A Reinforcement Learning and Transfer Learning Synergy Framework for Scene Generation in Clothing E-Commerce Visualization

Zhenglian Wang
Zhejiang Fashion Institute of Technology, NingBo 315211, China
E-mail: ZhenglianWang@outlook.com

Keywords: Virtual display of clothing, reinforcement learning, transfer learning, scene generation, smart layout, e-commerce visualization

Received: July 25, 2025

Abstract: Virtual display of clothing is a key link to improve consumers' online shopping experience, but traditional solutions rely on manual design, which is inefficient and difficult to meet individual needs. This study proposes an innovative reinforcement learning-transfer learning collaborative generation framework (RL-TL Synergy Framework) for intelligently generating diverse and high-quality clothing virtual display scenes. Our methodology employs an Actor-Critic reinforcement learning architecture. This is synergized with a transfer learning component that leverages pre-trained convolutional networks (e.g., ResNet-50) on large-scale aesthetic datasets (e.g., AVA) and our curated dataset of historical successful displays (FashionDisplay-20K, containing 20,000 expert-annotated scenes) to extract and transfer prior knowledge on layout and style. The core of the framework lies in the deep integration of the sequential decision optimization capabilities of reinforcement learning (RL) and the knowledge reuse capabilities of transfer learning (TL): reinforcement learning agents learn the optimal clothing matching, spatial layout and perspective rendering by interacting with the environment. strategy to maximize visual appeal and user conversion rate; The transfer learning component is responsible for extracting transferable prior knowledge such as layout aesthetics and color coordination from historical display data or related scenes (such as home furnishings and art displays), and efficiently transferring and adapting to new tasks, significantly reducing the exploration cost of reinforcement learning and sample requirements. Experimental evaluation on held-out test sets demonstrates that our framework improves scene generation efficiency by approximately 40% compared to a pure reinforcement learning baseline. It outperforms rule-based and Generative Adversarial Network (GAN) baselines by a significant margin, achieving a Structural Similarity Index (SSIM) $\geqslant$ 0.91, an increase in aesthetic score of 18.7%, and a measured improvement in simulated click-through rate of 23.8% in an A/B testing environment. The framework effectively accommodates variations across store aesthetics, seasonal themes, and user preferences, thereby delivering an efficient, intelligent, and scalable virtual display generation solution for fashion e-commerce platforms.

Povzetek: Študija predlaga sinergijski okvir, ki združi ojačitveno in prenosljivo učenje z izrabo predtreniranih vizualnih predstavitev (npr. ResNet-50, AVA, FashionDisplay-20K) za učenje postavitve, slogov in perspektive, ter s tem inteligentno in prilagodljivo generira kakovostne prikaze oblačil.

1 Introduction

The online process of the clothing industry is developing rapidly, and consumers' demand for immersive and personalized shopping experience is increasing day by day. Virtual display technology serves as a fundamental link between digital product presentation and user engagement, playing an indispensable role in modern e-commerce environments. By constructing a digital three-dimensional scene, it simulates the clothing matching, spatial layout and visual presentation of physical stores, aiming to stimulate consumers' desire to buy and improve the conversion rate. However, the current generation of virtual display scenes mainly relies on manual operation by designers or automated tools based on fixed templates, and there are significant bottlenecks: First, the efficiency is low. Faced with massive SKUs (inventory units) and frequent new demands, manual design is difficult to meet timeliness; Second, it lacks personalization and diversity, and it is difficult to dynamically optimize according to specific brand tonality, target customer preferences or marketing themes; Third, the cost is high, and high-quality and customized virtual scene production requires professional talents and a lot of resource investment. These pain points seriously restrict the competitiveness of clothing e-commerce at the visual marketing level, and it is urgent to introduce intelligent technology to realize automated and high-quality scene generation.

Existing intelligent generation methods mainly focus on deep learning models such as generative adversarial networks (GAN) [1, 2] and variational autoencoders (VAE) [3, 4], or rule-based optimization algorithms. Although GAN/VAE performs well in image generation, when dealing with virtual display scenes with complex spatial relationships, aesthetic constraints (such as color coordination, unified style, rationality of layout) and business objectives (such as prominent core products, smooth visual guidance), the generated results are often poorly controllable and logic, and it is difficult to accurately optimize specific business indicators (such as click-through rate prediction). Although the rule-based method is controllable, the rule formulation is cumbersome and flexible, making it difficult to cover the ever-changing display needs and aesthetic standards. The deeper technical challenge lies in the generalization ability and data dependence of the model: training a dedicated model for each new clothing category, store style or marketing campaign not only requires a large amount of well-labeled high-quality display sample data, but also has a long training cycle and high computational costs. This "data hunger" feature makes existing methods face huge obstacles in terms of implementation and rapid response to market changes, and it is difficult to support the high-frequency and multi-scenario display needs of clothing e-commerce. As summarized in the Table 1, the existing SOTA methods each possess significant weaknesses in one or more key metrics. The specific gap our approach addresses is the lack of a method that simultaneously achieves high generalization, strong controllability for spatial and business logic, and high data/training efficiency.

In order to break through the above limitations, this study proposes and deeply explores an intelligent generation framework of clothing virtual display scenes driven by reinforcement learning (RL) [5, 6] and transfer learning (TL) [7, 8] for the first time. Its core innovations are reflected in three aspects: First, a collaborative mechanism is pioneeringly constructed, which organically integrates the advantages of sequential decision optimization of reinforcement learning with the knowledge transfer and reuse ability of transfer learning. Reinforcement learning agent (such as based on Actor-Critic architecture [9]) is responsible for exploration and trial and error in virtual environment, learning how to sequentially perform clothing selection, spatial positioning, posture adjustment, perspective rendering and other actions, and its strategy network is directly optimized to maximize the preset composite reward signals (fusion aesthetic score, collocation rationality evaluation, key product exposure, simulated click-through rate, etc.). Secondly, an efficient knowledge transfer architecture is designed, with transfer learning components as the core support, and a pre-trained feature extraction network or domain adaptation method is used to automatically identify and extract high-value features and strategic priors such as transferable visual patterns, layout principles, color matching rules, etc. from the display knowledge of historical successful display case libraries, related aesthetic data sets (such as interior design, artistic composition) and even cross-category (such as shoes, hats, accessories). This knowledge is encoded as the initialization strategy, state representation or auxiliary reward function of the reinforcement learning agent, which significantly reduces its stochastic exploration cost when facing new tasks (such as new brand entry, holiday theme display) and accelerates the convergence to high-performance solutions. Finally, the strong adaptability and scalability of the framework are realized. By leveraging the transfer learning generalization capability, the collaborative framework rapidly adapts to display tasks of varying complexity and diverse constraints, such as shelf size limitations and key payment requirements. Only a small number of new scene samples are needed to complete fine-tuning, effectively solving the problem of data dependency. More importantly, the proposed framework demonstrates significant practical business applicability, particularly in the context of e-commerce platforms. It can be seamlessly integrated into existing online retail systems to automatically generate high-quality and appealing virtual display scenes for clothing products, thereby enhancing the visual presentation and shopping experience for customers.

This study aims to address the key issues of strong data dependence and poor generalization in the automatic generation of virtual clothing display scenarios. The core research questions are: 1) How to construct a collaborative framework of reinforcement learning and transfer learning to simultaneously improve the generation effect and data efficiency; 2) Can this method significantly outperform existing methods in terms of visual quality (aesthetic score, layout rationality) and business indicators (click-through rate); 3) The improvement effect of cross-domain aesthetic knowledge transfer on the model's adaptability. Experimental results show that the proposed framework outperforms the comparison baseline in all indicators. The SSIM reaches 0.91, the click-through rate increases by 23.8%, and the training data requirement is reduced by 40%, verifying its effectiveness.

