaboration between input and generation. This mechanism supports users to input information in a progressive logic of “sketch draft-text addition-color adjustment,” where after each input step, the model updates only the feature dimensions corresponding to the newly added information, generating intermediate results in real-time. This design not only aligns with the cognitive process of gradually deepening design creativity but also reduces the computational power consumption for real-time generation on the edge through incremental computation. Sketches, as design drafts, provide spatial structure constraints; text additions clarify style and functional requirements; color adjustments refine visual expression. The progressive input of these three aspects gradually focuses the model’s generation goal, enhancing the match between results and design intentions.

The design specification embedding mechanism is supported by the design discipline knowledge base, realizing real-time normative guidance and optimization suggestion output during the generation process. The knowledge base covers core content such as graphic design principles and UI design specifications, which, after structured analysis, are transformed into a computable rule base containing 23 key design indicators such as layout proportion, font readability, and color matching. During model generation, the specification detection module extracts the visual features of the generated results in real-time and compares them with the rule base. When issues such as font size being below the mobile readability threshold or layout not following the golden ratio are detected, the system immediately outputs normative suggestions and provides specific optimization plans, such as suggesting

increasing the font size from 12 pt to 14 pt to meet readability requirements. This mechanism transforms implicit constraints of design norms into explicit guidance, aligning with the core needs of normative teaching in design education.

The creative divergence guidance mechanism is designed to address common creative bottlenecks in design education by generating variants to expand the boundaries of design thinking. When the model detects that the user's input is too singular, it automatically triggers the variant generation module to produce multi-dimensional variant solutions based on the design style library and element library. Style variants cover 12 mainstream design styles, such as minimalist, retro, and futuristic, achieved by adjusting the style feature weight in the generation model. Element variants focus on adjustments of core design elements such as icon replacement, layout reconstruction, and color restructuring, with the most optimal variant combinations selected based on the semantic correlation of design elements. The variant solutions are presented in parallel, with explanations of the differences in style and elements. This not only provides students with diverse creative references but also helps them understand the expression logic of design styles and elements through comparative analysis, synchronizing creative inspiration with design knowledge learning.

3 MODEL TRAINING AND OPTIMIZATION

3.1 Dataset construction and preprocessing

To ensure the professionalism and educational adaptability of the dataset, this study adopts a combination of public datasets and self-built educational datasets, forming a dataset system covering multiple design fields and educational scenarios. The public datasets focus on authoritative subsets in the design field, including COCO-Design, which provides a wealth of finished design image samples, UIUCSketchDataset, which supplements sketch and finished product paired data, and PinterestDesignCollection, which covers samples of diverse design styles. A total of 82,000 valid samples were selected from these three datasets. The self-built dataset, named the Design Education AIGC Dataset, was collected in collaboration with the design schools of three universities. It includes student design assignments' sketch and final draft paired data, teacher feedback annotation texts for assignments, and design discipline specification documents. This results in a high-quality educational dataset of 50,000 samples. The two types of datasets complement each other in sample types, ensuring that the data distribution aligns with actual design education scenarios.

Dataset annotation follows a dual-annotation model involving design field experts and professional teachers, with cross-validation ensuring annotation consistency. The annotation dimensions are set around educational adaptability and design professionalism, including design field, style type, educational adaptability, and specification compliance. Educational adaptability focuses on the support ability of generated results for creative inspiration, while specification compliance corresponds to the basic specification requirements of design disciplines. Data preprocessing implements differentiated strategies for different modal features: sketch data undergoes rotation, scaling, and Gaussian noise addition to enhance the model's robustness to sketch drawing differences; text data is expanded using synonym replacement and sentence rephrasing to avoid generation bias due to singular demand descriptions. All image samples are resized to 512×512 resolution and normalized to the range [0, 1], while text data is encoded into fixed-length feature vectors using the CLIP Tokenizer.

