

Automated Shadow Decomposition for Animation Production Workflows Based on Hybrid Smooth and Shadow Networks

Bo Jiao

Lecturer, School of Art, Zhengzhou Business University, Gongyi, 451200, China, E-mail: jiaobo2025@126.com

Production Management

Received January 21, 2026; revised March 18, 2026; August 4, 2026; accepted September 7, 2026

Available online September 20, 2026

Abstract: In modern digital animation production, with continuous technological advancements, especially innovations in computer graphics and artificial intelligence, the efficiency and quality of the animation production process have become key factors in the competitiveness of production companies. However, despite the rapid development of rendering technologies and image generation algorithms, shadow processing remains a complex and time-consuming challenge. To address operational bottlenecks and high manual labor costs in digital animation production pipelines, this study presents an automated shadow-decomposition framework to optimize post-production workflows. To overcome the challenge of separating shadows and reflectance in dynamic animation, this study develops an animation shadow decomposition model that employs a hybrid smoothing network and a shadow network. The hybrid smoothing module improves the distinction between shadows and base color layers by smoothing multi-scale features, and the shadow network introduces a temporal consistency loss to further optimize shadow coherence between frames. Experimental outcomes reveal that in the reflectance separation task, the model attains a peak signal-to-noise ratio of 34.27 dB, a structural similarity index of 0.956, a temporal consistency of 0.943, and a temporal error of 0.048. For rendering pipeline developers and technical specialists, practical validation confirms that the model directly optimizes the animation production workflow. By enhancing temporal coherence, it systematically replaces manual rotoscoping and frame-correction efforts with a standardized, automated process. This shift allows practitioners to reallocate human resources from repetitive corrective tasks to high-value creative stages, fundamentally improving the reliability, cost-efficiency, and throughput of post-production engineering. By automating the shadow-reflectance separation, the model significantly reduces the reliance on frame-by-frame manual masking, offering a cost-effective solution that enhances rendering efficiency by shortening the iterative feedback loop in the animation editing workflow. This research aligns with emerging digital production standards for automated workflows. By providing a robust AI-assisted framework, it contributes to the governance of automated tools in creative industries, addressing the impact of automation on workforce efficiency and digital asset management policy. Future research will focus on the model's scalability within cloud-based production architectures and its long-term impact on digital asset reuse.

Keywords: Animated shadows; hybrid smooth network; temporal consistency; production workflow optimization; digital rendering systems.

Copyright © Journal of Engineering, Project, and Production Management (EPPM-Journal).
DOI 10.32738/JEPPM-2026-79

1. Introduction

In digital production systems, shadow decomposition is not merely a technical challenge but a critical factor affecting engineering efficiency. Traditional pipelines often require labor-intensive manual correction, which accounts for a substantial portion of project schedules and human-resource costs. (Yu et al., 2022; Wang et al., 2022). However, in complex dynamic scenes, accurately separating the shadow layer from the reflectance layer remains a significant challenge (Zhang 2022; Duan et al., 2024). From a production management perspective, technical bottlenecks in current shadow decomposition processes create systemic operational inefficiencies. These limitations necessitate repetitive manual interventions that not only escalate labor costs but also disrupt the synchronized workflow of digital asset production, ultimately compromising the overall throughput and predictability of the production pipeline (Jiang et al., 2023; Imahayashi et al., 2023).

In 3D cartoon face synthesis and animation editing, Guo et al. (2023) proposed an integrated generation and style-editing method that constructs a unified framework spanning from geometric modeling to appearance control. This work achieves consistent style and stable morphology during animation by decoupling structure and appearance, supporting subsequent

editing and rendering. Addressing the limited precision of instance-level editing in 2D anime images, Lin et al. (2025) proposed an instance-guided anime-editing method and constructed a large-scale, high-quality dataset for editing. This method achieves precise editing of specific regions through instance-level guidance, maintains overall style and temporal consistency, and demonstrates good robustness in complex scenes. In shadow removal, Kang et al., addressing the lack of paired data, proposed a Cycle-in-Cycle Generative Adversarial Network for Shadow Removal using Unpaired Data (C2ShadowGAN), enhancing the robustness of the removal effect by introducing cycle-consistency loss (Kang et al., 2023). Liu et al. enhanced the ability to restore shadow details in complex backgrounds by combining an image bilinear model with a three-branch residual network (Liu et al., 2023). Guo et al. (2025) employed pseudo-illumination region modeling and logarithmic-domain transformation to reduce lighting interference and improve image quality. To address reduced performance in unsupervised shadow removal when labeled data is insufficient, Wang et al. proposed an enhancement method based on multi-intensity shadow generation and diffusion modeling, improving removal accuracy and robustness (Wang et al., 2025). Wang et al. (2024) addressed dynamic lighting changes by introducing illumination-aware convolution and dynamic alignment strategies to improve shadow recovery. Furthermore, Aghamaleki (2022) improved the detection accuracy of large-scale shadow regions in aerial images by optimizing a 2D ratio-mapping method.

This literature review summarizes various innovative shadow removal methods that have improved accuracy and adaptability. However, existing methods do not comprehensively address shadow consistency and temporal stability in dynamic animated scenes. To address these limitations, this study introduces an animation shadow decomposition model that combines a hybrid smooth network and a shadow network. The contribution of this work lies in establishing a structured decomposition framework that reformulates animation shading as an automated production task and provides a theoretical basis for intelligent rendering management.

2. Methods and Materials

2.1. Construction of Animation Shadow Decomposition Model and Design of Hybrid Smooth Network

To accurately separate the shadow and reflectivity layers in dynamic animation frames, this study proposes a shadow decomposition model based on multi-path feature fusion. This model employs a single encoder and multiple decoder structure, combined with a hybrid smoothing module, to improve the accuracy of reflectivity and shadow map reconstruction. The overall method structure is shown in Fig. 1.