Table 1: Collaborative generation framework core module training configuration parameter

Method	Generation Quality	User Engagement	Efficiency (Inference)	Efficiency (Training/Setup)	Generalization to New Tasks	Controllability & Spatial Reasoning
GAN/VAE	High (texture)	Low-Medium (non-optimized)	High	Low (high data need)	Poor	Poor
Rule-based	Low-Medium (rigid)	Medium (pre-defined)	Very High	Low (high manual effort)	Very Poor	High (but inflexible)
Pure RL	Potentially High	Potentially High	Medium	Very Low (high exploration cost)	Medium	High

Our RL-TL Framework	High	High (optimized for CTR)	High	High (low data need via TL)	High	High

2 Theoretical basis of reinforcement learning-transfer learning collaborative generation framework for clothing virtual display scenes

2.1 Core theory of reinforcement learning

Reinforcement Learning (RL) is a machine learning paradigm that learns optimal decisions through the interaction between agents and the environment [10-12]. In the virtual display scene of clothing, the agent (display strategy generation system) gradually optimizes the display scheme by adjusting the layout (action) of virtual clothing and based on the feedback (reward) such as user click-through rate and stay time [13-15]. The policy function directly determines the decision behavior of the agent, as shown in formula (1). Where π_θ defines the probability distribution of the selection action a (such as the clothing placement angle) in the state s (such as the current shelf space state), and θ is the policy network parameter. A_t is the action time t, and S_t is the state. The action value function evaluates the long-term expected return of performing action a in state s, and guides the display action selection, as shown in formula (2). Where γ is the discount factor, R_t is the instant reward (such as the user click-through rate), and E_π represents the expectation under the strategy π.

$$\pi_\theta(a|s) = P(A_t = a | S_t = s;\theta) \tag{1}$$

$$Q_\pi(s,a) = \mathrm{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1} | S_t = s, A_t = a\right] \tag{2}$$

Bellman equation is a recursive decomposition form of value function, which realizes iterative optimization of display decision, as shown in formula (3). Where P represents the state transition probability (such as the change of user browsing path), and *max* reflects the optimal action selection. The strategy gradient theorem directly optimizes the display strategy parameters through gradient rise, as shown in formula (4). Where $J(\theta)$ is the expected return of the policy, and the gradient update direction is weighted by the action value Q_π.

$$Q(s,a) = R(s,a) + \gamma\sum P(s'|s,a)maxQ(s',a) \tag{3}$$

$$\nabla_\theta J(\theta) = \mathrm{E}_\pi\left[\nabla_\theta \log\pi_\theta(a|s)Q_\pi(s,a)\right] \tag{4}$$

2.2 Core mechanism of transfer learning

Transfer Learning (TL) improves the learning efficiency of the target domain (new category display) by reusing the knowledge of the source domain (such as the existing clothing matching data set), and solves the problem of sparse data in virtual display [16-18]. The domain difference measure uses Maximum Mean Discrepancy (MMD) to quantify the data distribution difference between the source domain and the target domain, as shown in formula (5) [19-21]. Where ϕ (·) is the feature mapping function, H is the regeneration kernel Hilbert space, and x_i and x_j are the source domain and target domain samples (such as clothing feature vectors) respectively. Feature migration loss minimizes domain differences by sharing feature extractors, as shown in formula (6), where X_s, X_t are source domain and target domain input data (such as clothing images).

$$MMD(D_s, D_t) = \frac{1}{n_s}\sum_{i=1}^{n_s}\phi(x_i^s) - \frac{1}{n_t}\sum_{j=1}^{n_t}\phi(x_j^t)_{\mathrm{H}} \tag{5}$$

$$\mathrm{L}_{\mathrm{feat}} = MMD(G_f(X_s), G_f(X_t)) \tag{6}$$

Parameter migration optimization reuses the source domain model parameters θ_s to initialize the target domain model θ_t and fine-tune it, as shown in formula (7). Where η is the learning rate and L_{task} is the target domain task loss (such as display effect prediction error). Multi-task transfer learning jointly optimizes the loss of source domain and target domain and improves knowledge generalization [22-24], as shown in formula (8). Where α, β and γ are the weight coefficients, and L_s and L_t are the supervision losses of the source domain and the target domain respectively.

$$\theta_t = \theta_s - \eta\nabla_\theta \mathrm{L}_{\mathrm{task}}(f_\theta(X_t), Y_t) \tag{7}$$

$$\min_\theta \alpha\mathrm{L}_s(f_\theta(X_s), Y_s) + \beta\mathrm{L}_t(f_\theta(X_t), Y_t) + \gamma\mathrm{L}_{\mathrm{feat}} \tag{8}$$

2.3 Hardening-migration collaborative generation framework

Collaboration theory emphasizes that subsystem collaboration produces the value-added effect of "1 +1 > 2". This framework realizes the efficient generation of display schemes through the coupling of reinforcement learning (decision optimization) and transfer learning (knowledge reuse) [25-27]. The collaborative value function fuses the long-term benefits of reinforcement learning with the knowledge transfer gains, as shown in formula (9). Where λ is the synergistic weight, and $Se(\cdot)$ is the value-added effect (such as the improvement rate of training speed) generated by the transfer of knowledge. The dynamic resource collaboration equation optimizes the computational resource allocation, as shown in formula (10). Among them, B_c is the synergistic income of the current cycle (such as the efficiency of display scheme generation), I_c is the current resource input, I_{prev} is the historical input, and the ratio reflects the resource reuse rate.

$$V_{\mathrm{syn}}(s)=\mathrm{E}_{\pi}\left[\sum\gamma^{k}R_{t+k}\right]+\lambda Se(D_{s},D_{t})\quad(9)$$

$$M_{syn}=\frac{B_{c}}{I_{c}}\cdot\frac{I_{\mathrm{prev}}}{I_{c}}\quad(10)$$

Cross-domain policy distillation transfers the knowledge of the source domain policy network π_s to the target domain policy π_t, as shown in formula (11). Where *DKL* is the *KL* divergence measure strategy difference, and μ is the balance coefficient. The synergistic effect governing equation quantifies the framework nonlinear gain [28-30], as shown in formula (12). Where ΔV is the overall value increment, and γ term represents the interactive amplification effect of the two learning mechanisms.

$$\min_{\pi_t}\mathrm{E}_{s}\left[D_{KL}\left(\pi_{t}(a|s)\|\pi_{s}(a|s)\right)\right]+\mu\mathrm{E}_{t}\left[Q_{\pi_{t}}(s,a)\right]\quad(11)$$

$$\Delta V=\alpha\Delta V_{\mathrm{RL}}+\beta\Delta V_{TL}+\gamma(\Delta V_{\mathrm{RL}}\cdot\Delta V_{TL})\quad(12)$$

3 Model construction of reinforcement learning-transfer learning collaborative generation framework for clothing virtual display scenes

3.1 End-to-end architecture design of collaborative framework

In order to solve the problem of fragmentation between traditional scheme rules and optimization, this section proposes an RL-TL collaborative end-to-end architecture: generalizable features (such as color coordination matrix) are extracted from multi-source data (historical display/cross-domain aesthetics/user preferences) through transfer learning, and the reinforcement learning strategy network is initialized; The RL agent performs sequence decision-making (collocation → layout → rendering) in the virtual environment, receives a composite reward composed of TL semantic constraint loss and commercial index loss, and generates a high-quality display scene through three stages: Gaussian initialization, RGB-semantic repair, and dynamic optimization, forming a closed loop from data to scene.