3.2 Training strategy design

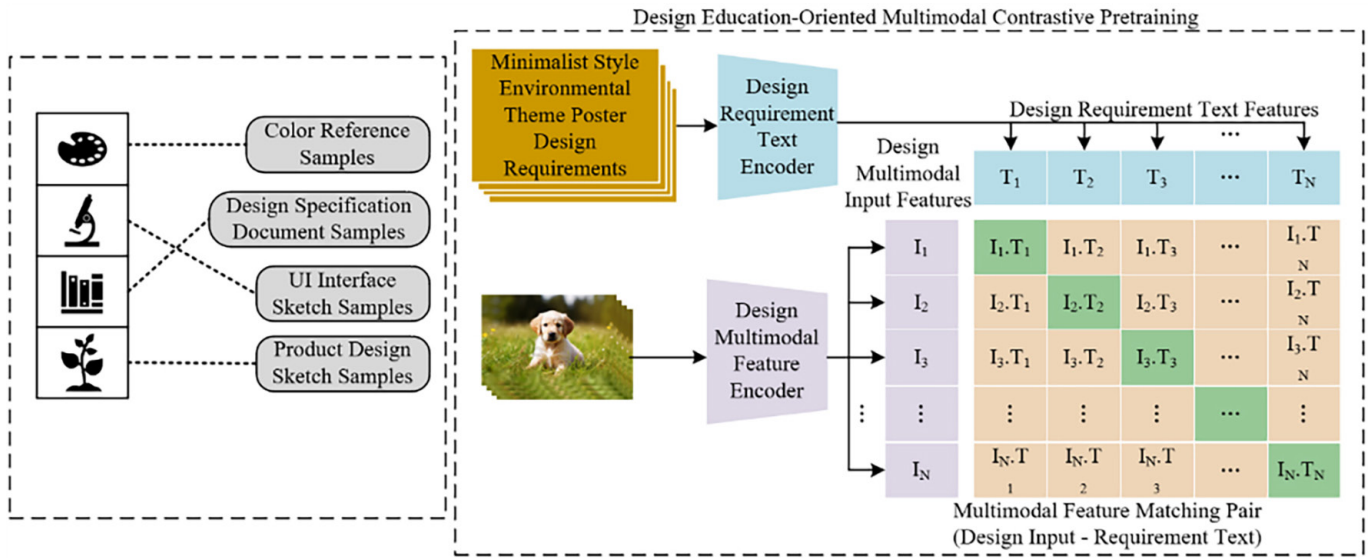


Fig. 2. Principle of mobile AIGC multimodal interactive pre-training model for design education

To achieve deep adaptation between model performance and edge-side deployment requirements, this study designs a progressive training strategy of “cloud-side pre-training-distillation adaptation-edge-side fine-tuning,” with each stage forming a complete training process through goal focus and parameter collaboration. Figure 2 presents the principle of the mobile AIGC multimodal interactive pre-training model for design education. By using a design requirement text encoder and multimodal feature encoder, the model matches features from color, specifications, sketches, and other multimodal inputs and requirement texts in the design education scene, laying the foundation for model adaptation to design education scenarios. The pre-training phase is conducted on a cloud-side server equipped with an NVIDIA A100 GPU, using a dataset composed of the DEA-Dataset and public datasets. The teacher model is built based on Stable Diffusion 1.5 as the framework. The training process uses the AdamW optimizer, with an initial learning rate of 5e-5, a weight decay coefficient of 1e-4, and a training cycle set to 100 epochs. An early stopping mechanism is introduced—when the educational adaptability score on the validation set does not significantly improve for 5 consecutive epochs, training is terminated to avoid overfitting and ensure training efficiency.

The distillation phase uses the pre-trained teacher model as the knowledge transfer entity, training a lightweight student model suitable for edge-side deployment. The core task is to construct a multi-objective loss function to accurately transfer design domain features and educational adaptability. The loss function expression is:

$$L_{total} = \alpha L_{MSE} + \beta L_{style} + \gamma L_{edu}$$

where, L_{MSE} is the pixel-level mean square error loss, used to constrain the basic visual differences between the generated image and the real sample; L_{style} is the design style feature loss, calculated by the cosine distance of the feature maps output by the style feature extraction layer of the teacher and student models, ensuring that the student model inherits the design style expression ability of the teacher model; L_{edu} is the educational adaptability loss, using a cross-entropy loss function to fit the difference between the model output and the manually annotated educational

adaptability score. Hyperparameters are determined by grid search as $\alpha=0.4$, $\beta=0.3$, $\gamma=0.3$, balancing visual quality and educational adaptability training objectives.