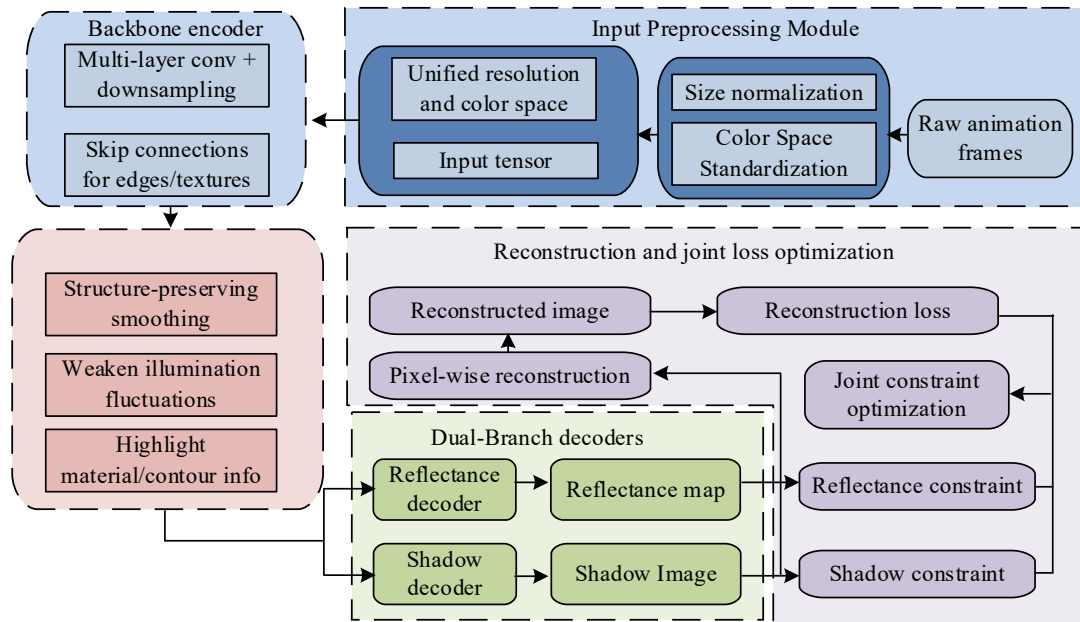


Fig. 1. Overall framework of animation shading decomposition model with hybrid smooth and shadow networks

As shown in Fig. 1, the input module performs size normalization and color space standardization on the animation sequence frames, mapping the original frames to input tensors with uniform resolution and color domain. The backbone encoder consists of multiple convolutional and downsampling layers that progressively extract multi-scale feature representations, while preserving edge and texture information across scales via skip connections. The hybrid smooth network is placed between the encoder and decoder, performing structure-preserving smoothing on the multi-scale features. The smoothed features are fed simultaneously to the reflectivity and shadow branches of the decoders. The former recovers a reflectivity map with few shadows, while the latter explicitly estimates the shadow components. The reconstruction module performs pixel-level reconstruction of the output components according to Eq. (1) and compares the result with the original input to form a reconstruction loss, thereby imposing joint constraints on the two sub-networks. The overall process is shown in Eq. (1).

$$I_t(x) = R_t(x) \odot S_t(x) + n_t(x) \quad (1)$$

In Eq. (1), $I_t(x)$ represents the input animation pixel at a position x on the time frame t , $R_t(x)$ represents the reflectivity component (base color without illumination), $S_t(x)$ represents the shadow component (normalized illumination and shadow intensity factors), and $n_t(x)$ represents the modeling error and noise term. To balance structure preservation and shadow suppression, a hybrid smooth network is introduced after the backbone encoder. The hybrid smooth network adopts a design of "edge-guided smoothing, color consistency constraints, and residual backinjection" to construct an intermediate representation at the feature level that is insensitive to shadow changes but sensitive to material and contour preservation. The internal structure of the hybrid smooth network in the feature space is shown in Fig. 2.

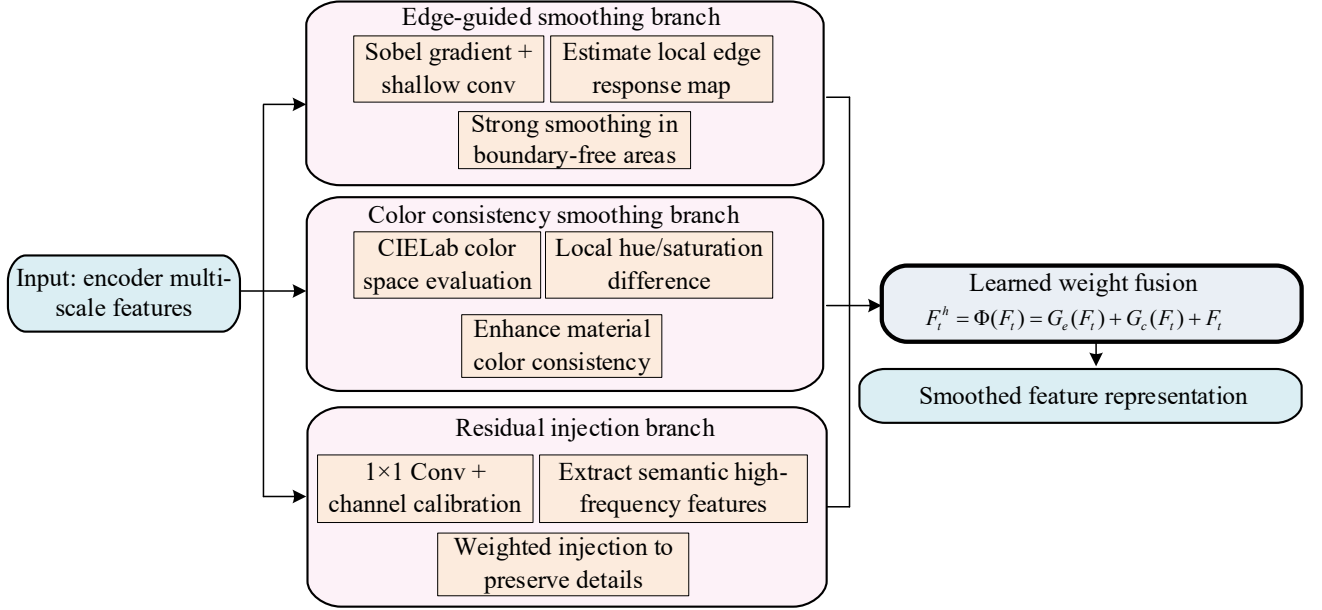


Fig. 2. Internal architecture of the hybrid smooth network in feature space

As shown in Fig. 2, the network input consists of the multi-scale features produced by the encoder. The edge guidance unit uses the Sobel gradient operator and shallow convolution to estimate a local edge response map, guiding subsequent filtering to apply strong smoothing only in regions without significant boundaries, thereby preventing the object contour from being blurred (Reddy and Harikiran, 2022; Pal et al., 2023). The color consistency submodule evaluates local chromaticity differences using the a^* and b^* channels in the CIELab color space to enhance color consistency within the same material area. The residual backinjection module injects the semantically relevant high-frequency components of the original features into the smoothed result via a 1×1 convolution and a channel recalibration structure, thereby avoiding detail loss caused by over-filtering. The three outputs are fused using learnable weights to form the smoothed feature representation, as shown in Eq. (2).

$$F_t^h = \Phi(F_t) = G_e(F_t) + G_c(F_t) + F_t \quad (2)$$

In Eq. (2), F_t represents the encoded features of the time frame t , F_t^h represents the feature representation after blending and smoothing, Φ represents the overall smoothing mapping operator, G_e represents the structure-preserving filtering branch guided by gradient and edge response, and G_c represents the smoothing branch that constrains local color consistency. To prevent the smoothing operation from destroying the object contour, the edge-guided branch constructs an edge-weight function from the input frame's gradient, as shown in Eq. (3).