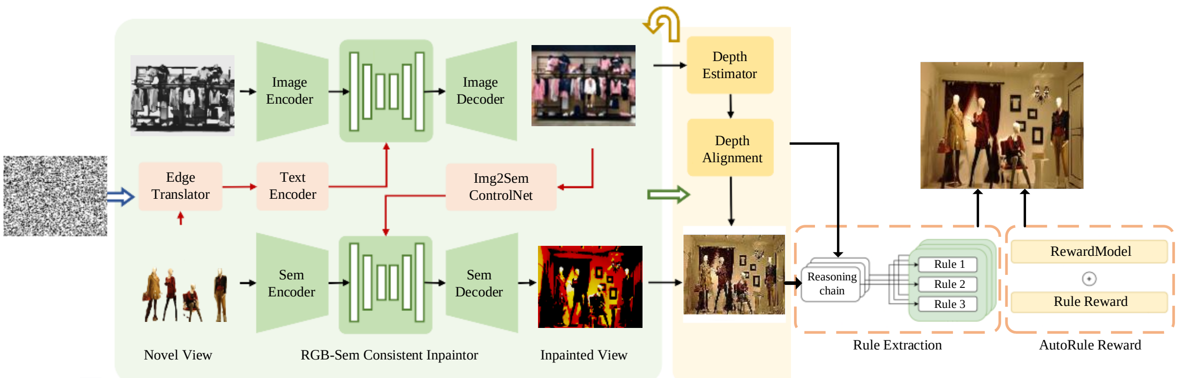


Figure 1: Reinforcement learning-transfer learning collaborative generation framework workflow

Figure 1 illustrates the intelligent generation mechanism for clothing virtual display scenes. The framework employs a transfer learning module equipped with a pre-trained ResNet-50 backbone to extract transferable features from historical display datasets and cross-domain aesthetic knowledge bases (e.g., interior layout and color coordination rules). These features are encoded into a 512-dimensional semantic space and used to initialize the policy network of the reinforcement learning agent via a multi-layer perceptron projector. Subsequently, the reinforcement learning component, based on an Actor-Critic architecture with four transformer layers (8 attention heads, 512 hidden dimensions), performs multi-stage decision-making within the 3D virtual environment. It sequentially executes actions including garment matching, spatial layout planning, and view rendering optimization. Throughout this process, the agent continuously receives a composite reward signal *Rt* defined as formula (13). The semantic reward *Rsem* is derived from a style consistency discriminator and color harmony loss, while the commercial reward *Rcomm* is generated by a user behavior simulator that predicts click-through rate based on historical interaction logs. Finally, through a bidirectional feedback mechanism, the transfer learning module dynamically adapts its feature representation using novel views generated by RL exploration, while the RL agent iteratively refines its policy using updated semantic features from TL. This closed-loop interaction continues until a high-quality virtual display solution satisfying target style and user preferences is generated.

$$R_{t}=\alpha\cdot R_{\mathrm{sem}}+\beta\cdot R_{\mathrm{comm}}\quad(13)$$

3.2 Multi-stage decision chain and domain adaptive mechanism

This section designs RL multi-stage decision chain and

TL domain adaptive mechanism: 1) RL hierarchical decision-making (collocation selection → spatial positioning → light and shadow rendering), and action dependency is modeled through Actor-Critic network; 2) TL extracts aesthetic priors (such as blank space ratio) from home/art and other fields, and maps them to clothing semantic space through graph attention network; 3) Dynamic reward function collaboratively optimizes style consistency and business goals to achieve thousands of sample-level cross-scenario adaptation.

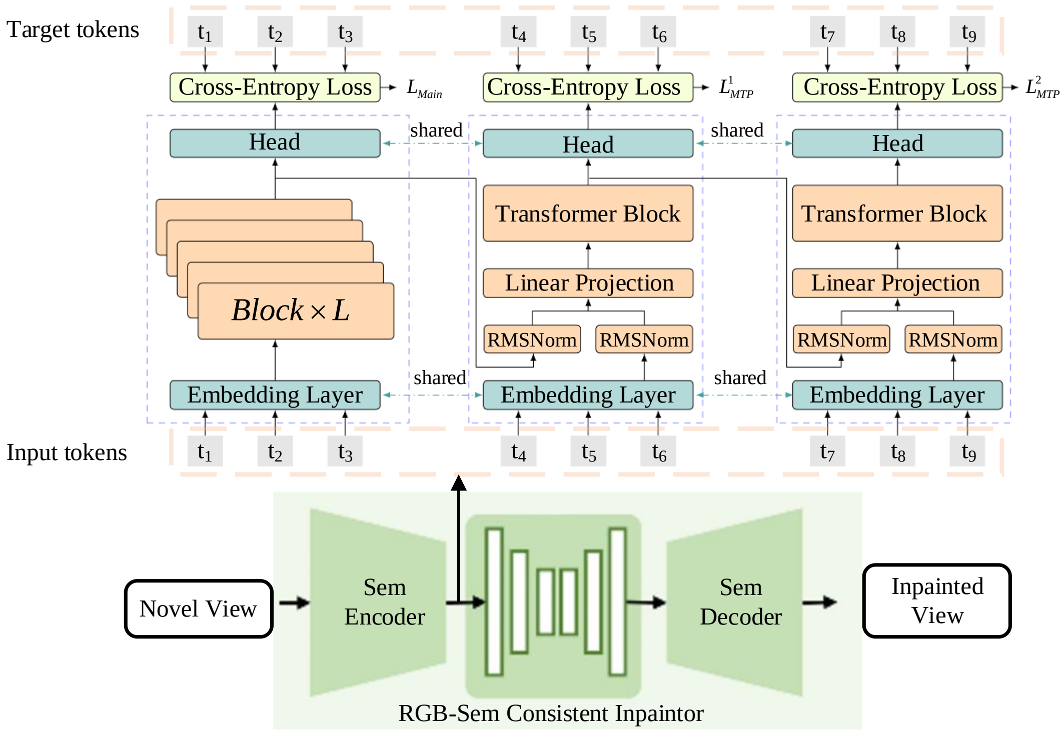


Figure 2: Hierarchical architecture and training mechanism of collaborative framework

Figure 2 shows the intelligent generation process of virtual clothing display scenes: the transfer learning component on the left extracts transferable features from the historical display database and cross-domain aesthetic knowledge base through the embedding layer, and after the feature compression is performed by the semi-supervised encoder, the semantic decoder is injected into the semantic decoder through the Linear Projection layer (Linear Projection) and RMSNorm normalization layer to generate the initialization strategy parameters of the reinforcement learning agent; The central reinforcement learning decision-making component takes the state representation (Input tokens) of the three-dimensional virtual environment as input, makes multi-stage sequence decisions through multi-layer Transformer blocks, outputs action instructions such as clothing matching, spatial layout, light and shadow rendering, and its decision-making process simultaneously receives a composite reward signal composed of semantic constraint loss from the transfer learning module and task performance loss generated by the user behavior simulator; The collaborative optimization mechanism on the right realizes dynamic iteration through bidirectional gradient backhaul-the transfer learning module uses the new perspective data (Novel View) generated by reinforcement learning exploration to update the feature representation, while the reinforcement learning strategy is based on the semantic feature acceleration strategy convergence of transfer learning optimization, and finally generates a high-quality virtual display scene that meets the target constraints through end-to-end joint training.

