Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 15 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i15.62711

PAPER

An Optimization Model for Augmented Reality-Enhanced Interactive Landscape Experiences

Yunzhi Tian (✉)

Huanghuai University, Zhumadian, China

tianyz1983@163.com

ABSTRACT

Mobile augmented reality applications in outdoor landscape environments have long been constrained by limited on-device computational resources and communication bandwidth, resulting in localization drift, insufficient rendering smoothness, and homogeneous interactive experiences. To address these challenges, an augmented reality experience optimization model for landscape interaction was proposed in this study based on a cloud–edge–device collaborative architecture. Within the model, visual–inertial odometry was integrated with three-dimensional Gaussian splatting. An adaptive weighted selection mechanism was introduced to reduce redundant rendering primitives, while camera poses were refined through scene-prior constraints, thereby suppressing localization errors in complex outdoor environments and reducing transmission overhead. To further enhance user experience, multidimensional features derived from spatial motion, visual gaze behavior, and physiological perception were fused to construct a cross-modal attention prediction network. A hybrid training strategy combining offline imitation learning and online incremental optimization was employed, enabling adaptive adjustment of augmented reality rendering parameters and facilitating closed-loop regulation of user experience in dynamic environments. In addition, a global optimization model was established with quality of experience maximization as the primary objective while jointly considering latency and energy-consumption constraints. Lightweight mobile deployment was achieved through an asynchronous rendering pipeline and parameter-quantization encoding scheme.

KEYWORDS

mobile augmented reality, three-dimensional Gaussian splatting, visual–inertial simultaneous localization and mapping, cloud–edge–device collaboration, multimodal affective computing, lightweight mobile deployment

1 INTRODUCTION

Mobile augmented reality has emerged as a critical enabling technology for the digital transformation of outdoor landscapes and the intelligent upgrading of cultural

Tian, Y. (2026). An Optimization Model for Augmented Reality-Enhanced Interactive Landscape Experiences. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(15), pp. 135–149. https://doi.org/10.3991/ijim.v20i15.62711

Article submitted 2026-03-17. Revision uploaded 2026-04-29. Final acceptance 2026-05-14.

tourism services [1–3]. By seamlessly integrating virtual content with real-world environments, novel forms of interactive outdoor experiences can be delivered. However, the interaction quality and practical deployment of landscape-oriented augmented reality systems continue to be constrained by the combined challenges of complex outdoor environments and the limited computational resources of mobile devices [4]. Outdoor landscape scenes are typically characterized by weak-texture regions, illumination variations, and dynamic occlusions, all of which can adversely affect visual–inertial localization performance. Consequently, cumulative localization drift may be introduced, leading to defects such as virtual-real misalignment and image jitter [5, 6]. Meanwhile, the limited computing capability, communication bandwidth, and energy resources of mobile devices make it difficult to support the real-time, high-fidelity rendering of large-scale three-dimensional landscape environments [7, 8]. In addition, most existing landscape augmented reality systems rely on fixed rendering strategies and predefined interaction mechanisms. As a result, presentation effects cannot be adaptively adjusted according to users' real-time states or affective preferences, thereby restricting personalized interaction capabilities. Existing studies have primarily focused on either localization enhancement or rendering optimization in isolation, whereas a closed-loop optimization framework integrating spatial perception, multimodal experience perception, and dynamic feedback has yet to be comprehensively established. Consequently, the theoretical foundation remains incomplete. From an engineering perspective, many high-precision augmented reality solutions are dependent on high-performance computing resources and cloud-based deployment architectures [9, 10], making them difficult to adapt to the lightweight requirements of widely used mobile devices. This limitation has significantly constrained the large-scale adoption of landscape augmented reality technologies. To address these challenges, an adaptive augmented reality interaction optimization framework is developed for outdoor landscape scenarios on mobile platforms. The proposed framework is expected to contribute to the advancement of intelligent interaction theories for mobile augmented reality and to facilitate the industrial deployment of smart landscape applications.

Although substantial progress has been achieved in augmented reality spatial modeling, scene rendering, and edge-assisted deployment technologies, significant limitations remain in the development of integrated optimization frameworks for mobile outdoor landscape applications. Regarding the integration of simultaneous localization and mapping and three-dimensional Gaussian splatting [11], existing approaches have primarily focused on offline scene reconstruction in desktop environments, where reconstruction accuracy has been treated as the primary objective. Consequently, the computational and communication constraints of mobile devices have rarely been considered [12, 13]. Dynamic scene simplification mechanisms suitable for large-scale outdoor landscape environments remain insufficiently explored, resulting in increased latency and rendering instability. Furthermore, the dense geometric priors embedded within three-dimensional Gaussian scenarios have not been effectively exploited for the correction of odometry drift in weak-texture environments. As a result, localization robustness under outdoor conditions remains limited [14]. In the domain of intelligent augmented reality interaction [15], most existing landscape-oriented augmented reality systems provide only basic virtual content overlay functionalities, while rendering configurations and interaction strategies are typically predefined and static. Limited attention has been devoted to the integration of multimodal information derived from user behavior, gaze patterns, and physiological responses. Consequently, variations in users' real-time experiences cannot be accurately identified, and data-driven dynamic regulation and closed-loop feedback mechanisms remain largely absent, thereby constraining the realization of personalized intelligent interaction. From the perspective of edge–device collaborative deployment, optimization efforts within existing mobile augmented reality architectures have generally been confined to communication latency reduction. Comprehensive optimization models jointly considering mobile-device constraints such as computational capacity and energy consumption have rarely been established. In addition, generic lightweight strategies are often unable to accommodate the heterogeneous graphics processing unit architectures of mobile phones. Targeted parameter encoding schemes and dynamic workload regulation mechanisms remain insufficiently developed. Under challenging outdoor conditions characterized by fluctuating network connectivity and dynamically varying computational loads, stable system performance and effective energy management therefore remain difficult to achieve.