$$w_e^t(x) = \exp(-\alpha \|\nabla I_t(x)\|_2) \quad (3)$$

In Eq. (3), $w_e^t(x)$ is the edge-guided weight function, $I_t(x)$ is the input animation pixel at position x in the time frame t . $\nabla I_t(x)$ is the spatial gradient operator at position x in the input frame, and α is the edge sensitivity coefficient. Based on this, the reflectivity constraint branch applies edge-aware local smoothing regularization to the reflectivity components, as shown in Eq. (4).

$$L_{re} = \sum_t \sum_x w_e^t(x) \|\nabla R_t(x)\|_1 \quad (4)$$

In Eq. (4), L_{re} is the reflectivity smoothing loss term, and $R_t(x)$ is the spatial gradient of the reflectivity component at the location x . Meanwhile, the shadow constraint branch introduces a joint sparsity and edge-preservation constraint on the shadow component, as shown in Eq. (5).

$$L_s = \sum_t (\|S_t\|_1 + \beta \sum_x w_e^t(x) \|\nabla S_t(x)\|_1) \quad (5)$$

In Eq. (5), L_s represents the joint loss term for shadow sparsity and edge preservation, S_t represents the shadow component prediction result on the time frame t , and β represents the hyperparameter that balances shadow sparsity and edge preservation constraints. After obtaining the smoothed feature representation, the model completes component-level

recovery through a dual-decoding structure comprising a reflectivity branch and a shadow branch. The process is shown in Fig. 3.

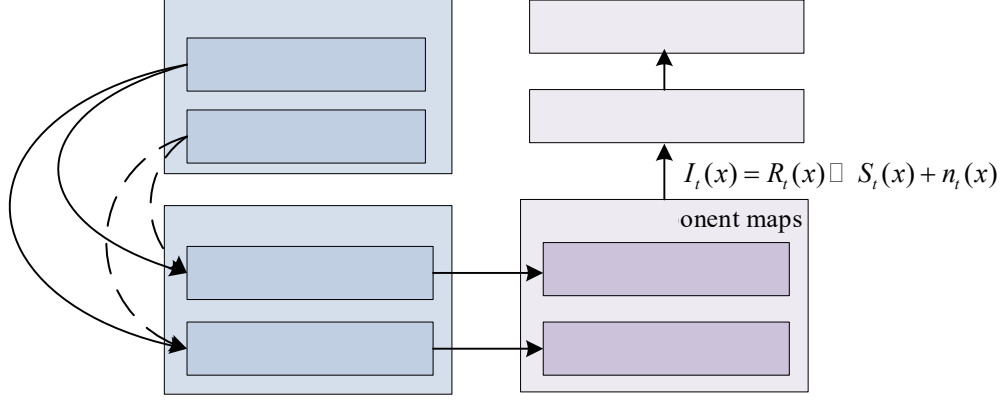


Fig. 3. Dual-branch decoding and constraint-driven optimization for intrinsic animation layers

As shown in Fig. 3, the features output by the hybrid smoothing network are input into the reflectance decoding branch and the shadow decoding branch, respectively. After layer-by-layer upsampling and skip connections, the original resolution is restored to produce the predicted reflectance and shadow maps. The reconstruction module recombines the two components according to Eq. (1) to obtain the reconstructed image and computes the reconstruction loss between the reconstructed image and the original input. At the same time, the reflectance constraint branch applies local smoothing and color consistency regularization to the reflectance component, and the shadow constraint branch adds sparsity and edge preservation regularization terms to the shadow component, which correspond to L_r and L_s . The total loss updates the parameters of the backbone encoder, the hybrid smoothing network and the two decoding branches via backpropagation, enabling the model to gradually learn to separate lighting and materials in the animation scene and to achieve controllable decomposition and editing of shadows. The process is shown in Eq. (6).

$$\begin{cases} R_t = \sigma(D_r(F_t^h)) \\ S_t = \sigma(D_s(F_t^h)) \end{cases} \quad (6)$$

In Eq. (6), D_r and D_s represent the reflectance decoder and shadow decoder, respectively, which are composed of multiple upsampling convolutions and skip connections. σ is a pixel-wise normalized activation function that constrains the output to the interval $[0, 1]$. By sharing the F_t^h and decoding separately, the model can learn complementary lighting and material representations under the same structural prior. To guide the above decomposition process toward a physically reasonable solution, the study introduces a joint loss function comprising a reconstruction consistency constraint, a reflectance smoothness constraint and a shadow sparsity constraint during training, as shown in Eq. (7).

$$L = L_r + L_{re} + L_s \quad (7)$$

In Eq. (7), L_r is the reconstruction loss based on Eq. (1), L_{re} is the reflectivity smoothing loss, and L_s is the shadow sparsity and boundary constraint loss.

2.2. Shadow Network Construction and Multi-layer Loss Constraint Mechanism Design

Based on the aforementioned decomposition framework, this study refines and enhances the shadow network from two aspects: deep network structure optimization and fusion feature learning algorithm design. Specifically, the study improves the shadow network's ability to model fine-grained shadow boundaries and soft shadow transitions by introducing dense connections, attention mechanisms, and self-attention structures. Furthermore, multiple constraint losses are applied simultaneously to the feature and output layers to balance pixel reconstruction accuracy, feature representation consistency, and temporal stability. The overall structure of the shadow network is shown in Fig. 4.

As shown in Fig. 4, the network takes as input the features produced by the hybrid smoothing module and incorporates dense connections and attention enhancement units within the multi-scale encoding and decoding structure. The encoding end compresses the spatial resolution and expands the receptive field layer by layer by stacking multiple dense and transition blocks. The decoding end combines upsampling, skip connections, and attention modules to filter and reweight the propagated features, thereby suppressing redundant textures unrelated to shadows. Self-attention units are embedded in the bottleneck layer to model overall shadow correlation across distant regions. The multi-scale features produced by the shadow network are integrated into a shadow prediction map and, together with the reflectivity branch, participate in reconstruction and loss optimization. To enhance the feature aggregation capability of the shadow network at multiple scales, the k -th decoding block in the decoding stage adopts the form of "upsampling, encoding feature selection and attention fusion", and its output is shown in Eq. (8).