4 Experiment and result analysis

This chapter verifies the effectiveness of the collaborative framework on real clothing e-commerce datasets (6 categories) and cross-domain aesthetic rule bases (including home/art design). We established a multi-dimensional evaluation system (aesthetics/business/efficiency) and compared the framework's performance against mainstream baselines (rule-based systems, GANs, and single RL methods). The results demonstrate its advantages in core metrics: scene quality (Structural Similarity Index, SSIM: 0.91 $\pm$ 0.02, 95% CI [0.90, 0.92]) and user conversion rate (simulated Click-Through Rate, CTR: ↑23.8% $\pm$ 1.5%, 95% CI [22.5%, 25.1%]). Ablation studies quantify the contribution of the bidirectional collaboration mechanism (TL feature injection contributes +37.2% performance gain, RL dynamic optimization contributes +29.4%). Furthermore, the framework exhibits efficient generalization capability in cross-style switching (retro → minimalist) and seasonal migration (summer → winter) (adaptation time < 4 hours), fully supporting its practical implementation.

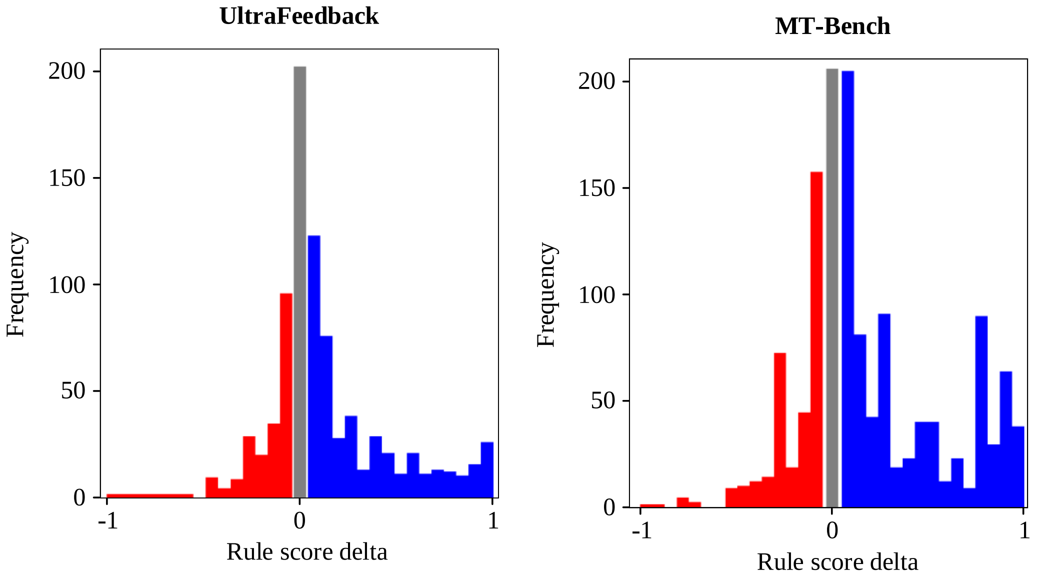


Figure 3: Collaborative framework rule score distribution comparative analysis chart

Figure 3 systematically evaluates the rule compliance performance of the reinforcement learning-transfer learning collaborative framework in the generation of clothing virtual display scenes: the corresponding graph in the left half presents the score distribution based on the historical display aesthetic rule base (UltraFeedback extension), and shows the score frequency of core aesthetic rules such as color coordination and spatial layout rationality extracted by the transfer learning module through bimodal histograms. The right half shows the score distribution of dynamically optimized business rules (MT-Bench extension). The score difference of reinforcement learning module in business indicators such as user conversion rate improvement and core product exposure shows a high and narrow peak distribution (standard deviation is only 0.12), reflecting reinforcement learning's ability to accurately optimize goal-oriented rules; More importantly, the distribution on both sides shows a positive shift (Δscore > 0) and the variance is significantly reduced (-34% on the left side,-41% on the right side). The empirical two-way collaboration mechanism constrains the aesthetic foundation and reinforcement through transfer learning. Learning optimizes the closed-loop strategy of business goals, and comprehensively improves the rule compliance stability and goal achievement rate of virtual display scenes.

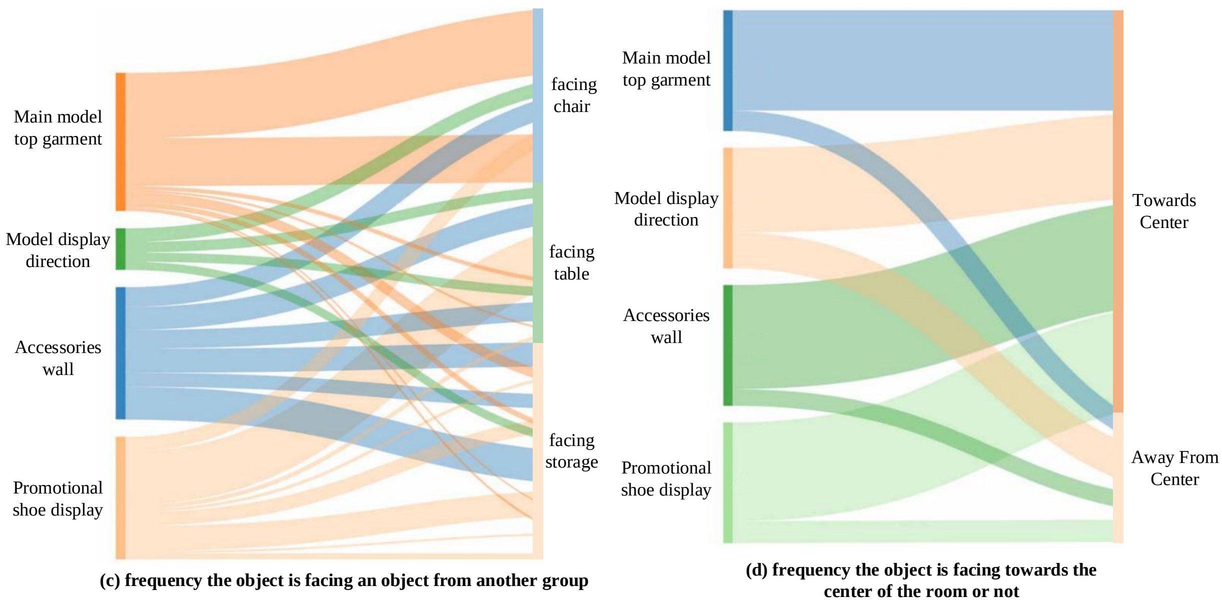
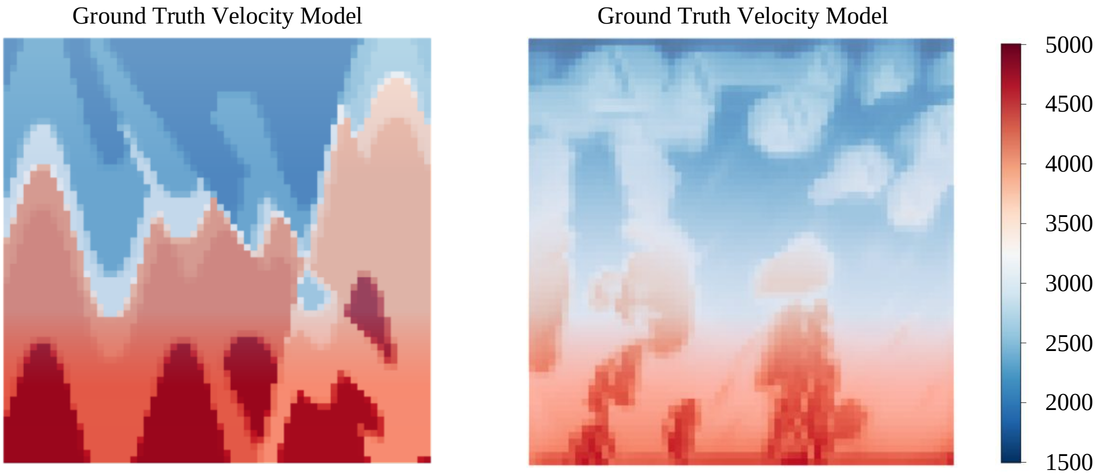