The introduction of transfer learning technology aims to reduce training complexity and improve convergence efficiency. Specifically, the strategy of “fixed base modules—training core adaptation modules” is adopted. The model reuses the pre-trained CLIP-ViT-L/14 text encoder and VAE decoder, with the parameters of these modules fixed during training, and only the lightweight U-Net and design style feature distiller are updated. This strategy reduces the number of training parameters from the original Stable Diffusion’s 1.2B to 200M, effectively reducing gradient computation overhead. Meanwhile, the stable feature extraction and image decoding capabilities provided by the pre-trained modules lay the foundation for adapting the training to design education scenarios, resulting in a significant reduction in training cycles.

Notably, Figure 3 presents a small-sample fine-tuning process for the mobile AIGC model for design education. Starting from the existing mobile base model, it integrates the design style point feature space, design specification vertical linear feature space, and educational features from graphic and product design, ultimately producing a mobile lightweight model adapted to design education scenarios. Figure 4 illustrates the multi-domain feature fusion training process for the mobile AIGC model for design education. Through the design education feature Lora adapter, features from graphic design, UI/UX design, and product design fields are integrated, combining design style and specification feature spaces to complete model training, followed by edge-side inference optimization, supporting the model’s efficient operation on mobile devices.

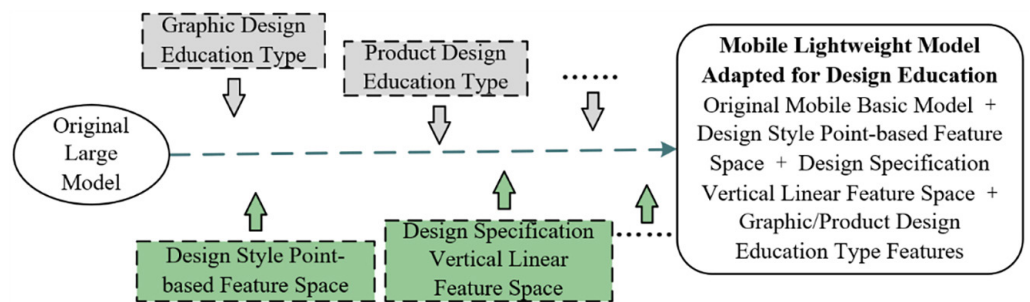


Fig. 3. Schematic diagram of the mobile AIGC model small-sample fine-tuning training process for design education

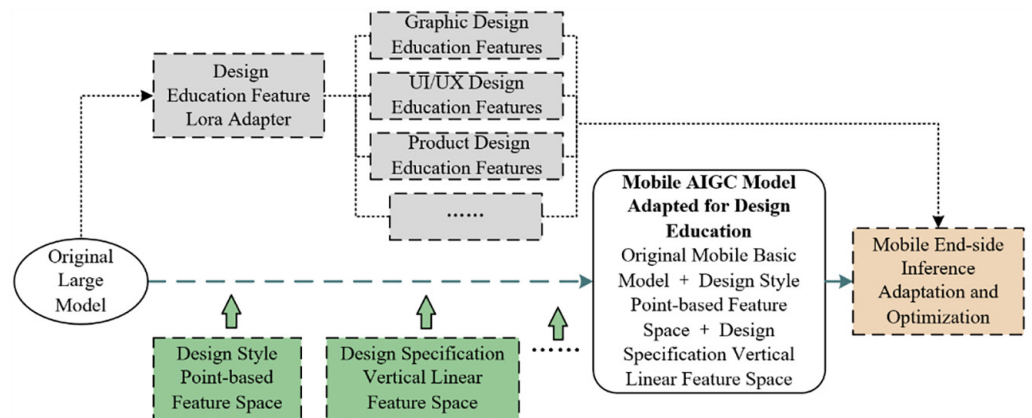


Fig. 4. Schematic diagram of the multi-domain feature fusion training process for the mobile AIGC model for design education

3.3 Mobile optimization implementation

Model compression adopts a “knowledge distillation + quantization optimization” collaborative approach, controlling the model size while selectively retaining core design domain features and educational adaptability. The knowledge distillation phase constructs the knowledge transfer channel between the teacher and student models through a customized design style feature distiller. The distiller selects the intermediate feature map from the U-Net of the teacher model, which is responsible for style feature extraction, as the transfer object. By minimizing the cosine distance between the corresponding feature map of the student model and the teacher model, style knowledge transfer is achieved. During the training process, the style feature difference between the two models is monitored and controlled to ensure it does not exceed 5%. Quantization optimization uses the INT8 symmetric quantization strategy to quantize model weights and activation values. To compensate for the accuracy loss during the quantization process, quantization-aware training is introduced before quantization—injecting quantization noise into the training process so that the model adapts to quantization errors during training, ensuring that the quantized model still retains core capabilities for design specification recognition and creative generation.