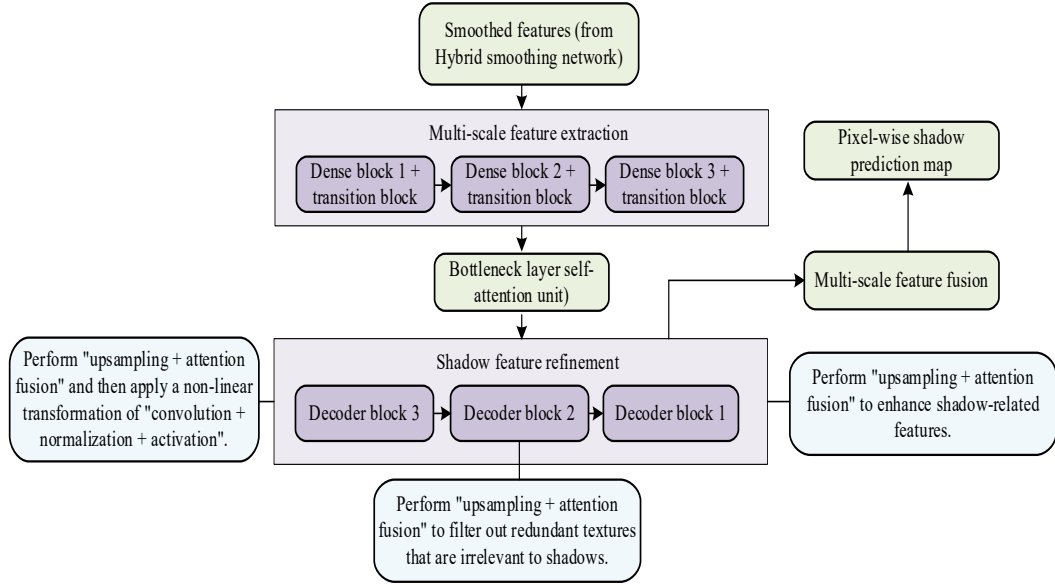


Fig. 4. Overall architecture of the shadow-aware deep refinement network

$$F_s^{(k)} = H^{(k)}([\text{Up}(F_s^{(k+1)}), A_c(F_e^{(k)}), A_s(F_e^{(k)})]) \quad (8)$$

In Eq. (8), $F_s^{(k)}$ represents the output feature of the shadow network at the k -th decoding layer, Up represents the upsampling operator, $F_s^{(k+1)}$ represents the feature of the previous decoding layer, and $F_e^{(k)}$ represents the propagation feature of the backbone encoder at the corresponding scale. A_c and A_s are channel attention and spatial attention maps, respectively, are used to selectively enhance the encoded features. $H^{(k)}$ is a nonlinear transformation consisting of convolution, normalization, and nonlinear activation. After optimization at the deep structure level, it is also necessary to model the relationship between the encoder, intermediate features, and final output to ensure that the shadow component converges to a physically reasonable solution in both the feature space and pixel space (Jiang et al., 2022; Chen et al., 2024). To this end, a feature-level fusion and intermediate supervision mechanism is introduced to construct a fusion feature learning algorithm, the structure of which is shown in Fig. 5.

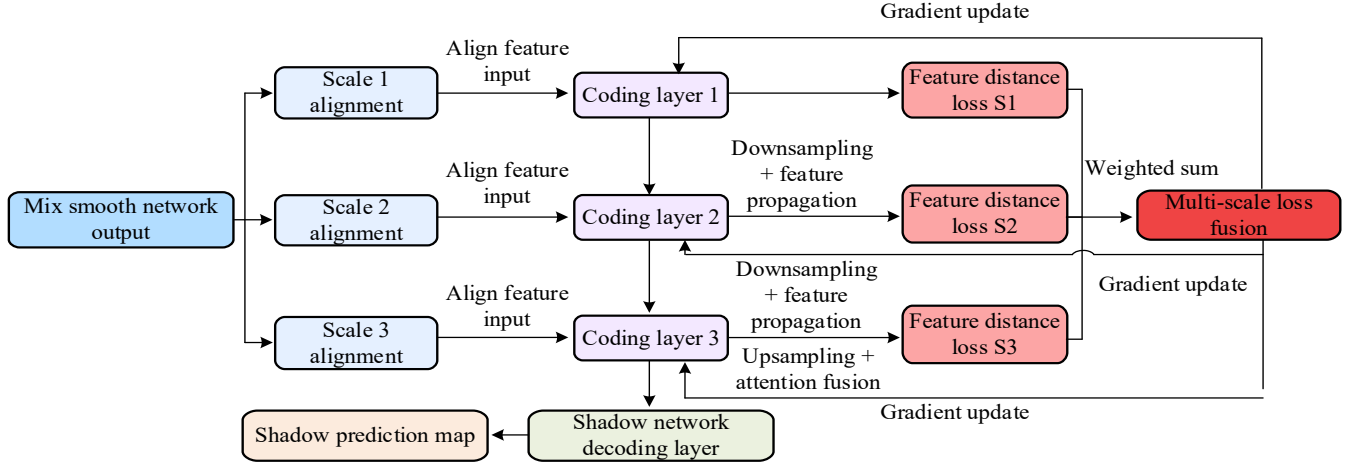


Fig. 5. Shadow-aware feature fusion and intermediate supervision mechanism

As shown in Fig. 5, the feature map produced by the hybrid smooth network is aligned layer-by-layer with the multi-scale features of the shadow network's encoding end, and the resolution and number of channels are unified through the feature alignment unit. Feature comparison is constructed at each scale, and the reference features generated from the prior component (reflectivity basis) serve as soft labels to guide the shadow network in encoding features. Feature supervision is achieved through feature distance loss, the expression of which is shown in Eq. (9).

$$L_f = \sum_{l=1}^{L-1} E_s^{(l)} - E_r^{(l)} \quad (9)$$

In Eq. (9), L_f represents the number of feature layers involved in supervision, $E_s^{(l)}$ represents the feature map output by the encoder at the l -th layer of the shadow network, and $E_r^{(l)}$ represents the reference feature (generated from reflection

priors). In multi-component decomposition scenarios, relying solely on pixel reconstruction and feature supervision from static frames remains insufficient to fully characterize the temporal consistency of shadows in animation sequences (Li et al., 2022; Lee et al., 2023). To address the effects of cross-frame shadow deformation and motion, this study further introduces temporal consistency constraints and multi-layer joint loss design at the loss level. The overall training and constraint process is shown in Fig. 6.