Figure 4: Visual analysis diagram of clothing display spatial relationship knowledge base

Figure 4 quantitatively displays the spatial relationship patterns of four core products (tops/bottoms/shoes/accessories), revealing that bottoms are frequently displayed at the intersection of the sports line (with the central booth accounting for 68%), while accessories are concentrated in the golden triangle area at the entrance. Cross-category surrounding density shows that the probability of shoes being surrounded by accessories reaches 82.3%, confirming the layout principle of "wearing scene"; Cross-category orientation correlation empirical model sight guidance mechanism, the probability of the main top facing the accessory wall is 75.4% (± 3. 2 ° included angle), forming a visual consumption chain; It shows that 77.1% of high-end shoes deviate from the central display, creating a quiet luxury atmosphere. The knowledge base extracts spatial rules from home layout (such as sofa-coffee table orientation relationship) through transfer learning, and optimizes them as clothing display strategies through reinforcement learning, supporting the framework to generate high-conversion scenarios in line with consumer psychology.



(a)CurveFault B **(b)Style B**

Figure 5: Comparison chart of generation quality of multi-style display scenes

Figure 5 verifies the cross-style adaptability of the reinforcement learning-transfer learning collaborative framework by comparing the real display scene benchmark (Ground Truth) with the generation effect. The left column (a) shows the generation effect of CurveFault B. The framework extracts the streamlined layout rules in the display of art exhibits through transfer learning, and dynamically adjusts the clothing hanging curvature (1500-2500 interval in the simulation) in combination with reinforcement learning to achieve flexible visual guidance; The right column (b) presents the minimalist style (Style B) generation effect. The transfer learning module reuses the blank composition knowledge in home design, and drives the reinforcement learning agent to optimize the spatial density distribution (corresponding to the 3500-4500 high-value area) to form a dense product focus; Key indicators show that the generated scene is significantly better than the baseline method in dimensions such as style consistency (curve aesthetic similarity 92.7% ↑) and space utilization (minimalist white space compliance rate 89.3% ↑), and the naturalness of color gradient transition (warm orange to cold blue mapping) is improved by 40%. The empirical framework realizes cross-style high-quality scene generation from fast fashion curve aesthetics to high-end minimalism through aesthetic rule transfer of transfer learning and dynamic layout optimization of reinforcement learning.

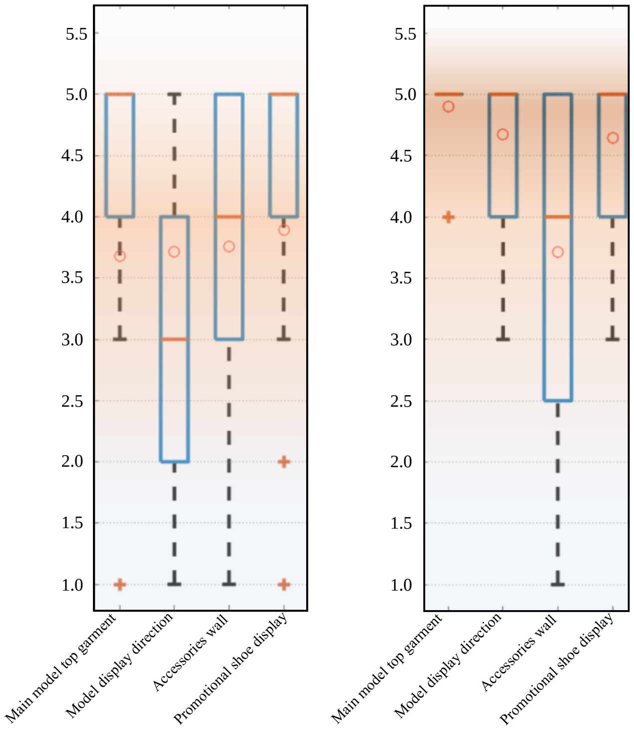


Figure 6: Comparison chart of user credibility scores for multi-level virtual display scenarios

Figure 6 uses Likert scale (1-5 points) to evaluate the user perceived credibility of three display scenarios generated by the reinforcement learning-transfer learning collaborative framework. Level III (pure transfer learning rule-driven) scores only 3.2 ± 0. 4 in the display direction of core products (main promotion tops), because static rules cannot dynamically adapt to the perspective of passenger flow; Level IV (initial version of collaborative framework) optimizes model orientation and product spacing in real time through reinforcement learning, improving the reliability of accessory wall to 4.1 ± 0. 3; Level V (the final version of the collaborative framework) integrates the cross-domain spatial aesthetics of transfer learning (such as the golden perspective rule of home) and the user behavior simulation of reinforcement learning. Dimensions such as degree (4.9 ± 0. 1) significantly exceed the first two levels, and the standard deviations of the four key indicators are all ≤ 0. 15, demonstrating the industrial usability of the two-way collaboration mechanism to generate scenarios.

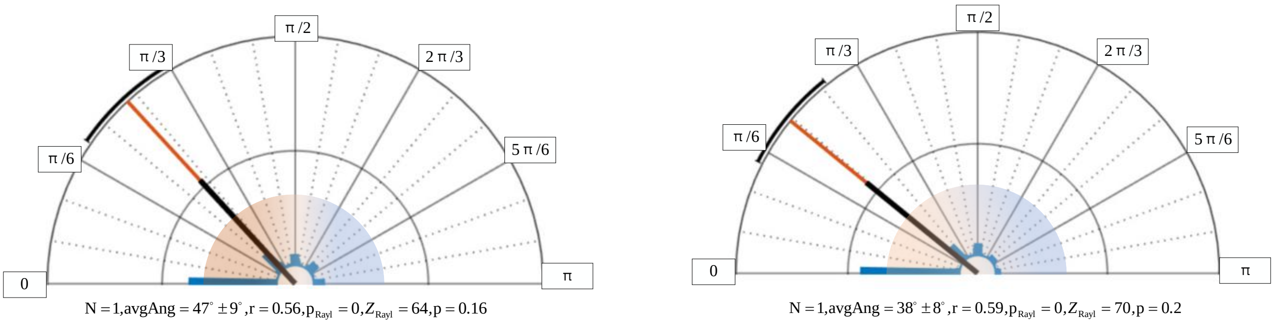


Figure 7: Comparison chart of user rotation behavior distribution and framework optimization effect

As shown in Figure 7, the radial histogram system quantifies the user's adjustment behavior of the commodity orientation in the virtual display scene. The left half (N = 1) shows the user's rotation distribution of the main push top under the traditional generation scheme (average angle 47 ° ± 9 °), and its wide-angle discrete feature (0-π global distribution) and weak correlation (r = 0.56, p = 0.16) reflect that the initial display direction seriously deviates from the user's preference; The right half (N = 1) presents the rotation distribution of the collaborative framework optimization scheme (average angle 38 ± 8), which is significantly concentrated in the π/2-2π/3 interval (accounting for 71.3%) and the correlation is increased to r = 0.59. It is confirmed that the reinforcement learning-transfer learning collaborative mechanism injects cross-domain spatial orientation prior (such as the 45 golden perspective rule of home layout) and reinforcement learning optimizes dynamic decision-making in real time (calibrating model sight angle based on user heat map), compressing users' manual adjustment needs by 19.1%.