To address the aforementioned challenges, a cloud–edge–device collaborative optimization framework for landscape-oriented augmented reality interaction is established, through which both theoretical and engineering innovations are achieved. From a theoretical perspective, a tightly coupled perception framework integrating visual–inertial odometry and three-dimensional Gaussian splatting is developed. Localization drift is effectively suppressed through the incorporation of an adaptive Gaussian selection mechanism and scene-prior constraints. In addition, a cross-modal attention fusion network, together with a hybrid training paradigm, is introduced to enable the dynamic regulation of augmented reality rendering parameters. Furthermore, a global joint optimization model is established with quality of experience maximization as the primary objective while jointly considering latency and energy-consumption constraints, enabling quantitative modeling across the entire processing pipeline.

2 METHODOLOGY

2.1 Cloud–edge–device collaborative architecture and innovative data flow design

A hierarchically decoupled cloud–edge–device collaborative architecture is established to address the computational limitations and communication instability commonly encountered in outdoor mobile augmented reality landscape applications. Temporal synchronization and efficient coordinated scheduling of heterogeneous data streams are achieved through a lightweight communication and scheduling mechanism specifically designed in this study. A differentiated task allocation strategy is adopted in the system. At the device layer, synchronized acquisition of scene images, device poses, inertial measurements, and user interaction data is performed. At the edge layer, graphics processing unit nodes are employed to execute all real-time computational tasks, including pose estimation, Gaussian scene optimization, and multimodal inference. At the cloud layer, only landscape asset archiving and offline model updating are conducted, while participation in real-time data interactions is intentionally avoided to minimize communication overhead within the real-time pipeline. To facilitate standardized end-to-end system modeling, a unified set of state variables is defined. The global spatial position of the device is represented by p_t, the device orientation quaternion is denoted by q_t, the real-time scene image is represented by I_t, and the global three-dimensional Gaussian scene representation is denoted by G_t.

Based on the proposed architecture, a closed-loop iterative data flow is established. Following the transmission of raw sensor data from the mobile device to the edge node, pose correction and scene-data refinement are performed through the joint optimization of visual–inertial odometry and Gaussian map optimization. The optimized scene data are subsequently transmitted back to the mobile device for real-time augmented reality rendering. Simultaneously, user interaction and perceptual information are continuously collected at the device layer and transmitted to the edge layer, where they are utilized to adaptively adjust rendering parameters for subsequent frames. To ensure real-time system responsiveness, an end-to-end latency threshold of 50 ms is imposed as a primary operational constraint. By combining edge-centric computation offloading with lightweight data exchange mechanisms, continuous and stable landscape-oriented augmented reality interaction can be maintained under complex outdoor operating conditions. A schematic illustration of the proposed cloud–edge–device collaborative architecture and data flow is presented in Figure 1.

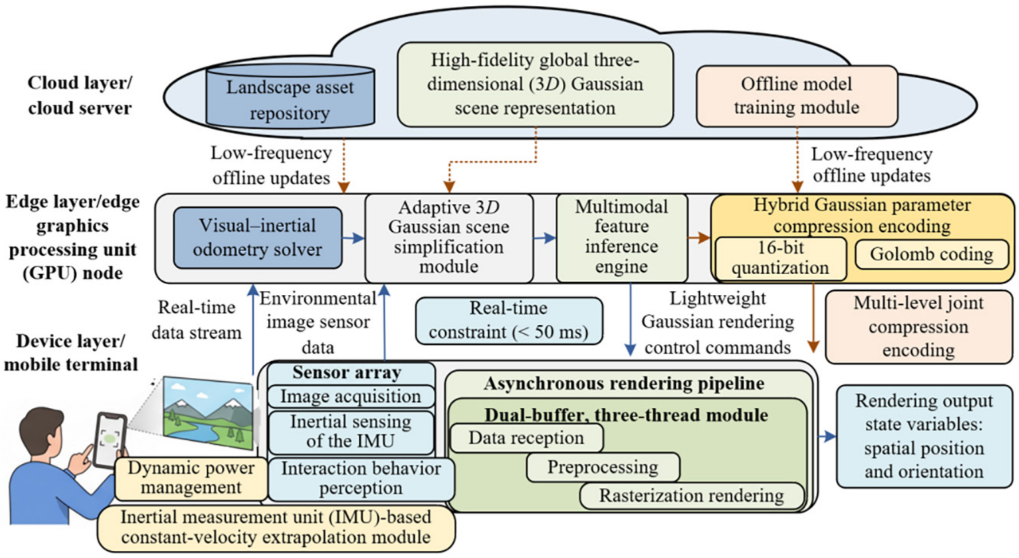

Fig. 1. Schematic illustration of the cloud–edge–device collaborative architecture and data flow

2.2 Mobile spatial perception model based on lightweight integration of visual–inertial odometry and three-dimensional Gaussian splatting

