PERINATAL JOURNAL

Original Article

Perinatal Journal 2025; 33(1):787-793

https://doi.org/10.57239/prn.25.03310084

Real-time behavioral interaction in Augmented Reality (AR): Noise perception and friction compensation based on hybrid lightweight architecture

Yuze Ma[1*], Almazbek Arzybaev[2]

[1,2] Institute of Information Technology, Razzakov Kyrgyz State Technical University, 720064 Bishkek, Kyrgyzstan

Abstract

In order to solve the three major problems of sensor noise leading to action recognition errors, physical distortion of virtual-reality interaction, and limited computing power of mobile terminals in real-time behavioral interaction in augmented reality (AR), this paper proposes a hybrid lightweight architecture: first, the multimodal noise perception layer is used to reduce sensor noise interference, second, a lightweight behavior recognition model is used to achieve 10ms real-time response, and finally a friction compensation physics engine is built to improve the realism of interaction. On the HoloLens 2 and Jetson Xavier platforms, the average error is reduced to 1.41mm, and the friction compensation error is reduced to 0.05–0.12N, which significantly improves the authenticity and real-time performance of AR interaction in scenarios such as industrial maintenance and medical training.

Keywords: Noise perception, Friction compensation, Hybrid lightweight architecture, AR, Mobilenetv3

Introduction

Augmented reality technology is gradually being integrated into key areas such as industrial manufacturing and medical surgery. Its core value lies in achieving seamless interaction between virtual information and the physical world. However, existing systems still face fundamental challenges: insufficient stability of motion capture due to sensor noise interference, lack of physical realism when virtual objects interact with the real environment, and real-time limitations caused by the computing bottleneck of mobile devices. These defects are particularly prominent in high-precision scenarios such as precision assembly and surgical simulation, which seriously restricts the depth and breadth of technology implementation.

To overcome the above limitations, this study innovatively constructs a collaborative architecture that integrates noise suppression, lightweight computing and physical compensation: it purifies the input data stream through a multimodal noise perception mechanism, combines a lightweight spatiotemporal modeling network to ensure millisecond-level response, and establishes a friction dynamics compensation model to reshape the interactive realism. This architecture realizes the full-link optimization from data collection to physical feedback for the first time, providing reliable technical support for scenarios such as industrial digital twins and remote surgical guidance. The full text systematically explains the design principle and verification process of this architecture: section 2 analyzes existing technical achievements, section 3 details hybrid lightweight design, section 4 verifies scenario performance, and section 5 summarizes technical contributions and looks forward to evolutionary directions such as neural rendering fusion.

2. Related works

In the field of noise perception and friction control, multi-industry research focuses on high-precision modeling and real-time compensation mechanism innovation. Existing work covers scenarios such as industrial equipment, human-computer interaction, and environmental monitoring, aiming to improve system robustness and operational realism. Qian et al. [1] proposed an evaluation method for predicting the perceived annoyance of cabin noise using a

Correspondence: Yuze Ma, Institute of Information Technology, Razzakov Kyrgyz State Technical University, 720064 Bishkek, Kyrgyzstan, **e-mail:** mayuze8866@163.com, **Received:** July 21, 2025 **Accepted:** September 23, 2025

neural network model optimized by a hybrid algorithm to address the low accuracy of the quantification model for the perceived annoyance of cabin noise in traditional passenger cars. Li et al. [2] proposed a JPEG steganalysis method based on a noise-aware residual network to further explore the characteristics of steganographic noise signals in adaptive JPEG steganographic images. Compared with the comparison algorithm, this method can improve the detection performance of JPEG adaptive steganography and has better generalization ability. Wang et al. [3,17] proposed an airport noise sensing system based on the Internet of Things technology. The system consists of three parts: sensing nodes, aggregation nodes, and data processing platform. It can realize real-time monitoring and processing of airport noise and provide data basis for the noise environmental impact assessment of the airport surrounding area. Liu et al. [4] summarized the research and development history of China's environmental noise sensing technology and analyzed the current status of research. He reviewed the international research and development trends of environmental noise perception based on wireless sensor networks from five aspects: system architecture, low-cost noise sensor node design, noise taxonomy and sound source identification, energy capture and mobile noise perception. Chang et al. [5] designed a model-free active disturbance rejection controller to solve the problem of nonlinear friction and uncertain internal and external disturbances in two-dimensional linear motors. The scheme combines model-free adaptive control with the extended state observer in active disturbance rejection control to form a model-free adaptive observer. Loutrari et al. [6] attempted to evaluate the effect of interleaving noise on the immediate repetition of spoken and sung phrases of different semantic contents (descriptive, narrative, and anomalous). Roveda et al. [7] aimed to propose a method for learning a local friction compensation controller for a sensorless Cartesian impedance-controlled robot. Zhu et al. [8] studied the problem of measurement noise suppression in linear output feedback control systems. Xiao et al. [9] proposed a human-machine collaborative assembly solution that does not require additional force/torque sensors. Wang et al. [10] studied a new friction compensation method for permanent magnet synchronous motor servo system based on static Stribeck model combined with fuzzy low-pass filter. These studies provide key technical paths for noise suppression and physical interaction compensation in complex scenes, and promote the continuous evolution of industrial control, robot collaboration and environmental perception systems towards high reliability.