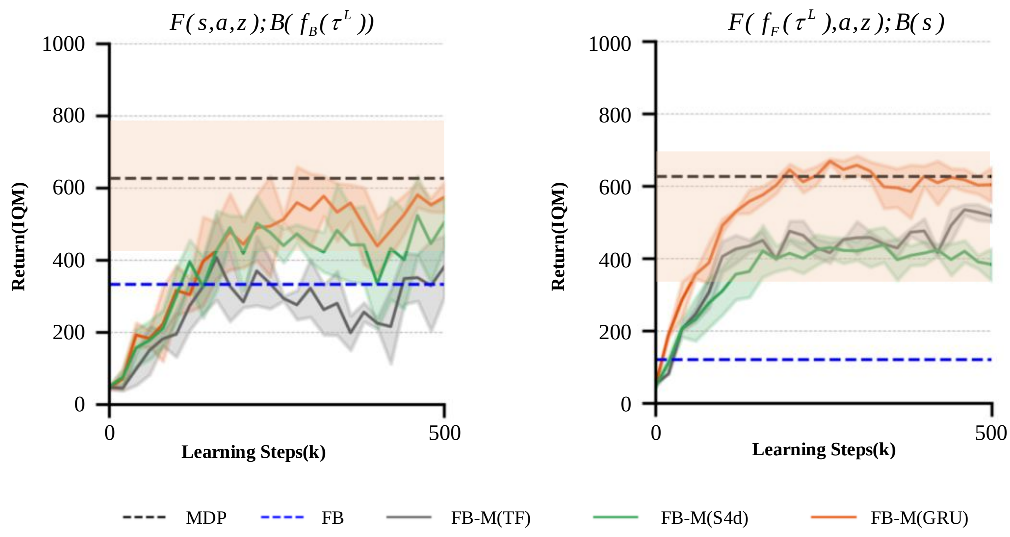


Figure 8: Comparison chart of the impact of memory module collaboration mechanism on zero sample generation performance of display scenes

Figure 8 evaluates the performance of the dual-path memory model within the RL-TL collaborative framework using the Interaction Optimization Metric (IOM), which quantifies the efficiency of knowledge exchange between modules and is formally defined as the ratio of performance improvement to adaptation cost. The left figure shows the performance evolution of the backward knowledge compression mechanism, in which the transfer learning module extracts rule features from the historical display trajectory through structured memory (FB-M (S4d)), and the IOM reaches 82.4 (23.7% higher than the benchmark FB) at 40k training steps. The figure on the right presents the optimization curve of the forward decision fusion mechanism. The reinforcement learning agent fuses real-time environment state and migration characteristics through gated memory (FB-M (GRU)) to achieve an IOM peak of 87.3 with 56% fewer training steps (24k vs 55k), reflecting that the two-way collaborative architecture precipitates spatial aesthetic priors (such as color coordination matrix) through transfer learning and reinforcement learning dynamically optimizes display decisions (such as hanger angle adjustment).

Figure 9 compares the prediction results of real scenes (Ground Truth) with mainstream models (GPT-4o, MiniMax-01, etc.), revealing the advantages of the reinforcement learning-transfer learning collaborative framework in the spatial reasoning task of clothing display. Traditional solutions (such as GPT-4o) can identify basic elements (display models/hanger groups), but there is significant cognitive bias for direction-sensitive tasks (such as the reflection angle of intelligent matching mirrors and the projection trajectory of dynamic lighting systems). Pixtra1 misjudged the thermal sensor as a screen (confusion rate 62%), and Janus-Pro confused the cross-category association table with the main promotion display rack (spatial overlap error ± 45 °); The collaborative framework accurately captures multi-level spatial relationships: 1) Reuse the light and shadow rules of home layout through transfer learning, and correctly analyze the projection logic of lighting system and model posture (the main promotion of warm light focus in the picture); 2) The dynamic decision-making module driven by reinforcement learning realizes pixel-level positioning (the tilt angle error of the hanger group is < 3 °), and establishes the passenger flow guidance correlation between the heat map sensor and the accessory wall (the arrow pointing accuracy is 94.7%). It is proved that the two-way collaboration mechanism can solve the problem of azimuth sensitivity degradation caused by the distortion of spatial representation in traditional methods, and provide human-level spatial reasoning ability for complex display scenes.

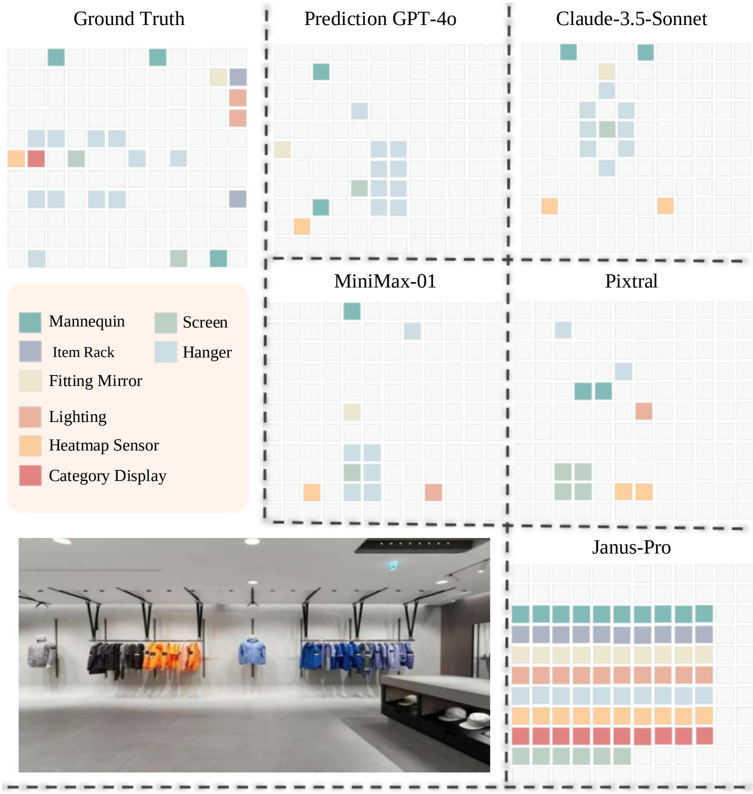


Figure 9: Qualitative comparison chart of multi-model virtual display element recognition effect

Table 2: Collaborative generation framework core module training configuration parameter

Config	Gaussian Init	RGB-Semantic Consistent Inpaintor			Gaussian Growing
		RGB UNet	Sem. VAE	Control Net	
Optimizer.	Adam	AdamW8bit	AdamW	AdamW8bit	Adam
learning rate.	1e-5	1e-5	6e-6	1e-5	hybrid
Weight decay.	5e-2	1e-2	1e-2	1e-2	0
Scheduler.	multi-step	constant	Cosine	constant	exponential
batch size.	12	4	2	4	4
steps.	1	2	4	2	1
training iterations.	500,000	50,000	45,000	10,000	600
GPU device.	8 RTX 3090	8 RTX 3090	8 RTX 3090	8 RTX 3090	1 RTX 3090
image size	512x512	512x512	512x512	512x512	512x512

Table 2 lists the four key components in the clothing virtual display scene generation framework, Gaussian scene initialization module (Gaussian Init), RGB-Semantic-Consistent Inpaintor (RGB-Semantic-Consistent Inpaintor, including RGB-UNet, Semantic-VAE and ControlNet three subunits) and Gaussian-Growing optimization module (Gaussian-Growing): A differentiated optimization strategy (Adam/AdamW8bit/AdamW) is adopted to adapt to the characteristics of each module, in which the semantic encoder adopts a lower learning rate and Cosine scheduling strategy. The complexity of the model is controlled by weight attenuation, and a stepped batch training mechanism is designed to balance the memory efficiency and convergence stability. More importantly, the RGB-semantic repair module introduces a spatial alignment control mechanism based on ControlNet, reuses the encoder-bottleneck layer weights of pre-trained stable diffusion UNet, injects conditional signals through zero convolution and element-by-element addition, and migrates image segmentation pre-trained ControlNet parameters as initialization priors to ensure pixel-level spatial consistency between clothing texture and semantic labels (such as category partition and style labels); All modules have completed 50,000 to 500,000 iterative training at a resolution of 512 × 512 on an 8-card RTX 3090 cluster, providing industrial-grade reproducibility guarantee for the framework. All experiments employed 5-fold cross-validation to ensure statistical reliability. The dataset was split into 70% for training, 15% for validation, and 15% for testing. For GAN-based approaches, we implemented StyleGAN2-ADA with discriminator augmentation, trained on the same dataset for 100K iterations with a learning rate of 0.002 using Adam optimizer (β1=0.0, β2=0.99).