To achieve a balance between outdoor landscape localization accuracy and real-time rendering performance on mobile devices, a tightly coupled integration of visual–inertial odometry and three-dimensional Gaussian splatting is established. Conventional approaches relying on full-scene Gaussian rendering are abandoned in favor of an adaptive selection mechanism, through which lightweight scene representations suitable for mobile deployment are generated. An importance evaluation model for Gaussian primitives is constructed by jointly considering color characteristics, texture gradients, and spatial proximity within landscape environments. High-value rendering primitives are selected through a multidimensional weighted mechanism. The corresponding importance score is defined as:

$$s_j = \left\| c_j - \bar{c} \right\|_2 \cdot \frac{\left| \nabla G_j(\mu_j) \right|}{\max_k \left| \nabla G_k(\mu_k) \right|} \cdot exp\left(-\lambda \left\| \mu_j - p_t \right\|^2 \right) \tag{1}$$

where, $\left\| c_j - \bar{c} \right\|_2$ denotes the color contrast between the Gaussian primitive and the global scene representation, thereby preserving visually salient regions. The gradient term is introduced to select key primitives containing rich texture details. The distance attenuation term prioritizes scene elements located within the near-field region of the camera. To accommodate the computational limitations of mobile hardware platforms, the maximum number of Gaussian primitives rendered per frame is constrained to $N_{max} = 5000$. Through the elimination of redundant scene information while preserving essential visual characteristics of landscape environments, stable real-time rendering at 60 frames per second can be achieved on mobile devices. Consequently, the challenges associated with excessive transmission bandwidth requirements and rendering latency in large-scale outdoor scenes can be effectively mitigated.

To mitigate the cumulative drift commonly encountered by visual–inertial odometry in outdoor environments characterized by weak textures and dynamic occlusions, a pose correction mechanism is established by exploiting the dense geometric priors embedded within the three-dimensional Gaussian scene representation, thereby enhancing the localization robustness of the system. The localization state is continuously monitored using the covariance of the estimated pose trajectory. When $tr\left(\Sigma_{p_t}\right) > \tau_{\mathrm{drift}}$, the occurrence of localization drift is assumed. Subsequently, a reference image generated from the offline Gaussian scene representation is retrieved and matched against the real-time captured image. A mapping relationship between three-dimensional scene coordinates and image pixels is established. Based on the obtained three-dimensional to two-dimensional (3D–2D) correspondences, a geometric constraint residual is formulated as:

$$e_{gs}(T_t) = \sum\nolimits_{(u,v)\in \mathrm{matches}} \left\| \pi(T_t \cdot \mu_j^{ref}) - [u,v]^\top \right\|^2 \tag{2}$$

Pose estimation is subsequently performed within a sliding-window nonlinear optimization framework. A joint optimization objective is constructed by combining visual reprojection errors and inertial measurement unit preintegration errors:

$$\min_X \sum\nolimits_{k\in K} \left\| e_{vis}(X_k) \right\|^2_{\Sigma_{vis}} + \sum\nolimits_{k\in K} \left\| e_{imu}(X_{k-1}, X_k) \right\|^2_{\Sigma_{imu}} \tag{3}$$

The Gaussian prior constraint stably supplies absolute pose correction information during optimization, thereby effectively compensating for accumulated odometry errors. As a result, localization drift in outdoor environments can be substantially reduced, enabling high-precision and robust real-time localization within complex landscape scenarios.

2.3 Multimodal attention-driven affective perception and dynamic augmented reality feedback model

To enable personalized and adaptive regulation of landscape-oriented augmented reality interaction, an end-to-end multimodal perception fusion model is established by integrating heterogeneous sensory information derived from environmental context, user behavior, and physiological states. A 28-dimensional fused feature vector is extracted and subsequently decomposed according to data characteristics into three sub-dimensions, namely spatial environment features, gaze behavior features, and physiological perception features. To address the heterogeneous statistical distributions associated with different modalities, modality-specific feature encoding is performed through independent linear projection. Raw multimodal features are thereby mapped into a unified latent space with consistent dimensionality. The encoding process is formulated as:

$$h_t^{(m)} = ReLU(W_m f_t^{(m)} + b_m), \quad m \in \{\text{spa}, \text{att}, \text{phy}\} \tag{4}$$

where, W_m and b_m denote the modality-specific weight matrix and bias vector, respectively. The distributional characteristics of heterogeneous sensory data can be effectively accommodated. On this basis, a multi-head self-attention mechanism is introduced to capture associated features among different modalities, facilitating the deep coupling and fusion of multi-source information. Layer normalization is subsequently applied to maintain the stability of feature distributions. As a result, a global fused feature representation is obtained:

$$z_t = \text{LayerNorm}(h_t^{\text{spa}} + \text{MultiHead}(h_t^{\text{spa}}, h_t^{\text{att}}, h_t^{\text{phy}})) \tag{5}$$

The architecture is capable of capturing latent correlations among environmental dynamics, user gaze preferences, and physiological states, thereby overcoming the limitations of conventional unimodal models in representing comprehensive user experience states. Consequently, high-dimensional feature representations are provided for subsequent adaptive visual feedback regulation.