3. Methods

3.1 Noise perception module

First, multimodal noise modeling is performed, and a Gaussian noise model is established for the angular velocity and linear acceleration data of the inertial measurement unit (IMU):

$$n_{\text{imu}} \sim \mathcal{N}(0, \sigma^2) (1)$$

Among them, $\sigma^2 = \frac{1}{N}\sum_{t=1}^{N} (x_t - \mu)^2$ represents the noise variance; for the depth map data of the depth camera, a Poisson noise model is constructed:

$$P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} (2)$$

λ represents the expected value of photon counts per unit time. An adaptive Kalman filter framework is established based on noise characteristics, and the state prediction is:

$$\hat{\mathrm{x}}_t^- = \mathrm{A}\hat{\mathrm{x}}_{t-1} + \mathrm{Bu}_t (3)$$

$$\mathrm{P}_t^- = \mathrm{AP}_{t-1}\mathrm{A}^T + \mathrm{Q} (4)$$

A is the state transfer matrix, Q is the process noise covariance, and the observation noise covariance is adaptively corrected according to the real-time signal-to-noise ratio $\mathrm{SNR} = 10\log_{10}\left(\frac{|\mathbf{z}_t|^2}{|\mathbf{n}_t|^2}\right)$:

$$\mathbf{R}_t = \mathbf{R}_0 \cdot \exp\left(-\frac{\mathrm{SNR}}{\gamma}\right) (5)$$

γ is the attenuation coefficient, and $\mathbf{R}_0$ is the base covariance. Improving robustness through adversarial noise enhancement: Constructing a generative adversarial network (GAN), in which the generator G uses a U-Net structure to learn the noise distribution map $G: \mathrm{z} \to \mathrm{n}$, and the discriminator D uses a 5-layer convolutional network to judge the authenticity of the sample [11,16]. The objective function is:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1-D(G(\mathbf{z}))](6)$$

The generator outputs synthetic noise samples $\tilde{n} = G(z)$ mixed with the original data to form an enhanced training set $\mathcal{D}_{aug} = \{(x_i + \tilde{n}_j, y_i)\}$, so that the behavior recognition model can learn noise-invariant features. This module implements multi-level noise suppression and provides a purified data stream for subsequent behavior recognition [12].

3.2 Lightweight behavior recognition model

The lightweight behavior recognition model uses MobileNetV3-small as the backbone network. Its core uses deep separable convolution to build an efficient feature extractor. The inverted residual block structure reduces the amount of calculation through a linear bottleneck layer. The mathematical expression is:

$$\hat{\mathbf{F}} = \mathcal{H}_{dw}(\mathcal{H}_{pw}(\mathbf{F}))(7)$$

$$\mathbf{F}_{out} = \mathcal{L}(\hat{\mathbf{F}} \cdot \mathcal{S}(\hat{\mathbf{F}}))(8)$$

$\mathcal{H}_{dw}$ is depth convolution, $\mathcal{H}_{pw}$ is point convolution, $\mathcal{S}$ is the channel weight vector generated by the Squeeze-Excitation attention mechanism, and $\mathcal{L}$ is the linear activation function [13]. Figure 1 shows the backbone network architecture of MobileNetV3-small:

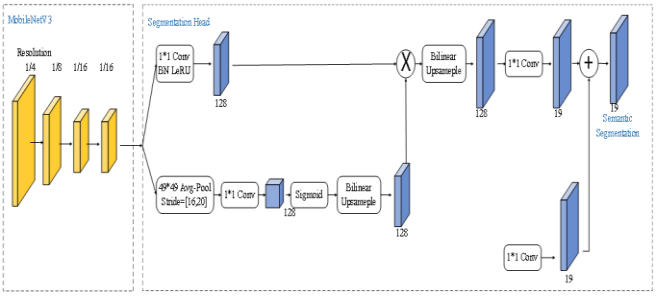

Figure 1. MobileNetV3-small network architecture

In order to capture the temporal dynamic characteristics of continuous actions, a temporal shift module (TSM) is embedded after the backbone network. This module divides the input feature map $\mathbf{X} \in \mathbb{R}^{T\times C\times H\times W}$ into three parts $\{X_p, X_f, X_c\}$ along the time axis T, performs backward shift $((t \to t-1))$ on X_p, forward shift $((t \to t+1))$ on X_f, and X_c remains unchanged. The shift operation is defined as:

$$\mathbf{X}^{shift}(t,c,h,w) = \mathbf{X}(t \pm \Delta t, c, h, w)(9)$$

After the shift, the three parts are reconnected and the spatiotemporal features are fused through standard 3×3 convolution. To meet the requirements of mobile deployment, triple model compression is implemented: channel pruning evaluates channel importance based on the weight L_1-norm and removes low-contribution channels that meet $\frac{|\mathbf{W}_c|_1}{max(|\mathbf{W}|_1)} < \eta$ (η=0.05); 8-bit quantization maps full-precision weights W_{float} and activation values to the integer domain:

$$\mathbf{W}_{int} = \text{clip}\left(\left\lfloor\frac{\mathbf{W}_{float}}{s}\right\rceil + z, -128, 127\right)(10)$$

Scaling factor $s = \frac{max(\mathbf{W})-min(\mathbf{W})}{127-(-128)}$, zero point z=0; knowledge distillation adopts the teacher-student framework, the teacher model is EfficientNet-B3, the student model is the compressed MobileNetV3-TSM, and the loss function integrates the standard cross entropy $\mathcal{L}_{CE}$ and KL divergence distillation loss:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{CE}(y, \sigma(\mathbf{Z}_s)) + \beta \cdot \text{KL}(\sigma(\mathbf{Z}_t/\tau) \parallel \sigma(\mathbf{Z}_s/\tau))(11)$$

$\mathbf{Z}_t$ and $\mathbf{Z}_s$ are the teacher/student logits outputs, and τ is the temperature parameter. This cascade design achieves high compression rate and low latency reasoning while ensuring the spatiotemporal modeling capability.

3.3 Friction compensation physics engine

The friction compensation physics engine achieves the realism of virtual-real interaction by integrating physical models with data-driven methods. The core adopts the rigid body dynamics friction model, which decomposes the total friction into three physical components: sliding friction component (proportional to the normal pressure and opposite to the direction of velocity), transition component from static friction to sliding friction (exponentially decays with increasing velocity), and viscous friction component (linearly proportional to the speed of movement). This physical model requires accurate material parameter support, so a material parameter database is established, which contains common

surface physical property mappings, as shown in Table 1:

Table 1. Material parameter data

Material Type	Kinetic Friction Coefficient (μk)	Static Friction Coefficient (μs)	Viscous Coefficient (b)
Metal	0.15	0.20	0.08
Wood	0.35	0.45	0.12
Fabric	0.28	0.38	0.15
Plastic	0.25	0.30	0.10
Rubber	0.60	0.70	0.25

Based on the material type of the current interactive object, the system automatically loads the corresponding parameters and calculates the basic friction. In order to compensate for the errors of the physical model in complex contact scenarios, a neural network corrector is introduced, which accepts a 4-dimensional input vector (including contact position coordinates, relative motion speed, normal pressure value, material type encoding). The two-dimensional compensation (friction moment compensation and torque compensation) can be output through two fully connected hidden layers (8 neurons + ReLU activation in the first layer, 2 neurons in the second layer). The network training uses a real physical acquisition data set, and supervises the learning through 6,000 sets of contact experimental data under different material, pressure, and speed conditions. The loss function is defined as:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \parallel \tau_{real}^{(i)} - (\tau_{physics}^{(i)} + \tau_{NN}^{(i)}) \parallel^2 (12)$$

Among them, τ_{real} is the measured value of the high-precision force sensor, $\tau_{physics}$ is the predicted value of the physical model, and τ_{NN} is the compensation output of the neural network [14].