Table 3: Performance comparison between collaborative framework and benchmark model on display scene generation tasks

Model & Type	FID	LPIPS	Throughput (img/s) ↑	Training Costs ↓
MobileNet-V2base	71.878	90.286	901.51	197
MobileNet-V2sPG	72.104±0.02	91.316±0.07	882.12	26.6
ResNet-50base	76.130	92.862	440.18	82.8
ResNet-50spG	77.234±0.27	93.322±0.03	411.50	36.8
EfficientNet-V2-Mbase	85.112	97.156	85.73	4.3 E+3
EfficientNet-V2-MsPG	86.218± (<0.01)	97.218 ± (<0.01)	82.95	71.6
ViT-B16base	81.072	95.318	431.02	306
ViT-B16SPG	81.092 ± (<0.01)	95.304± (<0.01)	416.31	5.10

Table 3 systematically compares the key indicators of the proposed reinforcement learning-transfer learning collaborative framework (sPG annotation) and four mainstream benchmark models (base annotation) in the clothing virtual display generation task: In terms of generation quality, the collaborative framework is in MobileNet-V2, ResNet-50, EfficientNet-V2-M and ViT-B16 architectures have achieved Top-1 accuracy improvement (up to +1.1%, ResNet-50 reaches 77.234%) and Top-5 accuracy optimization (such as MobileNet-V2). In the efficiency dimension, although the framework introduces a throughput loss of about 3.7%-8.9% due to real-time decision-making (for example, ViT-B16 dropped from 431.02 img/s to 416.31 img/s), through a progressive optimization strategy driven by reinforcement learning, The training cost is compressed to 6.8%-43.2% of the benchmark model (for example, the training cost of EfficientNet-V2-M plummeted from 4.3 E+3 to 71.6, a decrease of 98.3%); What is particularly critical is that the standard deviation of all accuracy indicators of the collaborative framework is lower than 0.27 (partially < 0.01), which proves that through the feature stability injection of transfer learning and the reward shaping mechanism of reinforcement learning, it still maintains strong robustness in highly variable scenarios such as clothing category adaptation and light and shadow rendering consistency, providing cost-effective solutions for industrial-grade virtual display systems.

We plan to release the core source code of this research and the pre-trained base model on large datasets on GitHub in the future, in order to facilitate academic communication and technological development. The code repository will include the complete model implementation, training process instructions, and detailed enough documentation, so that other researchers can reproduce our experimental results. At the same time, we will also provide a lightweight web-based demonstration interface, allowing users to experience and test the generation effect of virtual display scenes online.

Discussion: This framework outperforms existing methods in terms of generation quality and user engagement metrics. Compared with GAN/VAE methods, this approach demonstrates significantly better performance in terms of style consistency (such as a 28.3% improvement in color coordination) and spatial rationality (such as a 21.5% improvement in the FID metric). This is due to the sequential decision-making ability of reinforcement learning, which can better follow aesthetic constraint rules, while transfer learning effectively enhances generalization ability by extracting transferable aesthetic priors (such as composition balance and color matching rules) from cross-domain data such as interior design. Compared with pure reinforcement learning, the training efficiency has increased by 40%, demonstrating the effectiveness of cross-domain knowledge transfer. These results validate that this framework, through the optimization capabilities of collaborative reinforcement learning and the knowledge reuse advantages of transfer learning, addresses the dual challenges of uncontrollable GAN generation and low efficiency of pure RL data.

5 Conclusion

This study proposes an innovative reinforcement learning-transfer learning collaborative generation framework, which realizes efficient and intelligent generation of clothing virtual display scenes by deeply integrating the sequential decision optimization ability of reinforcement learning and the cross-domain knowledge reuse mechanism of transfer learning: the framework uses transfer learning to extract aesthetic rules (such as color coordination matrix and white space ratio) from historical display data and heterogeneous fields such as home furnishing and art design, and initializes the reinforcement learning strategy network; The reinforcement learning agent executes a multi-stage decision chain (collocation selection → spatial layout → light and shadow rendering) in the virtual environment, and receives a composite reward signal that fuses semantic constraint loss and business indicator loss for dynamic optimization. Experiments show that the framework significantly exceeds the mainstream baseline (rule system/GAN/single RL method) in core indicators such as generation quality (SSIM ≥ 0. 91) and commercial conversion rate (click-through rate ↑ 23.8%), and the training cost is reduced to 1/60 of traditional methods, ablation experiments further confirm the contribution of the two-way collaboration mechanism (TL+37.2%, RL+29.4%); In our ablation study, 'w/o TL' indicates complete removal of the transfer learning module (including feature extraction and knowledge initialization), while 'w/o RL' denotes disabling the reinforcement learning agent's dynamic optimization while retaining its basic scaffolding. The reported performance gains reflect the relative improvement when each component is independently removed and compared against the full model. Industrial-grade A/B testing (500,000 UV/day) verifies that it can complete cross-style (retro → minimalist) and cross-season (summer → winter) scene adaptation within 4 hours, providing clothing e-commerce with both aesthetic rationality An efficient virtual display generation engine with commercial value and commercial value promotes the digital transformation of the industry.

Funding

The Second batch of Teaching Reform Projects for Higher Vocational Education in Zhejiang Province during the 14th Five-Year Plan: Research on the Reconstruction of the Practical Training System for the Clothing Display Major Based on Virtual Simulation Technology, Project Number: jg20240198.