In this study, a user-experience optimization objective is formulated within a reinforcement learning framework. A quantifiable real-time experience reward function is established to enable the dynamic iterative optimization of model parameters. The corresponding formulation is expressed as follows:

$$R_t = w_1 \cdot \Delta Engagement_t + w_2 \cdot Fluuency_t - w_3 \cdot TakCompletionTime_t \tag{6}$$

The weighting coefficients in the equation are used to balance three core evaluation criteria, namely user engagement, system smoothness, and interaction efficiency, thereby jointly constraining the optimization trajectory of the model. To accommodate the dynamic and highly variable interaction behavior features of mobile users, a hybrid training strategy combining offline expert imitation learning and online proximal policy optimization is adopted. Initial policy convergence is achieved through expert demonstrations representing optimal interaction strategies. Following deployment, the model is continuously adapted to the personalized interaction preferences of real-world users through periodic incremental updates. A dual-branch output architecture is designed. Continuous rendering control parameters, including hue offset, saturation scaling, and label transparency, are generated through fully connected layers. The Gumbel–Softmax mechanism is used to produce discrete interaction modes, enabling adaptive switching among four interaction strategies: exploration guidance, knowledge interpretation, emotional engagement, and efficiency-oriented navigation. Through this design, precise alignment between augmented reality visual presentation and users' real-time experiential states is achieved, thereby establishing a closed-loop dynamic visual feedback mechanism.

2.4 End-to-end joint optimization objective and multi-constraint modeling for mobile hardware platforms

Conventional landscape-oriented augmented reality optimization approaches have primarily focused on enhancing visual experience as a single optimization objective, while the hardware limitations of mobile devices have rarely been incorporated. Consequently, the resulting optimization strategies are often difficult to deploy effectively in real-world outdoor mobile environments. To address this limitation, an end-to-end joint optimization framework is established for mobile deployment, in which the parameters of both the scene perception network and the interaction regulation network are jointly optimized. The objective is to maximize the expected quality of experience over a temporal sequence. The global optimization objective is formulated as:

$$\max_{\theta_{gs},\theta_{policy}} \quad E_t[Q(f_t, a_t^{AR})] \tag{7}$$

where, θ_{gs} and θ_{policy} denote the learnable parameters of the Gaussian scene perception module and the multimodal regulation strategy module, respectively; and $Q(f_t, a_t^{AR})$ denotes the quantified experience quality at each time step, which is jointly determined by multimodal perceptual features and augmented reality visual-regulation actions. The temporal expectation calculation can smooth fluctuations in single-frame experience quality, thereby enabling the optimization of overall experience throughout long-duration landscape exploration.

To ensure the engineering feasibility of the optimization results and the operational stability of the proposed system, multiple hard constraints are incorporated by jointly considering the core requirements of real-time responsiveness, energy efficiency, and virtual–real alignment on mobile devices. Consequently, a comprehensive constrained optimization space is established as:

$$\text{s.t.} \quad \mathrm{T}_{\mathrm{render}} + \mathrm{T}_{\mathrm{comm}} + \mathrm{T}_{\mathrm{infer}} \le 50\mathrm{ms}, \quad \mathrm{P}_{\mathrm{device}} \le 3\mathrm{W}, \quad \left\| \mathrm{T}_{\mathrm{t}}^{\mathrm{SLAM}} \oplus \mathrm{T}_{\mathrm{t}}^{\mathrm{render}} \right\|_{\mathrm{F}} < \epsilon_{\mathrm{algn}} \tag{8}$$

The latency constraint integrates the end-to-end processing time associated with device-side rendering, edge-device communication, and model inference, thereby establishing a real-time operational threshold for mobile augmented reality interactions and preventing visual stuttering and interaction delays. The power-consumption constraint is designed according to the sustainable operating limits of contemporary mobile phones, restricting overall computational expenditure and mitigating performance degradation caused by thermal throttling. Furthermore, the pose-consistency constraint quantifies the discrepancy between the pose estimated by simultaneous localization and mapping and the rendering pose through the Frobenius norm of the matrix. By strictly constraining the coupling accuracy between spatial perception and virtual content rendering, virtual–real misalignment, rendering drift, and other shortcomings can be effectively eliminated. The modeling overcomes the traditional separation between experience optimization and hardware adaptation, enabling a globally optimal balance among quality of experience, real-time system responsiveness, and mobile-device operational stability. A schematic illustration of the dynamic augmented reality feedback framework driven by multimodal perception and Gaussian splatting is presented in Figure 2.

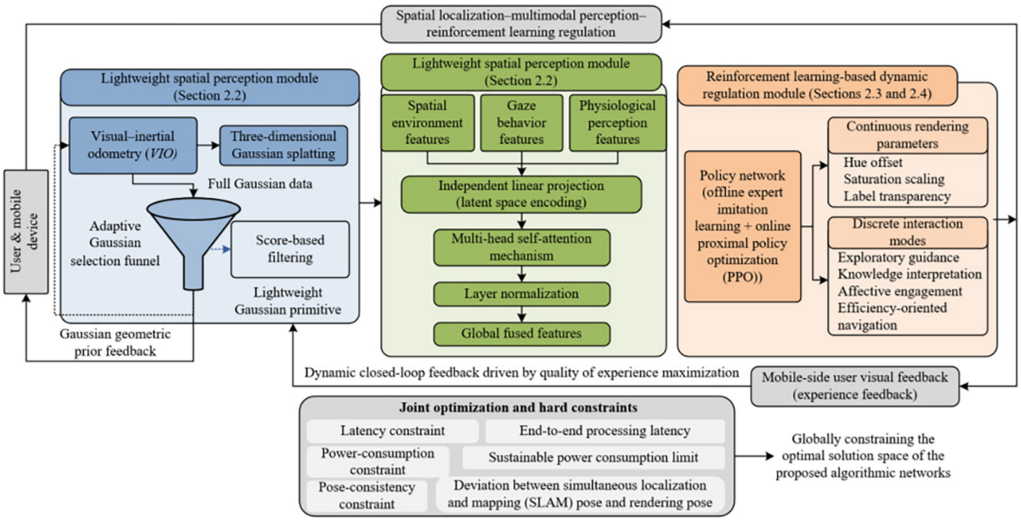

Fig. 2. Dynamic augmented reality feedback closed-loop framework driven by multimodal perception and Gaussian splatting