Inference speed optimization is approached from both operator reconstruction and hardware adaptation perspectives, improving underlying computational efficiency to meet edge-side real-time interaction demands. At the operator level, the core computational chain of the U-Net module is reconstructed, integrating the traditional “convolution-batch normalization-activation” three-step computation process into a single composite operator, reducing data transmission and memory read/write overhead between operators. Additionally, parallel computation optimization at the operator level improves the processing efficiency of computationally intensive tasks. Hardware adaptation uses a dynamic backend selection mechanism, where the model automatically detects the operating system and hardware acceleration capabilities of the running device: for iOS devices, the CoreML inference backend is called to fully utilize the acceleration capabilities of the Apple Neural Engine; for Android devices, the NNAPI backend is adapted, compatible with neural network processing units from different manufacturers, maximizing the utilization of hardware resources.

Resource consumption control is achieved through dynamic adaptation and memory management technologies to ensure the model runs stably on mobile devices of varying performance levels. The dynamic resolution adjustment mechanism automatically matches the generation resolution according to device hardware performance: by evaluating the terminal’s floating-point computing power through a device computing power detection module, mid-to-high-end devices use a default resolution of 512×512 to ensure generation quality, while low-end devices automatically reduce to 256×256 resolution to control computational power consumption. In terms of memory management, a tensor memory reuse strategy is adopted to plan the reuse of temporary feature tensors generated during the U-Net encoding and decoding process. A memory pool mechanism is used to manage the creation and release of tensors, reducing memory fragmentation and peak occupancy. Additionally, during model inference, unnecessary gradient computation and log recording functions are disabled to further reduce backend resource consumption, ensuring that the model’s resource usage remains within a reasonable range on mainstream mobile devices.

4 EXPERIMENT VERIFICATION AND APPLICATION CASES

To systematically verify the model's generation efficiency, quality, and resource consumption performance on mobile devices of different performance levels, three types of design images—flat posters, product sketch transformations, and UI interfaces—were selected as the generation targets. The core technical indicators of each model were tested on high-end, mid-range, and low-end devices, with results shown in Table 1.

Although the original Stable Diffusion model has the best generation quality, it cannot be deployed on the edge side. In contrast, the proposed model achieves a balance between quality and performance on edge devices. Regarding generation speed, the proposed model improves by 42.3% on high-end devices, 42.3% on mid-range devices, and 41.5% on low-end devices compared to Mobile Stable Diffusion. Even on low-end devices, the 6.8s generation time still meets real-time interaction demands. Regarding generation quality, the proposed model's FID is 26.1% lower and IS is 22.0% higher than Mobile Stable Diffusion, slightly lower than the non-open-source MidJourney Mobile. In terms of resource consumption, the proposed model reduces memory usage by 34.8% on high-end devices compared to Mobile Stable Diffusion and lowers power consumption by 28.0% compared to LiteGAN, with mid-range and low-end devices performing consistently better than the comparison models. These results show that the proposed model's lightweight optimization strategy effectively balances edge-side generation quality, speed, and resource consumption.

To verify the model's adaptability to different design education scenarios, typical scenarios were selected, including undergraduate flat poster design, UI/UX design login page design, and postgraduate product design handheld device sketch transformation. Sixty design students were recruited to complete design tasks, and subjective scoring and target matching rate were used to evaluate educational adaptability. The results are shown in Table 2.