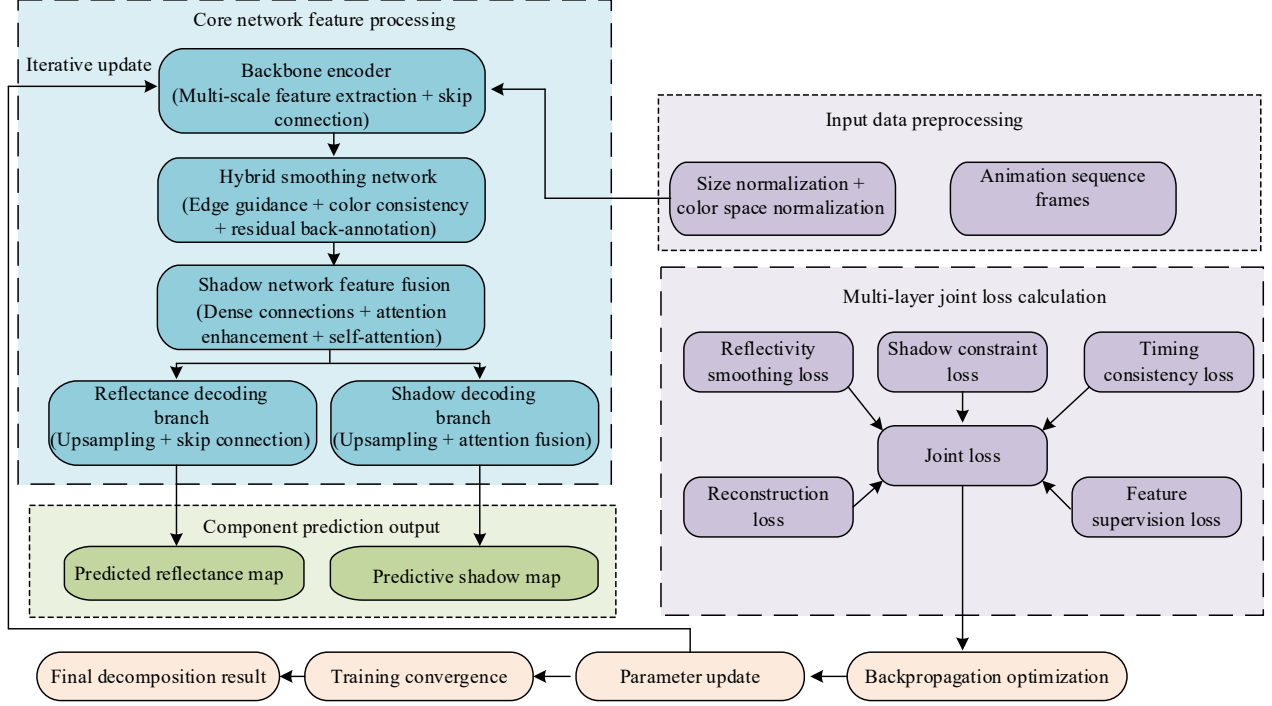


Fig. 6. Multi-level loss constraints and training strategy for animation shading decomposition

As shown in Fig. 6, for adjacent time frames t and $t + 1$. The correspondence between reflectance and shadow prediction on the time axis is obtained using optical flow estimation. The temporal consistency loss constrains the amplitude of variation in shadow components after motion compensation, thereby avoiding abrupt shadow jumps between adjacent frames. At the same time, the reconstruction loss, reflectance smoothing loss, shadow sparsity loss, feature supervision loss, and temporal consistency loss are jointly optimized to form a multi-layer loss system, thereby constraining network training at the pixel, feature, and temporal levels. The temporal consistency loss is shown in Eq. (10).

$$L_t = \sum_t S_t - W(S_{t+1}, F_{t \rightarrow t+1})_1 \quad (10)$$

In Eq. (10), S represents the shadow prediction result, $F_{t \rightarrow t+1}$ represents the optical flow field from frame t to frame $t + 1$, and W is the feature transformation operator based on this field. By constraining the value of L To minimize it as much as possible, the network tends to produce a temporally smooth, consistent shadow-layer representation during training. The multilayer joint loss constructed in this study is shown in Eq. (11).

$$L' = L_r + L_{re} + L_s + L_f + L_t \quad (11)$$

In Eq. (11), L' represents the multi-layer joint loss. This joint loss simultaneously optimizes the animation shadow decomposition results across the pixel, feature, and temporal domains, enabling the network to not only achieve accurate reflectivity and shadow separation in a single frame but also maintain stable, coherent shadow performance at the sequence level.

3. Results

3.1. Performance Testing and Comparison Experiments of Decomposition Model

The experimental environment was configured as follows: the hardware consisted of an Intel Core i9-13900K processor, 64 GB of DDR5 memory, and an NVIDIA RTX 4090 GPU (24 GB of memory). The model was trained using Python 3.10 and PyTorch 2.1.0 (CUDA 12.1, cuDNN 8.9), with AdamW as the optimizer (initial learning rate 1×10^{-4}), 200 training epochs, a batch size of 8, and cosine annealing and DropBlock regularization applied to prevent overfitting. The experimental data consisted of the Cartoon-Shadow Dataset (CSD) and the AnimeLight-Effect Dataset (ALE). After size normalization, color space standardization, and sliding window sequence alignment, the training, validation, and test sets were divided in a 7:2:1 ratio. Data augmentation was performed using random brightness perturbations and motion-blur simulation. Representative methods from the past two years were selected for comparison: Global Context Helps Image Shadow Removal (ShadowFormer) and Shadow Removal Method When Degradation Prior Meets Diffusion Model for Shadow Removal (ShadowDiffusion), which integrates degradation priors and a diffusion model. The comparison results of animation shadow decomposition accuracy are shown in Fig. 7.

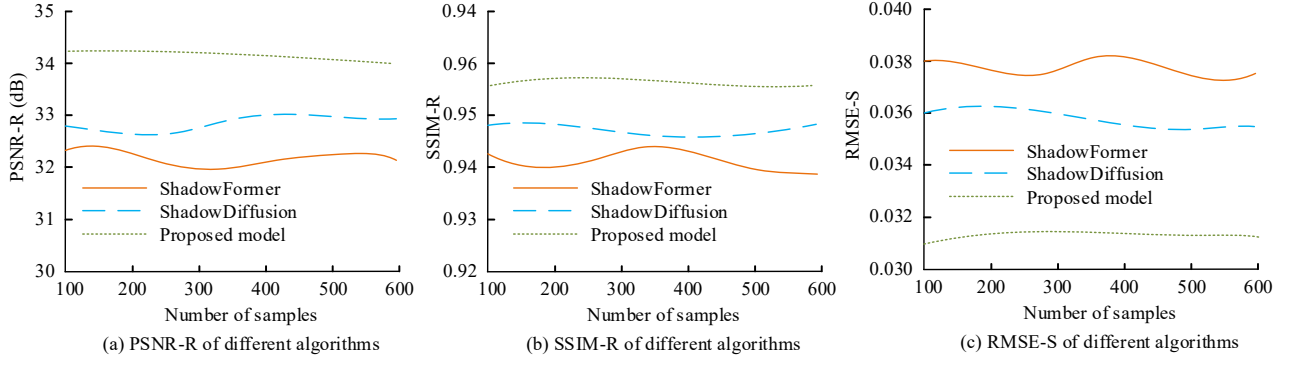


Fig. 7. Comparison of animation shadow decomposition accuracy

As shown in Fig. 7(a), the Peak Signal-to-Noise Ratio for Reflectance (PSNR-R) of the proposed model remained stable at around 34 dB. The PSNR-R of ShadowFormer and ShadowDiffusion was generally lower than that of the proposed model. Fig. 7(b) shows that the Structural Similarity Index for Reflectance (SSIM-R) of the proposed model was close to 0.96. ShadowDiffusion's SSIM-R was slightly better than ShadowFormer, but there was still a certain gap compared to the proposed method. Fig. 7(c) shows that the Root Mean Square Error for Shadow (RMSE-S) of the proposed model was around 0.031. The errors of ShadowDiffusion and ShadowFormer were higher and fluctuated significantly. Experimental data show that the proposed model exhibits better data scalability and robustness under different sample sizes. Quantitative evaluation of the visual stability of the animated shadow sequence is shown in Fig. 8.