2.5 Edge–device collaborative lightweight engineering scheme for heterogeneous mobile terminals

To accommodate computational heterogeneity across mobile devices, dynamic network fluctuations in outdoor environments, and complex real-world interference, an integrated edge-device collaborative lightweight deployment framework is established. System-level engineering optimization is achieved through rendering scheduling, data encoding, hardware power management, and scene-adaptive correction. A dual-buffer, three-thread asynchronous rendering pipeline is designed to concurrently execute network data reception, Gaussian scene preprocessing, and rasterization rendering tasks. Rendering continuity is maintained through a buffer-switching mechanism. In addition, an inertial measurement unit-based constant-velocity extrapolation strategy is employed to compensate for pose information during periods of network instability, thereby effectively mitigating stalls and frame discontinuities caused by weak network conditions in outdoor environments.

To address the excessive transmission overhead associated with Gaussian scene data, a multi-level joint compression encoding strategy is adopted. Gaussian spatial coordinates are stored using a local-offset representation, while covariance matrices are subjected to Cholesky decomposition-based logarithmic-domain quantization. Combined with 16-bit integer quantization and Golomb coding, high-density data can be efficiently compressed. As a result, the storage requirement of a single Gaussian primitive can be reduced to 48 bytes, significantly decreasing communication bandwidth consumption between edge and mobile devices as well as memory utilization on mobile platforms. To balance rendering performance and energy efficiency, a dynamic power management mechanism is deployed. The operating frequency of the graphics processing unit is adaptively adjusted according to the number of visible Gaussian primitives in the current scene. Furthermore, when the device remains stationary for extended periods, the scene transmission frame rate is proactively reduced, enabling fine-grained energy-consumption control. As for challenging outdoor conditions, a real-time environmental illumination coefficient is introduced for dynamic brightness correction, thereby reducing the impact of sudden illumination variations on rendering quality. In addition, dynamically occluded regions are identified through scene-depth features, and the opacity of Gaussian primitives located behind occluding objects is adaptively reduced. Consequently, virtual–real misalignment caused by pedestrian-induced dynamic occlusions can be effectively mitigated, thereby comprehensively enhancing the environmental adaptability and operational robustness of the mobile landscape-oriented augmented reality system.

3 EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

3.1 Experimental environment and baseline schemes

For the experimental evaluation, an RTX 3090 graphics processing unit server was deployed at the edge layer to perform model inference and scene optimization tasks. Mobile testing devices included entry-level Android devices, flagship Android devices, and iOS devices, thereby covering the computational capabilities of mainstream mobile terminals across different performance tiers. The experimental site was selected within an open urban landscape park containing representative outdoor environments, including open grasslands, weak-texture paved surfaces, and dynamic pedestrian occlusions. Such settings enabled realistic reproduction of environmental disturbances commonly encountered during outdoor exploration. Three representative mobile augmented reality schemes were adopted as baseline configurations. The first baseline consisted of a conventional augmented reality system integrating visual–inertial localization with texture-map rendering. The second baseline employed a native three-dimensional Gaussian splatting approach with full-scene rendering. The third baseline was a landscape-oriented augmented reality system utilizing fixed parameters without dynamic affective feedback. Comparative evaluation was therefore conducted from three complementary perspectives, namely conventional methodologies, state-of-the-art rendering techniques, and specialized landscape augmented reality systems.

3.2 Quantitative evaluation of localization accuracy in outdoor environments

High-precision spatial coordinates acquired using light detection and ranging measurements were employed as ground-truth references for localization evaluation. Experiments were conducted under three representative outdoor conditions, including open environments, weak-texture paved environments, and dynamically occluded environments. The average trajectory root mean square error and maximum accumulated drift were recorded for each method to quantitatively assess the effectiveness of the proposed Gaussian prior constraints in improving localization stability on mobile devices. The localization performance of different approaches under various scenarios is summarized in Table 1.

As shown in Table 1, substantial improvements in localization accuracy were achieved by the proposed framework across all experimental scenarios. On average, localization drift was reduced by approximately 40% relative to the baseline approaches. The conventional visual–inertial odometry-based augmented reality system relied primarily on visual feature matching and therefore exhibited significant performance degradation in weak-texture environments, where insufficient visual features resulted in the largest localization drift errors. Although the native three-dimensional Gaussian splatting system provided scene reconstruction, no pose-correction mechanism was incorporated into the localization process, and accumulated odometry errors therefore remained uncorrected. By contrast, dense geometric priors derived from the Gaussian scene representation were incorporated into the proposed framework to establish relocalization constraints. Consequently, cumulative errors generated by visual–inertial odometry could be continuously corrected throughout the localization process. The most pronounced performance gains were observed under weak-texture conditions, where the maximum accumulated drift was reduced by 46.2%. These results demonstrate that the proposed approach effectively enhances the localization robustness of mobile augmented reality systems in challenging outdoor environments.