In all scenarios, the educational adaptability indicators of the proposed model are significantly better than those of the open-source comparison models and approach the non-open-source MidJourney Mobile. In terms of subjective scoring, the proposed model's overall creativity support, compliance with norms, and practicality average scores are 8.7, 8.5, and 8.3, respectively, improving by 40.3%, 37.1%, and 38.3% over Mobile Stable Diffusion. In the UI login page design scenario, the compliance with norms reaches 8.8, reflecting the effectiveness of the design specification embedding mechanism. Regarding the target matching rate, the overall rate of the proposed model is 89.7%, and the UI login page design scenario reaches 92%, a 62.5% improvement over Mobile Stable Diffusion, indicating that the multi-modal interaction mechanism can precisely capture educational scene needs. These results demonstrate that the educationally-oriented design of the proposed model effectively matches the core needs of different design education scenarios.

To verify the necessity of the three core modules—design style feature distiller, multi-modal interaction layer, and feedback iteration module—variants of the model were created by removing individual modules. The performance was tested on high-end devices and in the flat poster design scenario. The results are shown in Table 3.

Removing any of the core modules results in a significant decline in model performance, confirming the irreplaceability of each module. After removing the design style feature distiller, the FID increased by 31.0%, and the educational adaptability score decreased by 45.7%, indicating the critical role of this module in preserving design domain features and enhancing educational adaptability. After removing the

multi-modal interaction layer, although generation speed improved by 15.6%, the target matching rate decreased by 40.4%, confirming the core role of multi-modal inputs in accurately capturing design requirements. After removing the feedback iteration module, the educational adaptability score dropped by 32.6%, even though the generation quality changed little, the adaptability to educational scenes was significantly weakened, highlighting the value of iterative optimization for long-term educational adaptability. The results show that the collaborative design of the three core modules is the key to balancing the model's technical performance and educational adaptability.

Table 1. Basic performance experimental results

Models	Generation Time (s)			Generation Quality		Memory Usage (GB)			Power Consumption (mAh/10 images)		
	High-End	Mid-Range	Low-End	FID	IS	High-End	Mid-Range	Low-End	High-End	Mid-Range	Low-End
Proposed Model	3.2	4.5	6.8	18.7	7.2	1.5	1.6	1.8	345	362	388
Original Stable Diffusion	–	–	–	16.2	7.5	–	–	–	–	–	–
Stable Diffusion Mobile	8.9	11.2	15.6	22.9	6.3	3.1	3.3	3.5	498	515	542
Mobile Stable Diffusion	5.5	7.8	11.8	25.3	5.9	2.3	2.5	2.7	479	496	523
LiteGAN	4.5	6.3	9.5	22.1	6.1	2.1	2.2	2.4	479	496	523
MidJourney Mobile	3.8	5.2	7.9	19.3	7.0	1.7	1.9	2.1	368	385	412

Table 2. Educational adaptability experimental results

Models		Proposed Model	Stable Diffusion Mobile	Mobile Stable Diffusion	LiteGAN	MidJourney Mobile
Flat Poster Design	Creativity Support	8.6	5.8	6.1	5.7	8.2
	Compliance with Norms	8.4	5.6	6.0	5.5	7.9
	Practicality	8.2	5.4	5.9	5.3	7.8
	Target Matching Rate (%)	89	52	56	51	85
UI Login Page Design	Creativity Support	8.9	6.1	6.4	5.9	8.5
	Compliance with Norms	8.8	6.3	6.5	6.0	8.2
	Practicality	8.5	5.9	6.2	5.6	8.0
	Target Matching Rate (%)	92	57	57	53	88
Product Sketch Transformation	Creativity Support	8.6	5.9	6.2	5.8	8.3
	Compliance with Norms	8.3	5.7	6.1	5.6	8.0
	Practicality	8.2	5.5	5.8	5.4	7.7
	Target Matching Rate (%)	88	54	55	50	84
Overall Average	Creativity Support	8.7	5.9	6.2	5.8	8.3
	Compliance with Norms	8.5	5.9	6.2	5.7	8.0
	Practicality	8.3	5.6	6.0	5.4	7.8
	Target Matching Rate (%)	89.7	54.3	56.0	51.3	85.7