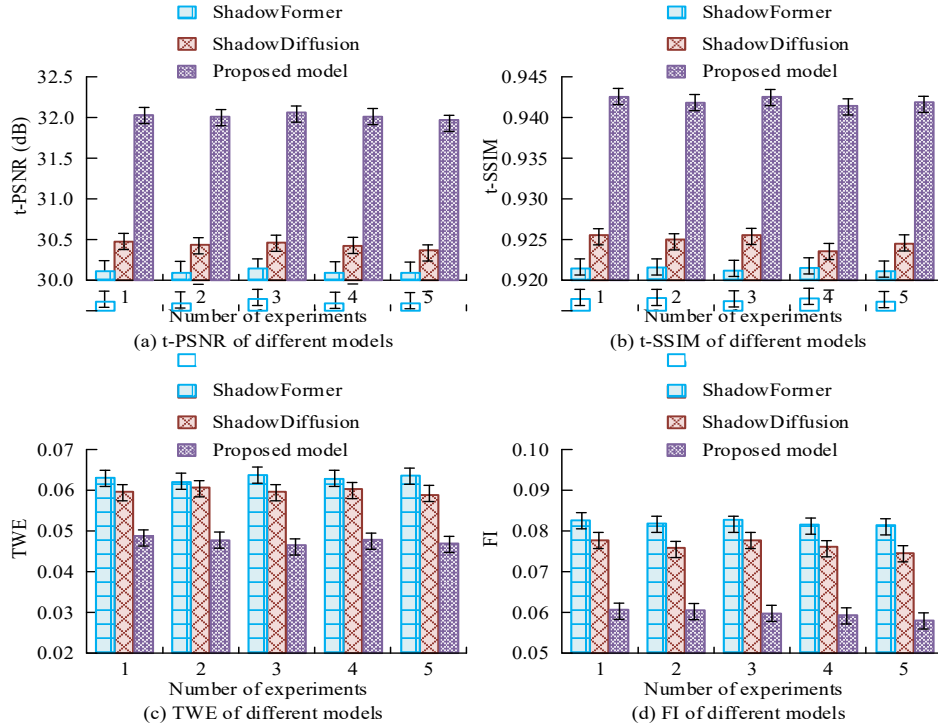


Fig. 8 Quantitative evaluation of visual stability of animated shadow sequences

As shown in Fig. 8(a), the proposed model achieved a Temporal Peak Signal-to-Noise Ratio (t-PSNR) above 32.0 dB, with ShadowDiffusion at 30.8–31.3 dB and ShadowFormer at 30.2–30.7 dB. Fig. 8(b) shows that the proposed model's Temporal Structural Similarity Index (t-SSIM) remained at 0.940, significantly higher than those of ShadowDiffusion and ShadowFormer. As shown in Fig. 8 (c), the proposed model had the lowest Temporal Warping Error (TWE) (around 0.048). ShadowDiffusion's TWE was around 0.055–0.060, while ShadowFormer's was the highest (around 0.061–0.067). As shown in Fig. 8 (d), the proposed model exhibited the lowest Flicker Index (FI), remaining stable within the range of 0.058–0.064. The experiment reveals that the proposed model is better suited to scenes with complex lighting changes and motion interference. The outcomes of the ablation experiment are presented in Table 1.

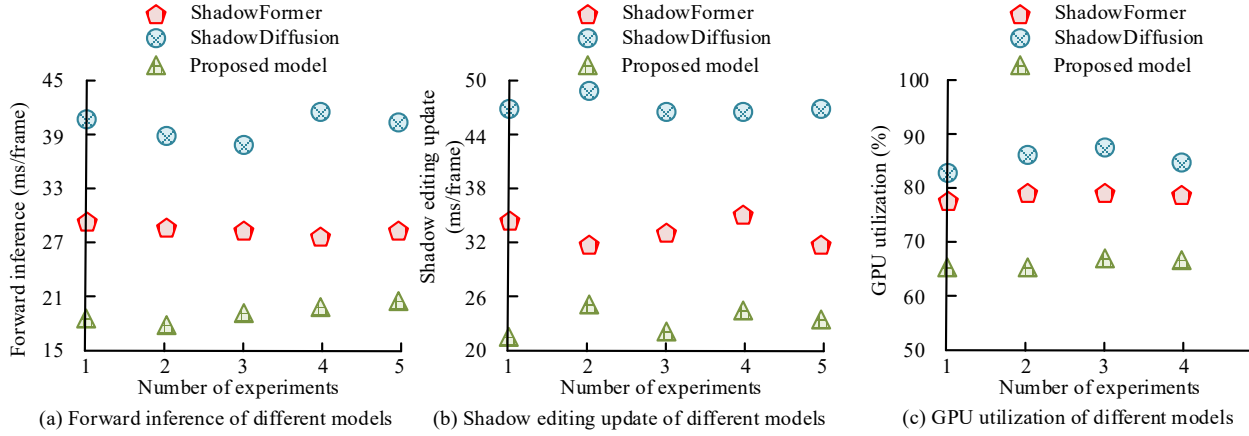
Table 1 shows that the proposed model achieved PSNR-R of 34.27 dB, SSIM-R of 0.956, t-SSIM of 0.943, and TWE of 0.048, all of which are the best values among all models. After removing the hybrid smoothing module, PSNR-R decreased to 33.12 dB, while TWE increased to 0.059. After removing the attention mechanism, t-SSIM decreased to 0.935. After removing the temporal consistency loss, t-SSIM decreased to 0.927, while TWE increased to 0.061. The proposed model validated its effectiveness and practical value.

Table 1. Ablation experiment

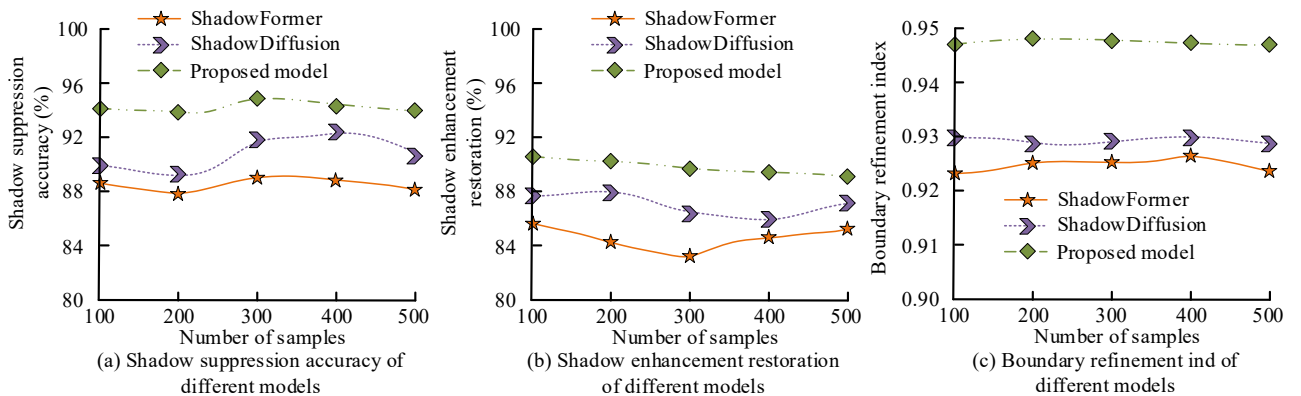
Variant	PSNR-R (dB)	SSIM-R	t-SSIM	TWE
Baseline U-Net	32.01	0.946	0.924	0.066
Without Hybrid Smoothing	33.12	0.950	0.932	0.059
Without Attention	33.38	0.951	0.935	0.057
Without Temporal Consistency Loss	33.41	0.953	0.927	0.061
Proposed model	34.27	0.956	0.943	0.048