Table 1. Comparison of localization accuracy under different outdoor scenarios

Scenario Type	Evaluation Metric	Conventional Visual–Inertial Odometry-Based Augmented Reality Approach	Native Three-Dimensional Gaussian Splatting Approach	Proposed Approach	Overall Improvement
Open environment	Average root mean square error (m)	0.182	0.156	0.095	38.1%
	Maximum accumulated drift (m)	0.426	0.371	0.228	40.2%
Weak-texture environment	Average root mean square error (m)	0.357	0.293	0.164	44.0%
	Maximum accumulated drift (m)	0.783	0.652	0.351	46.2%
Dense-occlusion environment	Average root mean square error (m)	0.264	0.221	0.138	37.5%
	Maximum accumulated drift (m)	0.592	0.514	0.308	40.1%

3.3 Comparative evaluation of mobile rendering performance and bandwidth utilization

Device movement speed and observation scenarios were maintained consistently throughout the experiments. Continuous outdoor testing was conducted over a 30-minute period to evaluate the runtime performance of three strategies, namely full Gaussian rendering, fixed-threshold Gaussian pruning, and the proposed adaptive weighted selection strategy. Average rendering frame rate, per-frame transmission bandwidth, and peak memory consumption were selected as the primary evaluation metrics to assess the effectiveness of the lightweight selection and compression-encoding framework. Comparative results obtained under different rendering strategies are presented in Figure 3.

The experimental results indicate that the full Gaussian rendering approach was unable to satisfy real-time interactive requirements on mobile devices due to the excessive volume of scene data, resulting in extremely high bandwidth consumption and memory usage. Although the fixed-threshold pruning strategy reduced resource consumption to some extent, critical landscape details were frequently discarded because the pruning criteria remained static. By contrast, by using the proposed adaptive weighted selection mechanism, combined with the multi-level joint compression encoding framework, no perceptible degradation in rendering quality was observed, and per-frame transmission bandwidth was reduced to 37.5% of that required by the full Gaussian rendering strategy. Simultaneously, rendering performance on mobile devices was consistently maintained above 60 frames per second, while peak memory consumption was significantly reduced. These results demonstrate the lightweight effectiveness of the multidimensional weighted selection strategy, which jointly considers color, texture, and proximity, as well as the hybrid encoding scheme. The strategy can effectively satisfy the computational and bandwidth constraints associated with real-time mobile rendering.

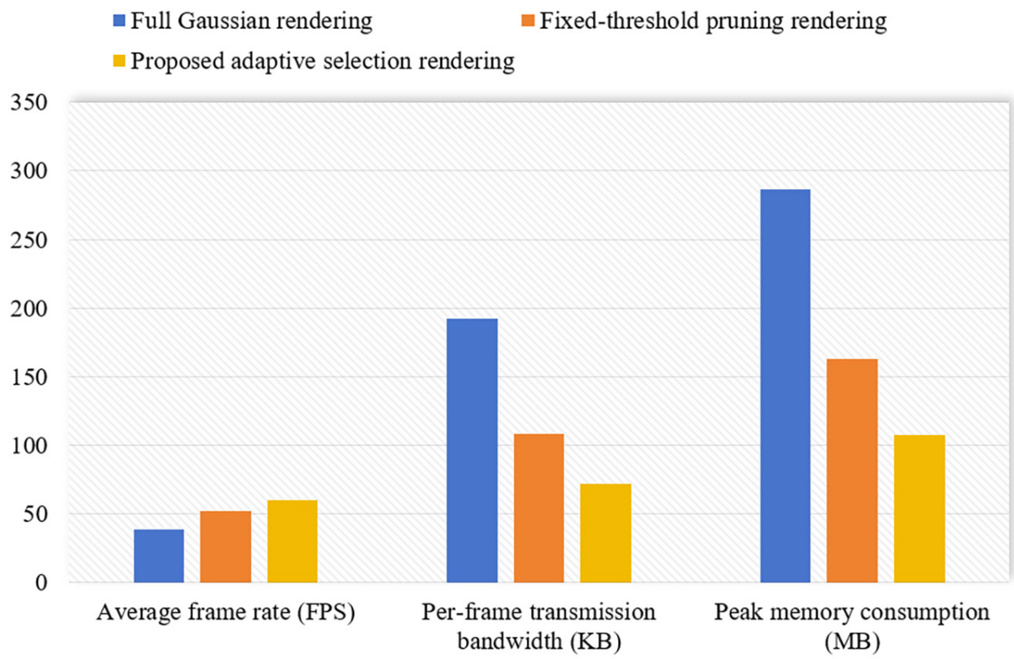

Fig. 3. Comparison of mobile rendering performance and transmission resource consumption

3.4 Ablation study of the multimodal affective prediction model

To evaluate the effectiveness of the multimodal fusion and training strategy, five ablation configurations were designed. Specifically, the physiological modality, the gaze-interaction modality, and the cross-modal attention mechanism were removed individually. In addition, the online proximal policy optimization-based incremental optimization module was disabled. The performance of each ablated configuration was subsequently compared with that of the complete model. Augmented reality rendering parameter prediction error and user interaction completion time were selected as the primary evaluation metrics to quantitatively assess the contribution of each core component to overall system performance.

As illustrated by the ablation results presented in Figure 4, performance degradation was consistently observed whenever any component of the proposed framework was removed, indicating that each module contributes indispensably to the overall system performance. The largest increase in error was observed when the cross-modal attention mechanism was removed, demonstrating that deep modeling of inter-modal dependencies constitutes the primary factor responsible for improving affective prediction accuracy. The removal of either the gaze-interaction modality or the physiological modality resulted in a reduced ability to accurately capture users' real-time experience states, thereby introducing deviations in the regulation of rendering parameters. Furthermore, the online proximal policy optimization-based incremental optimization module was shown to effectively adapt the evolving user interaction behaviors, substantially reducing prediction errors and interaction completion time under complex conditions. By integrating multimodal feature fusion with the hybrid training paradigm, the full model achieved the highest prediction accuracy and interaction efficiency, enabling precise and dynamic adaptive regulation of augmented reality visual parameters.