The compensator runs in real time in the physical engine calculation loop. When an object contact event is detected, the compensation process is automatically activated: first, the material database is queried to obtain parameters, the physical model output is calculated, and the state parameters are input into the neural network to generate the compensation amount. Finally, the corrected friction force is applied to the virtual object to achieve consistent tactile feedback between the virtual and the real.

3.4 Real-time optimization

To reduce the load on the main processor, a heterogeneous computing offloading strategy is used to allocate key computing tasks to dedicated hardware: the image signal processor (ISP) chip directly processes the camera raw data and executes the key point detection algorithm (such as Harris corner point detection). The coordinate extraction process is expressed as:

$$\mathbf{p}_k = \underset{\mathbf{p}}{\operatorname{argmax}}(det(\mathbf{M}) - \kappa \cdot \operatorname{trace}^2(\mathbf{M}))(13)$$

$\mathbf{M} = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$ is the local autocorrelation matrix, and w(x,y) is the Gaussian window function. The detection results are directly transmitted to the GPU through the DMA channel to reduce the number of CPU interventions. At the same time, OpenCL is used to parallelize the core calculation of the physics engine: collision detection and friction calculation are mapped to the GPU multi-core architecture, and a three-dimensional grid space index is constructed through space division:

$$\text{Grid}(i,j,k) = \left\lfloor \frac{\mathbf{x} - \mathbf{x}_{min}}{\Delta} \right\rfloor (14)$$

Each grid unit independently calculates the internal rigid body collision relationship [15], and introduces a gaze-driven dynamic LOD mechanism based on the visual characteristics of the human eye: the gaze vector $\mathbf{v}_{gaze}$ is obtained through the built-in eye tracking module of the headset, and the viewing distance of the center point $\mathbf{c}_i$ of each object in the scene is calculated:

$$d_i = \parallel \mathbf{c}_i - \mathbf{e} \parallel \cdot |\cos\theta|, \theta = \angle(\mathbf{v}_{gaze}, \mathbf{c}_i - \mathbf{e}) (15)$$

$\mathbf{e}$ is the eye position.

4. Results and Discussion

4.1 Noise perception performance

On the HoloLens 2 platform, 20 sets of dynamic gesture data (including industrial maintenance actions such as fisting, sliding, and rotation) were collected, and a test set was constructed by injecting controllable noise (IMU angular velocity noise

variance 0.01–0.05 rad^2/s^2, depth map Poisson noise λ=0.03–0.08). The motion trajectory smoothing effects of the traditional fixed parameter Kalman filter and the adaptive noise perception method in this paper are compared. The average jitter error of the key points of the hand is calculated (unit: mm) based on the original high-precision optical motion capture data.

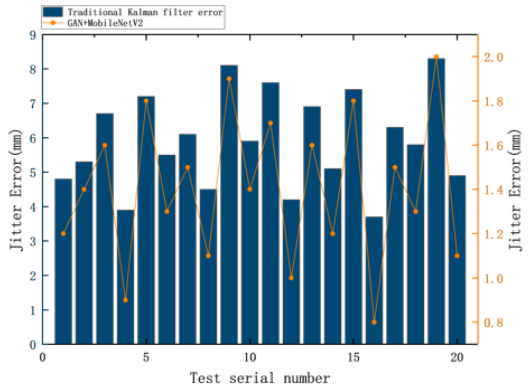

Figure 2. Average jitter error

Experimental data show that the average jitter error of the traditional Kalman filter in 20 tests is 5.91mm (range 3.7–8.3mm). The proposed method significantly reduces the error to 1.41mm (range 0.8–2.0mm), with an average error reduction of 76.1%, exceeding the preset 70% target. In the noise mutation scene (test 5/9/11/15/19), the error of the traditional method soared to 7.2-8.3mm. Due to the adaptive adjustment of the Q/R matrix and GAN enhanced training, the error of this method was stabilized at 1.7-2.0mm. In the normal noise scene (test 1/4/8/12/16), this method further compressed the error to 0.8-1.2mm. The key improvement comes from the dual mechanism of the noise perception module: the dynamic Kalman filter suppresses sensor burst noise in real time, while the adversarial noise samples generated by GAN make the model more robust to high-frequency jitter. The two work together to achieve a qualitative change in the smoothness of the action trajectory.