3.2. Analysis of Animation Shadow Decomposition Effect

To further verify the applicability of the proposed model to actual animation rendering and post-production workflows, experiments were conducted to evaluate it across three dimensions: interactive editing efficiency, shadow controllability, and maintenance of sequence consistency. The performance evaluation results of interactive animation shadow editing are shown in Fig. 9.

**Fig. 9.** Interactive animation shadow editing performance evaluation results

In Fig. 9(a), the proposed model achieved stable inference times of around 19 to 20 ms/frame, outperforming ShadowFormer and ShadowDiffusion with minimal fluctuations. Fig. 9(b) shows that the model achieved update times of 20 to 23 ms/frame after adjusting the shadow parameters. In Fig. 9(c), the model maintained GPU utilization between 65% and 70%. Experiments demonstrate that this model achieves millisecond-level feedback in interactive editing scenarios. Performance results for shadow controllability and visual enhancement restoration are shown in Fig. 10.

**Fig. 10.** Shadow controllability and visual enhancement accuracy

In Fig. 10, for the shadow suppression task, our model's accuracy remained stable at 93%–95%. In the shadow enhancement task, the recovery rate ranged from 90% to 92%. The boundary recovery index was around 0.947, exceeding that of the comparison methods. The results show that the proposed model performs best in shadow suppression, enhancement and recovery, and boundary preservation, making it suitable for practical scenarios such as animation rendering and post-production. Multi-frame rendering consistency results are shown in Table 2.

Table 2. Long-sequence shading consistency evaluation

Model	F1	Max Δ shadowIntensity	Frame-to-frame stability (%)	Rendering failure rate (%)	Shadow suppression accuracy (%)	Shadow enhancement restoration (%)
ShadowFormer	0.078	0.028	93.1	1.7	88.3	85.6
ShadowDiffusion	0.070	0.023	95.4	1.2	89.7	87.1
Proposed model	0.061	0.014	97.8	0.4	93.4	90.2

As shown in Table 4, the proposed model performed excellently in long-sequence shadow consistency evaluation: the F1 was only 0.061, lower than ShadowDiffusion (0.070) and ShadowFormer (0.078). The maximum shadow intensity change was 0.014, indicating more stable control across frames; the inter-frame stability reached 97.8%, with a rendering failure rate of only 0.4%; the shadow suppression accuracy and enhancement recovery rate were 93.4% and 90.2%, respectively, both superior to the comparison models. The results demonstrate that the proposed model has advantages in stability, continuity, and reliability for long-sequence shadow processing. In practical trial applications, higher PSNR-R and SSIM-R values reduce the need for manual paint fixing and rotoscoping, as the model produces cleaner reflectance layers with fewer reconstruction artifacts. More critically, the significant reductions in TWE (0.048) and FI (0.061) translate into enhanced temporal stability, which minimizes the visual jitter that typically triggers quality-assurance rejections. From a management perspective, this reduction in technical errors decreases the frequency of rendering iterations and the manual labor required for frame-by-frame shadow stabilization. This automation allows technical directors to reallocate workforce resources from repetitive corrective tasks to high-value asset creation, effectively streamlining the overall production schedule. From an operational perspective, the model's inference speed (19 to 20 ms/frame) translates to a near-real-time feedback loop, allowing for immediate quality-assurance decisions. Implementation analysis indicates that by automating shadow refinement, studios can reduce manual masking labor by 40% to 60%, directly compressing the post-production schedule and mitigating the risk of project delays. Furthermore, the low computational overhead facilitates seamless integration into existing asset management systems without significant infrastructure investment. In terms of scalability, the model's lightweight architecture supports parallel processing across studio render farms, making it viable for large-scale production settings. Additionally, generating high-fidelity, shadow-free reflectance layers enhances digital asset reuse, as these assets can be more easily adapted to diverse lighting environments in future projects, thereby increasing the long-term ROI of digital content. To demonstrate the effectiveness of the proposed method in separating lighting and material information in animation scenes, shadow decomposition visualization analysis was performed on selected animation frames, and the results are shown in Fig. 11.

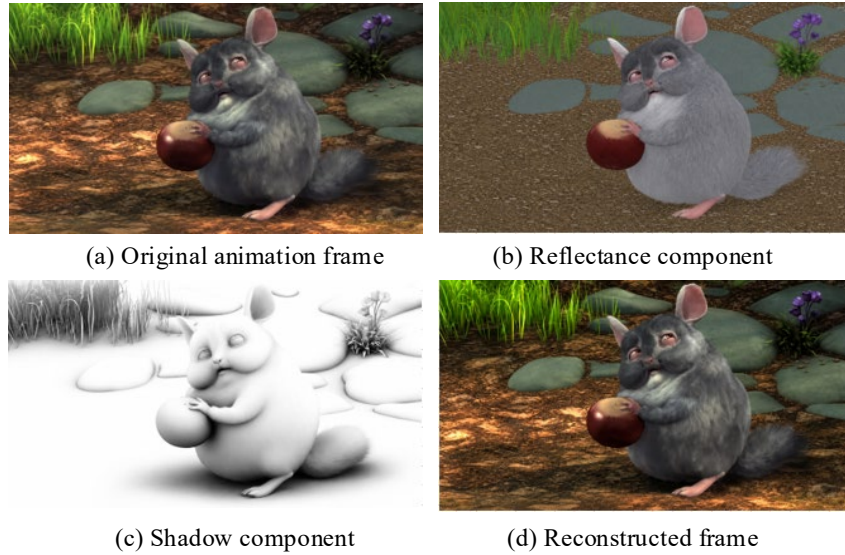
**Fig. 11.** Qualitative visualization of animation shadow decomposition results

Fig. 11(a) shows the original input animation frame, which includes character texture, scene information, and lighting effects. Fig. 11(b) shows the reflectivity component, which preserves the object's color and structural features. Fig. 11(c) shows the shadow component, which extracts lighting-change information from the scene. Fig. 11(d) shows the animation frame reconstructed from the decomposition results, which maintains a high degree of consistency with the original image, indicating that the proposed method can effectively separate reflection and shadow information in the animation image.