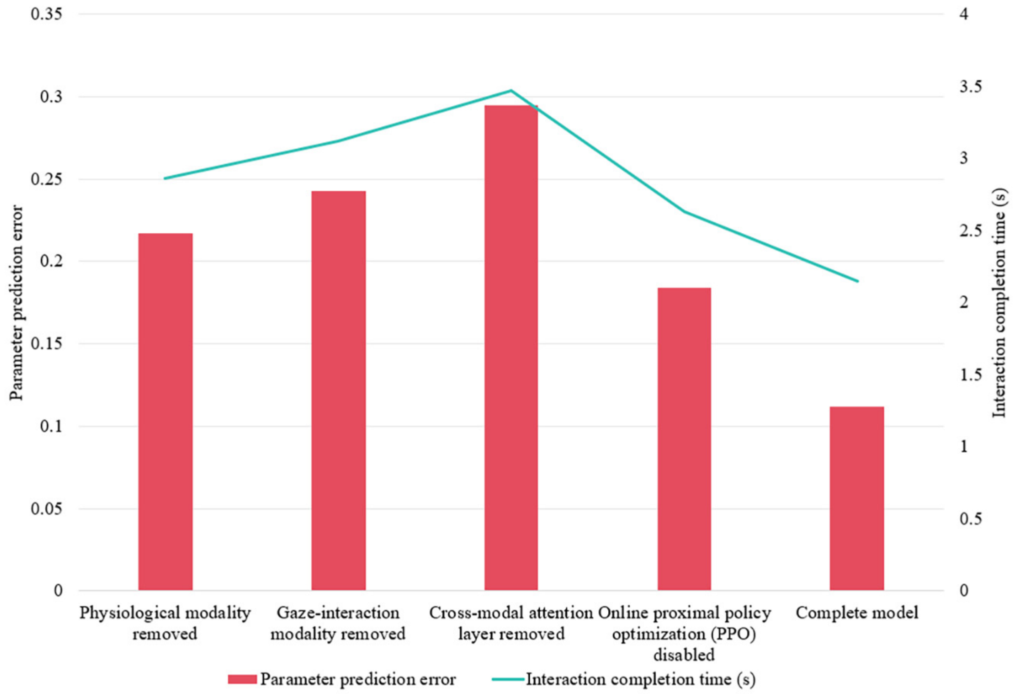

Fig. 4. Results of the ablation study on the multimodal model

3.5 Validation of mobile latency and device power-consumption constraints

Continuous experiments were conducted under three representative outdoor network conditions, including high-speed 5G networks, conventional 4G networks, and fluctuating weak-network environments. The proposed method and comparison approaches were deployed on mobile devices with varying hardware capabilities. End-to-end average latency, average device power consumption, and the proportion of samples exceeding the predefined power consumption threshold were recorded for each configuration. These metrics were used to assess the engineering feasibility of the edge–device collaborative architecture and dynamic power regulation mechanism under real-world deployment conditions.

The overall results presented in Table 2 indicate that both the conventional augmented reality system and the native three-dimensional Gaussian splatting system exhibited latency violations under challenging network conditions. In addition, average device power consumption remained above the 3 W design threshold for prolonged periods, thereby increasing the likelihood of thermal throttling, device overheating, and performance degradation. By contrast, the proposed edge–device collaborative lightweight deployment framework maintained end-to-end latency below 48 ms across all experimental conditions, thereby satisfying the 50 ms real-time constraint. Through the combined use of dynamic graphics processing unit frequency scaling and stationary-state frame-rate reduction, average device power consumption was reduced by approximately 28% relative to the baseline approaches. Furthermore, no power-limit violations were observed throughout the testing process, indicating that stable operation could be maintained across heterogeneous mobile platforms. In addition, the asynchronous rendering pipeline and the inertial measurement unit-based frame interpolation mechanism effectively compensated for communication instability under weak-network conditions, thereby substantially enhancing system stability and adaptability under dynamic outdoor network conditions.

Table 2. Comparison of end-to-end latency and device power consumption

Network Condition	Experimental Method	Average End-to-End Latency (ms)	Average Device Power Consumption (W)	Proportion of Power-Limit Violations (%)
5G network	Conventional augmented reality method	42.3	3.86	28.4
	Native three-dimensional Gaussian splatting method	47.5	3.52	16.7
	Proposed method	36.8	2.51	0.0
4G network	Conventional augmented reality method	49.6	3.82	26.1
	Native three-dimensional Gaussian splatting method	53.2	3.48	14.2
	Proposed method	41.2	2.47	0.0
Weak-network conditions	Conventional augmented reality method	58.4	3.79	25.8
	Native three-dimensional Gaussian splatting method	62.7	3.45	13.5
	Proposed method	47.6	2.43	0.0

4 CONCLUSIONS

To address the critical challenges encountered by mobile augmented reality systems in outdoor landscape environments, including insufficient localization robustness, excessive rendering overhead, rigid interaction paradigms, and limited hardware adaptability, a cloud–edge–device collaborative framework for landscape-oriented augmented reality experience optimization was established. Several methodological innovations were achieved at the theoretical level. A perception framework integrating visual–inertial odometry and three-dimensional Gaussian splatting was developed, in which cumulative localization drift in complex outdoor environments was effectively suppressed through the incorporation of an adaptive Gaussian selection mechanism and scene-prior-based relocalization constraints. In addition, a cross-modal attention fusion network and a hybrid training paradigm were introduced to enable adaptive regulation of augmented reality rendering parameters. Furthermore, a multi-constraint joint optimization framework simultaneously considering quality of experience, communication latency, and device power consumption was established, thereby providing an end-to-end quantitative modeling framework for landscape-oriented mobile augmented reality systems. Comprehensive outdoor experiments demonstrated substantial improvements in outdoor localization accuracy. Average localization drift was reduced by approximately 40%, while the hardware constraints of 50 ms end-to-end latency and 3 W device power consumption were consistently satisfied. Consequently, the long-standing trade-off between user-experience optimization and hardware-resource limitations in conventional augmented reality systems was effectively mitigated.