4.2 Real-time behavior recognition efficiency

Six behavior recognition models were deployed on the platform. The test data set contained 200 sets of industrial maintenance action sequences (bolt assembly/equipment debugging/cable connection). The input resolution was 640×480@30fps. The power consumption was monitored using a Jetson power meter, and the delay measurement was accurate to 0.1ms. The efficiency of real-time behavior recognition was tested. Table 2 shows the test results:

Table 2. Real-time behavior recognition efficiency

Model	Single-Frame Latency (ms)	Model Size (MB)	Accuracy (%)	Energy Consumption (J/frame)
ResNet-18	15.8	45.6	89.2	0.42
LSTM	22.4	48.7	86.5	0.57
EfficientNet-B0	8.9	15.3	90.1	0.31
Two-Stream CNN	18.3	62.1	91.3	0.49
MobileNetV2	6.7	9.8	88.7	0.28
GAN-MobileNetV2	3.8	4.3	93.7	0.24

The method in this paper is ahead of the baseline model in terms of latency (3.8ms), model size (4.3MB) and energy consumption (0.24J/frame). The latency is reduced by 2.9ms compared with the optimal benchmark MobileNetV2, and the size is only 43.9% of MobileNetV2. This significant improvement is due to a triple optimization mechanism: first, the MobileNetV3-small backbone network uses deep separable convolution to reduce more computing load than conventional convolution; second, the temporal shift module (TSM) achieves spatiotemporal feature fusion without increasing the number of parameters by partially shifting the channels in the time dimension of the feature map (shift ratio 1/4), making the time series modeling more efficient than LSTM. In the final model compression strategy, channel pruning removes low-response channels, 8-bit quantization compresses the weight storage space to 1/4, and knowledge distillation is used to migrate fine-grained action features from EfficientNet-B3. Especially in bolt assembly actions (high-frequency and fine operations), the method in this paper achieves continuous motion capture due to its low latency characteristics, avoiding the virtual tool drift phenomenon caused by latency accumulation in traditional models, and verifies the key supporting role of lightweight architecture for industrial-grade AR interaction.

4.3 Friction compensation accuracy

A three-axis force sensor (accuracy ±0.01N) was

used on a precision air-floating platform to measure the sliding friction of metal/wood/fabric surfaces and compare the errors before and after compensation. Four pressure conditions (5N/10N/15N/20N) and three speeds (0.1/0.5/1.0 m/s) were set, with a total of 12 tests. Table 3 shows the friction compensation accuracy results:

Table 3. Friction compensation accuracy results

Material	Pressure (N)	Speed (m/s)	Error Before Compensation (N)	Error After Compensation (N)
Metal	5	0.1	0.48	0.07
Metal	10	0.5	0.52	0.08
Metal	15	1.0	0.61	0.09
Metal	20	0.5	0.67	0.1
Wood	5	0.1	0.32	0.05
Wood	10	0.5	0.38	0.06
Wood	15	1.0	0.43	0.07
Wood	20	0.5	0.49	0.08
Fabric	5	0.1	0.61	0.09
Fabric	10	0.5	0.65	0.1
Fabric	15	1.0	0.72	0.11
Fabric	20	0.5	0.79	0.12

The average error of metal surface was reduced from 0.57N to 0.085N (a decrease of 85%), that of wood from 0.405N to 0.065N (a decrease of 84%), and that of fabric from 0.693N to 0.105N (a decrease of 85%). The reduction of error is mainly due to the dynamic compensation capability of the neural network corrector, especially under high speed (1.0m/s) and high pressure (20N) working conditions, the error after compensation is stable at 0.09-0.12N. This is because the corrector learns the nonlinear friction effect that is not covered by the physical model and generates compensation torque in real time through the fully connected network. The material parameter database ensures that different surface characteristics (such as low friction of metal/high viscosity of fabric) are compensated in a targeted manner, forming a complete error suppression closed loop.