Table 3. Ablation experiment results

Model Variant	Generation Quality		Educational Adaptability		Generation Speed (s)	Memory Usage (GB)	Target Matching Rate (%)
	FID	IS	Educational Adaptability Score	Creativity Support			
Complete Model (Proposed Model)	18.7	7.2	4.6	8.6	3.2	1.5	89
Remove Design Style Feature Distiller	24.5	6.0	2.5	6.5	3.0	1.4	65
Remove Multi-modal Interaction Layer	20.3	6.8	2.8	6.8	2.7	1.3	53
Remove Feedback Iteration Module	19.1	7.0	3.1	7.2	3.1	1.5	72

A practical application scenario was selected from a university's second-year undergraduate course in Visual Communication Design, specifically the quick prototype design module for posters in the "Basic Graphic Design" course. The module consists of four class hours, and the core teaching objective is to help students master the complete process of "theme determination—sketch drawing—plan optimization—final output." The experimental subjects were 48 students from two natural classes in the course, all of whom had basic hand-drawing and design software skills. The model was used throughout the teaching process as a creative assistance tool. The application process is as follows: students first import hand-drawn thematic sketches, add a design requirement description, and submit them to the model. The model then quickly generates minimalist, retro, and illustration style variants and outputs specification tips based on the graphic design knowledge base, such as suggestions for title font selection and color matching optimization. Students select the target variant and supplement detail adjustment requirements, and the model iterates to generate two optimized plans. Teachers provide targeted guidance based on the solution set generated by the model, focusing on creative logic and visual layout, reducing the time cost of correcting basic specifications. Figure 5 shows the generated images of the minimalist, retro, and illustration style variants. To quantify the application effect, an evaluation system was constructed based on course assignment data, teacher scores, and surveys. The results are shown in Table 4.

The application data in Table 4 shows that the model-assisted teaching mode significantly improved teaching efficiency and quality: students' prototype design time was reduced by 50%, thoroughly solving the core problem of "long conversion time from sketch to digital draft" in traditional teaching; teacher creativity scores and basic specification accuracy improved by 15.7% and 26.6%, respectively, confirming that the model's creative inspiration and specification guidance mechanisms enhance the quality of design output; student creativity satisfaction reached 87.5%, with most feedback indicating that the model's style variant generation function effectively expanded design ideas; teacher units of guidance per hour increased by 133.3%, indicating that the model freed teachers from basic specification corrections, allowing them to focus more on advanced creative guidance, aligning with the core teaching objectives of design education.



Fig. 5. Example images of the model generating minimalist, retro, and illustration style variants

Table 4. Application effect evaluation for undergraduate graphic design course

Evaluation Index	Before Use (Traditional Teaching)	After Use (Model-Assisted Teaching)	Improvement Percentage
Average Poster Prototype Completion Time (Class Hours)	3.0	1.5	50.0%
Teacher Creativity Score (Out of 10)	7.0	8.1	15.7%
Basic Design Specification Accuracy (%)	72.3	91.5	26.6%
Student Creativity Satisfaction (%)	58.3	87.5	49.9%
Teacher Units of Guidance per Hour	1.2	2.8	133.3%

5 CONCLUSION

This paper addresses the core pain points of mobile AIGC tools in design education scenarios, such as inadequate edge-side adaptation, disconnection between interaction and teaching objectives, and the inability to support creative iteration. It builds a “multi-modal educational interaction layer-lightweight AIGC generation model layer-feedback iteration optimization layer” three-layer architecture model. Through lightweight technologies such as design style feature distillation, INT8 quantization, and operator fusion, combined with progressive multi-modal input and design specification embedding as educational-oriented interaction mechanisms, it achieves a deep coupling of technical performance and educational needs. Experimental verification shows that the model performs excellently on high, mid, and low-end mobile devices: the generation time for a single 512×512 image is only 3.2s on high-end devices, with an FID of 18.7, memory usage of 1.5GB, and core technical indicators significantly outperforming comparison models such as mobile stable diffusion and LiteGAN. In terms of educational adaptability, in three scenarios—flat design, UI design, and others—the average subjective scores for creativity support, compliance with norms, and target matching rate all exceed 8.3 points, with a target matching rate of 89.7%. Ablation experiments confirm the key role of the three core modules, such as the design style feature distiller, in educational adaptability. The application

case in the undergraduate course further shows that the model reduced poster prototype completion time by 50%, and teacher units of guidance per hour increased by 133.3%, fully proving its practical value. Theoretically, this paper improves the interdisciplinary theoretical framework of “mobile AIGC-design education” and extends the application of constructivism theory in AI tools; practically, it provides a light-weight and educationally adaptive technical solution for the digital transformation of design education.