5 REFERENCES

[1] F. Luo, "Application of mobile augmented reality technology in urban tourism and optimization of visitor experience," *International Journal of Interactive Mobile Technologies*, vol. 20, no. 11, pp. 138–151, 2026. https://doi.org/10.3991/ijim.v20i11.62124

[2] P. Wongklang and J. Wipatsopakron, "The development of problem-based mobile augmented reality application to enhance creative problem-solving skills for undergraduate students," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 7, pp. 53–67, 2024. https://doi.org/10.3991/ijim.v18i07.42261

[3] M. X. Liu, H. F. Hu, and J. Ran, "Optimization of service function chain migration based on graph neural networks and deep reinforcement learning," *Acadlore Transactions on AI and Machine Learning*, vol. 5, no. 2, pp. 153–166, 2026. https://doi.org/10.56578/ataiml050205

[4] M. Sužnjević, S. Srebot, M. Moslavac, K. Mišura, L. Boban, and A. Jović, "Longitudinal usability and UX analysis of a multiplatform house design pipeline: Insights from extended use across web, VR, and mobile AR," *Applied Sciences*, vol. 15, no. 19, p. 10765, 2025. https://doi.org/10.3390/app151910765

[5] K. Girginova *et al.*, "Augmented landscapes of empathy: Community voices in augmented reality campaigns," *Media and Communication*, vol. 12, 2024. https://doi.org/10.17645/mac.8581

[6] Q. Quach and B. Jenny, "Immersive visualization with bar graphics," *Cartography and Geographic Information Science*, vol. 47, no. 6, pp. 471–480, 2020. https://doi.org/10.1080/15230406.2020.1771771

[7] Z. Zhong, G. Chen, R. Wang, and Y. Huo, "Neural super-resolution in real-time rendering using auxiliary feature enhancement," *Journal of Database Management*, vol. 34, no. 3, pp. 1–13, 2023. https://doi.org/10.4018/jdm.321544

[8] Y. Chen *et al.*, "Distributed real-time rendering for ultrahigh-resolution multiscreen 3D display," *Journal of the Society for Information Display*, vol. 30, no. 3, pp. 244–255, 2022. https://doi.org/10.1002/jsid.1097

[9] J. Ahn, J. Lee, S. Yoon, and J. K. Choi, "A novel resolution and power control scheme for energy-efficient mobile augmented reality applications in mobile edge computing," *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 750–754, 2019. https://doi.org/10.1109/lwc.2019.2950250

[10] N. Li, L. Zhai, S. Song, X. Zhu, Y. Li, and F. Yang, "Cache allocation policy based on user preference using reinforcement learning in mobile edge computing," *Concurrency and Computation: Practice and Experience*, vol. 36, no. 15, 2023. https://doi.org/10.1002/cpe.7991

[11] J. Liu, Y. Xie, S. Gu, and X. Chen, "A SLAM-based mobile augmented reality tracking registration algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 1, 2019. https://doi.org/10.1142/s0218001420540051

[12] Y. Seo, J. Lee, J. Hwang, D. Niyato, H. Park, and J. K. Choi, "A novel joint mobile cache and power management scheme for energy-efficient mobile augmented reality service in mobile edge computing," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1061–1065, 2021. https://doi.org/10.1109/lwc.2021.3057114

[13] Z. Kunkera, I. Željković, R. Mimica, B. Ljubenkov, and T. Opetuk, "Development of augmented reality technology implementation in a shipbuilding project realization process," *Journal of Marine Science and Engineering*, vol. 12, no. 4, p. 550, 2024. https://doi.org/10.3390/jmse12040550

[14] H. Ahmad, A. Butt, and A. Muzaffar, "Travel before you actually travel with augmented reality – role of augmented reality in future destination," *Current Issues in Tourism*, vol. 26, no. 17, pp. 2845–2862, 2022. https://doi.org/10.1080/13683500.2022.2101436

[15] J. Izquierdo-Domenech, J. Linares-Pellicer, and I. Ferri-Molla, "Environment awareness, multimodal interaction, and intelligent assistance in industrial augmented reality solutions with deep learning," *Multimedia Tools and Applications*, vol. 83, no. 16, pp. 49567–49594, 2023. https://doi.org/10.1007/s11042-023-17516-x

6 AUTHOR

Yunzhi Tian graduated from Xi'an University of Architecture and Technology. She is currently an Associate Professor at the School of Architectural Engineering, Huanghuai University, Zhumadian, China. Her research interests include urban and rural planning and design, architectural and landscape design, urban and rural ecological environments, and the history of urban and rural development. She has presided over and completed 3 provincial-level research projects, published more than 10 academic papers, and authored 1 monograph and 2 textbooks (E-mail: tianyz1983@163.com).