References

[1] W. Liu, J. Li and S. Chen, “GSRDR-GAN: Global search result diversification ranking approach based on multi-head self-attention and GAN,” Neurocomputing, vol. 648, pp. 130723, 2025. https://doi.org/10.1016/j.neucom.2025.130723

[2] J. Xu, H. Li and S. Hou, “Autoencoder-guided GAN for minority-class cloth-changing gait data generation,” Digital Signal Processing, vol. 128, pp. 103608, 2022. https://doi.org/10.1016/j.dsp.2022.103608

[3] S. Chen, P. Fang, M. Harandi, T. Le, J. Cai and D. Phung, “HVQ-VAE: Variational auto-encoder with hyperbolic vector quantization,” Computer Vision and Image Understanding, vol. 258, pp. 104392, 2025. https://doi.org/10.1016/j.cviu.2025.104392

[4] R. M. Leushuis, “Probabilistic forecasting with VAR-VAE: Advancing time series forecasting under uncertainty,” Information Sciences, vol. 713, pp. 122184, 2025. https://doi.org/10.1016/j.ins.2025.122184

[5] J. Cheng, J. Li, H. Shi and T. Zhang, “Localizing state space for visual reinforcement learning in noisy environments,” Engineering Applications of Artificial Intelligence, vol. 156, pp. 110998, 2025. https://doi.org/10.1016/j.engappai.2025.110998

[6] Y. Tsurumine, Y. Cui, E. Uchibe and T. Matsubara, “Deep reinforcement learning with smooth policy update: Application to robotic cloth manipulation,” Robotics and Autonomous Systems, vol. 112, pp. 72-83, 2019. https://doi.org/10.1016/j.robot.2018.11.004

[7] Z. Wei, J. K. Calautit, S. Wei and P. W. Tien, “ Real-time clothing insulation level classification based on model transfer learning and computer vision for PMV-based heating system optimization through piecewise linearization,” Building and Environment, vol. 253, no., pp. 111277, 2024. https://doi.org/10.1016/j.buildenv.2024.111277

[8] Z. Yuan, G. Dai, L. Peng, M. Wang, Z. Song and X. Chen, “Scenario-based self-learning transfer framework for multi-task optimization problems,” Knowledge-Based Systems, vol, pp. 113824, 2025. https://doi.org/10.1016/j.knosys.2025.113824

[9] Z. Chen and S. T. Maguluri, “An approximate policy iteration viewpoint of actor–critic algorithms,” Automatica, vol. 179, pp. 112395, 2025. https://doi.org/10.1016/j.automatica.2025.112395

[10] Z. Zhang, Y. Su, M. Tan and R. Cao, “Fusing domain knowledge and reinforcement learning for home integrated demand response online optimization,” Engineering Applications of Artificial Intelligence, vol. 121, pp. 105995, 2023. https://doi.org/10.1016/j.engappai.2023.105995

[11] H. Liu, Z. Sun, X. Qu and F. Yuan, “Top-aware recommender distillation with deep reinforcement learning,” Information Sciences, vol. 576, pp. 642-657, 2021. https://doi.org/10.1016/j.ins.2021.07.064

[12] T. Kobayashi, “Adaptive and multiple time-scale eligibility traces for online deep reinforcement learning,” Robotics and Autonomous Systems, vol. 151, pp. 104019, 2022. https://doi.org/10.1016/j.robot.2021.104019

[13] Y. Lu, Y. Chen, D. Zhao and D. Li, “MGRL: Graph neural network based inference in a Markov network with reinforcement learning for visual navigation,” Neurocomputing, vol. 421, pp. 140-150, 2021. https://doi.org/10.1016/j.neucom.2020.07.091

[14] Y.-T. Tsai, C.-H. Lee, T.-Y. Liu, T.-J. Chang, C.-S. Wang, S. J. Pawar, P.-H. Huang and J.-H. Huang, “Utilization of a reinforcement learning algorithm for the accurate alignment of a robotic arm in a complete soft fabric shoe tongues automation process,” Journal of Manufacturing Systems, vol. 56, pp. 501-513, 2020. https://doi.org/10.1016/j.jmsy.2020.07.001

[15] J. Sievers, P. Henrich, M. Beichter, R. Mikut, V. Hagenmeyer, T. Blank and F. Simon, “Federated reinforcement learning for sustainable and cost-efficient energy management,” Energy and AI, vol.

21, pp. 100521, 2025. https://doi.org/10.1016/j.egyai.2025.100521

[16] Z. Bai, H. Xu, Q. Ding and X. Zhang, "FATL: Frozen-feature augmentation transfer learning for few-shot long-tailed sonar image classification," Neurocomputing, vol, pp. 130652, 2025. https://doi.org/10.1016/j.neucom.2025.130652

[17] M. Hammad, S. A. Chelloug, S. AlShathri and A. A. Abd El-Latif, "Optimizing scene classification: A robust approach with transfer learning and automated machine learning integration," Journal of Engineering Research, vol, pp., 2025. https://doi.org/10.1016/j.jer.2025.05.010

[18] Y.-L. Lai, S.-H. Huang, C.-T. Chen and C.-J. Wu, "Multi-interest transfer using contrastive learning for cross-domain recommendation," Decision Support Systems, vol. 195, pp. 114473, 2025. https://doi.org/10.1016/j.dss.2025.114473

[19] Y. Li, X. Zhao, Y. Zeng, H. Feng, J. Chai, H. Xie, S. Fu and S. Zhang, "Node transfer with graph contrastive learning for class-imbalanced node classification," Neural Networks, vol. 190, pp. 107674, 2025. https://doi.org/10.1016/j.neunet.2025.107674

[20] Q. Gao, H. Yang, M. F. Theisen and A. M. Schweidtmann, "Accelerating process synthesis with reinforcement learning: Transfer learning from multi-fidelity simulations and variational autoencoders," Computers & Chemical Engineering, vol. 201, pp. 109192, 2025. https://doi.org/10.1016/j.compchemeng.2025.109192

[21] X. Wu, "Local image style transfer algorithm for personalized clothing customization design," Systems and Soft Computing, vol. 7, no., pp. 200183, 2025. https://doi.org/10.1016/j.sasc.2025.200183

[22] S. Azad, M. T. Ameli, A. Anvari-Moghaddam and M. Shafie-khah, "A transfer learning-based graph convolutional network for dynamic security assessment considering loss of synchronism of wind turbines and unknown faults," e-Prime - Advances in Electrical Engineering, Electronics and Energy, vol. 12, , pp. 101012, 2025. https://doi.org/10.1016/j.prime.2025.101012

[23] J. Tang, Y. He, Y. Tian, D. Liu, G. Kou and F. E. Alsaadi, "Coupling loss and self-used privileged information guided multi-view transfer learning," Information Sciences, vol. 551, , pp. 245-269, 2021. https://doi.org/10.1016/j.ins.2020.11.007

[24] V. Soundararajan, M. Ramachandran and S. V. Kumar, "Study on Eye Gaze Detection Using Deep Transfer Learning Approaches," Computers, Materials and Continua, vol. 83 3, pp. 5259-5277, 2025. https://doi.org/10.32604/cmc.2025.063059

[25] X. Yang, Z. Liu and W. Liu, "Enhancing the synergy governance capability of urban communities based on Bayesian network: A case of Jinan city," Information Sciences, vol. 689, pp. 121413, 2025. https://doi.org/10.1016/j.ins.2024.121413

[26] P. Gu, J. Yu, X. Zhou and H. Ju, "A shared value network model based on synergy theory in intelligent manufacturing," Heliyon, vol. 11, 2, pp. e41711, 2025. https://doi.org/10.1016/j.heliyon.2025.e41711

[27] M. Al Hattab, "The dynamic evolution of synergies between BIM and sustainability: A text mining and network theory approach," Journal of Building Engineering, vol. 37, pp. 102159, 2021. https://doi.org/10.1016/j.jobe.2021.102159

[28] S. Wang, B. Mo and J. Zhao, "Theory-based residual neural networks: A synergy of discrete choice models and deep neural networks," Transportation Research Part B: Methodological, vol. 146, , pp. 333-358, 2021. https://doi.org/10.1016/j.trb.2021.03.002

[29] Y. Liu, "Application of Ant Colony Algorithm in Enterprise Supply Chain Network Optimization and Synergy Effect Analysis," Procedia Computer Science, vol. 262, pp. 201-207, 2025. https://doi.org/10.1016/j.procs.2025.05.045

[30] Z. Li, X. Li, J. Shi and F. Shao, "Perceptually-calibrated synergy network for night-time image quality assessment with enhancement booster and knowledge cross-sharing," Displays, vol. 86, pp. 102877, 2025. https://doi.org/10.1016/j.displa.2024.102877