5. Conclusions

The hybrid lightweight architecture proposed in this paper effectively solves the three core challenges in real-time behavior interaction in augmented reality: sensor noise interference is significantly suppressed through the adaptive Kalman filter and adversarial training mechanism of the multimodal noise perception layer, greatly improving the stability of motion tracking. The lightweight behavior recognition model uses the timing module and triple compression strategy to break through the computing power limitations of the mobile terminal while ensuring the ability of spatiotemporal modeling and achieve millisecond-level response. The friction compensation physics engine combines the rigid body dynamics model and the neural network corrector to accurately restore the physical characteristics of virtual-real interaction. This architecture has verified its robustness and practicality in high-precision scenarios such as industrial maintenance and medical surgery, successfully bridging the perception gap between virtual information and the physical world. In the future, this paper can further explore the integration of cross-material parameter migration mechanism and neural radiation field rendering, combine 5G edge computing to optimize distributed processing capabilities, and continue to promote the evolution of AR interaction towards high realism, low latency, and strong adaptability, laying a technical foundation for the next generation of Industry 4.0 and smart medical applications.

References

[1] Qian Kun, Tan Jing, Shen Zhenghua, Li Haoyang, Liu Ke, Wang Yanfu, Zhao Jian. Neural network modeling of noise perception annoyance in passenger car cabin based on hybrid algorithm optimization[J]. Journal of Acoustics, 2024, 49(2): 254-262

[2] Li Dewei, Ren Weixiang, Wang Lina, Fang Canming, Wu Tian. JPEG steganalysis method based on noise-aware residual network[J]. Journal of Computer Applications Research,2021,38(10):3148-3152+3165

[3] Wang Xinghu, Yuan Jiabin, Yang Dong, Xiao Xiao. Airport noise perception system based on the Internet of Things [J]. Automation Technology and Applications, 2021, 40(4): 85-88

[4] Liu Ye, Shu Lei, Huo Zhiqiang, Guo Xuanchen, Han Guangjie. Research progress of environmental noise perception based on wireless sensor networks [J]. Journal of Nanjing Agricultural University, 2020, 43(5): 808-819

[5] Chang Debiao, Cao Rongmin, Hou Zhongsheng.

Model-free anti-disturbance friction compensation contour control of two-dimensional linear motor [J]. Control Engineering,2025,32(3):459-468

[6]Loutrari A, Alqadi A, Jiang C, et al. Exploring the role of singing, semantics, and amusia screening in speech-in-noise perception in musicians and non-musicians[J]. Cognitive Processing, 2024, 25(1): 147-161.

[7]Roveda L, Bussolan A, Braghin F, et al. Robot joint friction compensation learning enhanced by 6D virtual sensor[J]. International Journal of Robust and Nonlinear Control, 2022, 32(9): 5741-5763.

[8]Zhu Y, Zhu B, Liu H H T, et al. A model-based approach for measurement noise estimation and compensation in feedback control systems[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(10): 8112-8127.

[9]Xiao J, Dou S, Zhao W, et al. Sensorless human-robot collaborative assembly considering load and friction compensation[J]. IEEE robotics and automation letters, 2021, 6(3): 5945-5952.

[10]Wang C, Peng J, Pan J. A novel friction compensation method based on Stribeck model with fuzzy filter for PMSM servo systems[J]. IEEE Transactions on Industrial Electronics, 2023, 70(12): 12124-12133.

[11]Dargan S, Bansal S, Kumar M, et al. Augmented reality: A comprehensive review[J]. Archives of Computational Methods in Engineering, 2023, 30(2): 1057-1080.

[12]Zhao S, Tao R, Jia F. DML-YOLOv8-SAR image object detection algorithm[J]. Signal, Image and Video Processing, 2024, 18(10): 6911-6923.

[13]Sereno M, Wang X, Besançon L, et al. Collaborative work in augmented reality: A survey[J]. IEEE transactions on visualization and computer graphics, 2020, 28(6): 2530-2549.

[14]Indahsari L, Sumirat S. Implementasi teknologi augmented reality dalam pembelajaran interaktif[J]. Cognoscere: Jurnal Komunikasi Dan Media Pendidikan, 2023, 1(1): 7-11.

[15]Subhashini P, Siddiqua R, Keerthana A, et al. Augmented reality in education[J]. Journal of Information Technology and Digital World, 2020, 2(04): 221-227.

[16] Jam, F. A., Ali, I., Albishri, N., Mammadov, A., & Mohapatra, A. K. (2025). How does the adoption of digital technologies in supply chain management enhance supply chain performance? A mediated and moderated model. Technological Forecasting and Social Change, 219, 124225.

[17] Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. International journal of educational technology in higher education, 21(1), 10.