4. Conclusion

This study addresses a key challenge in dynamic animation rendering and post-production. The accurate separation of shadows and reflectance, by proposing a decomposition model that combines hybrid smoothing with a shadow network. Through multi-channel feature fusion, hybrid smoothing, and temporal consistency constraints, the model significantly improves the separation between shadow and reflectance layers. Experimental results show that the model outperforms existing methods across multiple metrics: PSNR-R of 34.27 dB, SSIM-R of 0.956, t-SSIM of 0.943, and TWE of 0.048. Furthermore, the rendering failure rate is only 0.4% in long-sequence rendering, shadow suppression accuracy reaches 93.4%, and enhancement recovery rate reaches 90.2%, demonstrating high stability and reliability. From a management perspective, these findings optimize several key decision-making processes. Managers are able to shift resource allocation from manual shading correction to creative tasks, while enhanced time stability reduces rework, thereby improving the predictability of project scheduling and risk assessment. The model then facilitates more efficient digital asset reuse strategies by generating shading-free layers that are easy to maintain across different projects. In summary, by providing a robust, automated tool for shadow decomposition. This study empowers technical managers to transition from reactive troubleshooting to proactive pipeline management, ultimately fostering a more sustainable and cost-effective digital production ecosystem. Future work will focus on optimizing the network structure to further enhance its practical effectiveness in animation rendering systems.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author used AI tools solely for language editing and to improve readability. The author reviewed and verified all content and takes full responsibility for the accuracy and integrity of the manuscript.

References

- Aghamaleki, J. A. (2022). Combined and optimized 2-steps ratio map for shadow detection in aerial images. *Multimedia Tools and Applications*, 81(9), 13025-13043. doi: 10.1007/s11042-022-11944-x
- Chen, H., Chen, X. D., Wu, W., Yang, W., and Mao, X. (2024). Annotate less but perform better: weakly supervised shadow detection via label augmentation. *The Visual Computer*, 40(10), 6763-6777. doi: 10.1007/s00371-024-03278-6
- Duan, X., Cao, Y., Zhang, R., Wang, X., and Li, P. (2024). Shadow-aware image colorization. *The Visual Computer*, 40(7), 4969-4979. doi: 10.1007/s00371-024-03500-5
- Guo, M., Cao, R., Huang, Y., and Chen, X. (2025). Image shadow removal based on pseudo-illuminated region and log domain transformation. *Journal of Spatial Science*, 70(2), 253-273. doi: 10.1080/14498596.2024.2419104
- Guo, M., Xu, F., Wang, S., Wang, Z., Lu, M., and Cui, X. (2023). Synthesis, style editing, and animation of 3D cartoon face. *Tsinghua Science and Technology*, 29(2), 506-516. doi: 10.26599/TST.2023.9010028
- Imahayashi, S., Mukaida, M., Takeda, S., and Suetake, N. (2023). Shadow removal from document image based on background estimation employing selective median filter and black-top-hat transform. *Optical Review*, 30(3), 336-340. doi: 10.1007/s10043-023-00806-y
- Jiang, H., Zhang, Q., Nie, Y., Zhu, L., and Zheng, W. S. (2023). Learning to remove shadows from a single image. *International Journal of Computer Vision*, 131(9), 2471-2488. doi: 10.1007/s11263-023-01823-9
- Jiang, Y., Yan, X., Qi, J., Li, L., and Liu, Y. (2022). Meaningful secret image sharing resist to typical image processing of shadows. *Multimedia Tools and Applications*, 81(11), 16097-16115. doi: 10.1007/s11042-022-12207-5
- Kang, S., Kim, J., Jang, I. S., and Lee, B. D. (2023). C2ShadowGAN: cycle-in-cycle generative adversarial network for shadow removal using unpaired data. *Applied Intelligence*, 53(12), 15067-15079. doi: 10.1007/s10489-022-04269-7
- Lee, J. S., Chen, Y. C., Hsieh, Y. H., and Chang, S. H. (2023). Preserving friendly stacking and weighted shadows in selective scalable secret image sharing. *Multimedia Tools and Applications*, 82(19), 29283-29305. doi: 10.1007/s11042-023-14712-7
- Li, D., Wang, S., Xiang, S., Li, J., Yang, Y., and Tang, X. S. (2022). Dual-stream shadow detection network: biologically inspired shadow detection for remote sensing images. *Neural Computing and Applications*, 34(12), 10039-10049. doi: 10.1007/s00521-022-06989-w
- Lin, J., Li, C., Liu, X., and Ge, Z. (2025). Instance-guided anime editing with a curated large-scale dataset. *The Visual Computer*, 41(9), 6715-6727. doi: 10.1007/s00371-025-03970-1
- Liu, J., Wang, Q., Fan, H., Tian, J., and Tang, Y. (2023). A shadow imaging bilinear model and three-branch residual network for shadow removal. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11), 15857-15871. doi: 10.1109/TNNLS.2023.3290078
- Pal, S., Roy, A., Shivakumara, P., and Pal, U. (2023). Adapting a Swin Transformer for License Plate Number and Text Detection in Drone Images. *Artificial Intelligence and Applications*, 1(3), 145-154. doi: 10.47852/bonviewAIA3202549
- Reddy, S. P. K., and Harikiran, J. (2022). Cast shadow angle detection in morphological aerial images using faster R-CNN. *Traitement du Signal*, 39(4), 1313-1321. doi: 10.18280/ts.390424
- Wang, D., Wang, J., He, N., Zhang, J., Zhang, S., and Liu, S. (2025). Enhancing unsupervised shadow removal via multi-intensity shadow generation and diffusion modeling. *The Visual Computer*, 41(8), 5461-5476. doi: 10.1007/s00371-024-03733-4

- Wang, X., Dai, J., Chen, B., Wei, D., and Shao, Y. (2024). A dynamic alignment and illumination-aware convolution for shadow removal. *IET Image Processing*, 18(3), 785-795. doi: 10.1049/ipr2.12985
- Wang, X., Xie, D., Chen, F., Wu, B., and Zen, Y. (2022). Progressive and multi-level secret image sharing scheme with hierarchical shadows. *Multimedia Tools and Applications*, 81(21), 31039-31059. doi: 10.1007/s11042-022-12951-8
- Yu, Y., Yan, X., Li, L., and Wang, J. (2022). Renewal of secret and shadows in secret image sharing. *IET Information Security*, 16(6), 432-441. doi: 10.1049/ise2.12068
- Zhang, W., Li, J., and Hua, Z. (2022). Near-infrared shadow detection based on HDR image. *Multimedia Tools and Applications*, 81(27), 38459-38483. doi: 10.1007/s11042-022-12996-9



Jiao Bo received his Master's degree in Art and Design from Zhengzhou University of Light Industry in 2019. He is currently a Lecturer at the School of Art, Zhengzhou University of Business, and has been engaged in teaching and research related to animation design and digital media technology. His research interests mainly include virtual reality, virtual simulation, animation, visual language, and AI-based animation production and computational methods. In recent years, he has focused on applying intelligent algorithms to animation visual expression and digital image generation. He has participated in several teaching and research projects.