6 REFERENCES

- [1] S. Jonnala, B. Swamy, and N. M. Thomas, “Geopolitical bias in sovereign large language models: A comparative mixed-methods study,” *Journal of Research, Innovation and Technologies*, vol. 4, no. 2, pp. 173–192, 2025. [https://doi.org/10.57017/jorit.v4.2\(8\).04](https://doi.org/10.57017/jorit.v4.2(8).04)
- [2] R. El Chaal, R. B. Seghir, and M. O. Aboutafail, “Psychometric evaluation of human-crafted and AI-generated multiple-choice questions for mathematics instruction,” *Education Science and Management*, vol. 3, no. 2, pp. 78–92, 2025. <https://doi.org/10.56578/esm030201>
- [3] J. Chen *et al.*, “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, 2025. <https://doi.org/10.1016/j.foar.2024.08.003>
- [4] M. Y. Chen, W. Wei, H. C. Chao, and Y. F. Li, “Robotic musicianship based on least squares and sequence generative adversarial networks,” *IEEE Sensors Journal*, vol. 22, no. 18, pp. 17646–17654, 2021. <https://doi.org/10.1109/JSEN.2021.3066200>
- [5] L. Meng, “Evaluating the impact of mobile-based generative AI tools on visual design education,” *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 19, no. 16, pp. 189–203, 2025. <https://doi.org/10.3991/ijim.v19i16.57603>
- [6] Z. Yang and J. Shin, “The impact of Gen AI on art and design program education,” *The Design Journal*, vol. 28, no. 2, pp. 310–326, 2025. <https://doi.org/10.1080/14606925.2024.2425084>
- [7] J. Li and Q. Li, “Enhancing educational design capabilities through interactive mobile and adaptive learning platforms: An empirical study,” *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 18, no. 8, pp. 87–101, 2024. <https://doi.org/10.3991/ijim.v18i08.48875>
- [8] V. Sigurdsson, N. M. Larsen, M. Folwarczny, M. Dubois, and A. Fagerstrøm, “Putting an artificial intelligence-generated label on it comes naturally,” *Psychology & Marketing*, vol. 42, no. 2, pp. 579–599, 2025. <https://doi.org/10.1002/mar.22137>
- [9] Y. Cao, A. Abdul Aziz, and W. N. R. Mohd Arshard, “Stable diffusion in architectural design: Closing doors or opening new horizons?” *International Journal of Architectural Computing*, vol. 23, no. 2, pp. 339–357, 2025. <https://doi.org/10.1177/14780771241270257>
- [10] M. Trabilisy *et al.*, “Utilizing generative text-to-image artificial intelligence models to explore race, gender, and age in plastic and aesthetic surgery,” *Aesthetic Surgery Journal*, vol. 45, no. 9, pp. NP172–NP178, 2025. <https://doi.org/10.1093/asj/sjaf084>
- [11] A. Melnik *et al.*, “Face generation and editing with stylegan: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 3557–3576, 2024. <https://doi.org/10.1109/TPAMI.2024.3350004>
- [12] S. F. Wang and C. C. Chen, “Exploring designer trust in artificial intelligence-generated content: TAM/TPB model study,” *Applied Sciences*, vol. 14, no. 16, p. 6902, 2024. <https://doi.org/10.3390/app14166902>

- [13] B. Nethravathi *et al.*, “Visually impaired person assistance based on Tensor FlowLite technology,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 11, pp. 609–614, 2022. <https://doi.org/10.14569/IJACSA.2022.0131171>
- [14] W. Niu *et al.*, “Mobile-3DCNN: An acceleration framework for ultra-real-time execution of large 3D CNNs on mobile devices,” *ACM Transactions on Architecture and Code Optimization*, vol. 22, no. 3, pp. 1–22, 2025. <https://doi.org/10.1145/3747842>

7 AUTHOR

Cheng Wang is an Associate Professor, with Master’s degree, at the General Education Department, Hubei Institute of Fine Arts, Wuhan 430205, China. His research interests include digital media technology and AIGC applications (E-mail: wangcheng@hifa.edu.cn).