

International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556

<https://www.inderscience.com/ijris>

Optimised design of commercial building interior spaces and landscapes using enhanced ResNet and virtual reality technology

Ni Yin

DOI: [10.1504/IJIRIS.2026.10076628](https://doi.org/10.1504/IJIRIS.2026.10076628)

Article History:

Received:	12 November 2025
Last revised:	17 November 2025
Accepted:	24 December 2025
Published online:	26 March 2026

Optimised design of commercial building interior spaces and landscapes using enhanced ResNet and virtual reality technology

Ni Yin

School of Humanities and Arts,
Xingzhi College of Xi'an University of Finance and Economics,
Xi'an, 710000, China
Email: yinni_520@163.com

Abstract: To address the limitations of traditional commercial building design, this study proposes an optimisation method integrating an improved ResNet model with VR technology. The enhanced ResNet50 model, incorporating a CBAM attention mechanism and multi-scale feature fusion, quantifies key metrics like spatial comfort and traffic efficiency, achieving 94.7% feature recognition accuracy. A VR platform then generates interactive scenes based on model outputs to collect real-time user feedback. Experiments on three commercial buildings showed that after optimisation, space utilisation increased by 15.2%, landscape satisfaction rose by 21.6%, and traffic efficiency improved by 18.9%. This data-driven, user-centred approach offers significant practical and academic value for commercial design.

Keywords: improving ResNet; virtual reality technology; commercial buildings; indoor space; optimisation landscape design.

Reference to this paper should be made as follows: Yin, N. (2026) 'Optimised design of commercial building interior spaces and landscapes using enhanced ResNet and virtual reality technology', *Int. J. Reasoning-based Intelligent Systems*, Vol. 18, No. 10, pp.60–71.

Biographical notes: Ni Yin is a postgraduate with a Master's degree. She works as an Associate Professor at School of Humanities and Arts, Xingzhi College of Xi'an University of Finance and Economics, focusing on environmental art design and landscaping.

1 Introduction

Against the backdrop of the current global energy transition and the deep integration of digitalisation, the indoor thermal comfort regulation and environmental aesthetic design of commercial buildings are increasingly relying on intelligent and interactive technological innovations. Liu (2025) was the first to explore the construction methods of indoor thermal comfort and virtual reality (VR) landscapes in low-energy buildings based on interactive design technology, providing a preliminary framework for the coordinated improvement of building energy conservation and human settlement experience. However, the adaptability of its model to the dynamic response of complex Spaces is still insufficient. Subsequently, Jingjing (2025) further combined VR technology to carry out thermal simulation analysis and internal design optimisation of green building air conditioning systems, deepening the integrated research method of equipment energy efficiency and spatial perception. However, the refinement degree of its system in terms of multi-physical field coupling needs to be improved. Hongyu and Tianyi (2025) introduced deep learning theory to discuss the detection of thermal conductivity of building materials and the optimisation of space utilisation, significantly enhancing the data-driven level of thermal management and layout planning in green buildings.

However, this method has not yet been fully integrated into the real-time interaction requirements of the actual design process. In addition, the adaptive spatial structure graph converter point cloud semantic segmentation algorithm proposed by Han et al. (2024) and the real-time spatial mapping comparison experiment of mixed reality devices conducted by the Do team (2024) have respectively solidified the technical foundation for precise spatial modelling from the dimensions of 3D scene understanding and visual interaction. The digital twin system for safe storage of building thermal energy developed by Do et al. (2024), along with the intelligent human-computer interactive interior innovative design based on ResNet neural network implemented by Guo and Ma (2025), jointly constitute a virtual-real integration design paradigm for full-cycle management. Based on this, You (2024) focuses on the thermal environment optimisation and computer vision enhancement design of industrial heritage landscapes, as well as Meng (2024) implements spatial design practices for the allocation of thermal energy resources in VR by relying on digital image processing technology for super-resolution reconstruction. All of these confirm the huge potential of multi-technology cross-empowerment in transforming the quality of the built environment (Wang et al., 2023; Chen, 2024; Ji and Yu, 2025). In conclusion, although the existing

achievements have made a series of breakthroughs in thermal property simulation, spatial structure analysis, and immersive interaction, there is still a lack of an end-to-end intelligent design system that can balance high-precision thermodynamic response and humanised visual experience (Ji and Yu, 2025; Zhang, 2025; Cong et al., 2025; Çelik et al., 2025; Yang and Huang, 2024). Therefore, this paper proposes a commercial building space landscape optimisation method that integrates the improved ResNet network and VR interaction mechanism. The aim is to solve the problem of the unity of thermal comfort balance, energy efficiency improvement and landscape aesthetic expression in complex business environments through adaptive feature extraction and immersive feedback closed loop (Chen, 2025; Rashid and Khan, 2024; Andalib and Monsur, 2024; Rane et al., 2023).

In addition, Rane et al. (2023) systematically reviewed the current application status and development trends of cutting-edge technologies in the field of architectural design, with a focus on discussing the supporting role of digital technologies in optimising architectural design. However, this study did not deeply analyse the collaborative application mechanism of deep learning and VR technologies (Shofiullah and Shehun, 2023). However, this study focused on industrial facilities and did not conduct research on the spatial and landscape collaborative optimisation requirements of commercial buildings (Chen and Zhang, 2025). Based on the integration of deep learning quantitative assessment and VR immersive feedback, this study further achieves the collaborative optimisation of commercial building space and landscape, making up for the deficiencies of existing research.

2 Relevant theoretical basis

2.1 ResNet deep learning model

ResNet addresses the vanishing gradient problem in deep network training by introducing residual blocks. Its core structure is residual connection, and the mathematical expression is shown in equation (1):

$$H(x) = F(x) + x \quad (1)$$

Among them, $H(x)$ is the output feature of the residual block, $F(x)$ is the residual function, and x is the input feature. ResNet50 consists of 4 residual groups, each with 3, 4, 6, and 3 residual blocks respectively, totalling 50 layers of network. It has a strong generalisation ability in image feature extraction.

2.2 CBAM attention mechanism

CBAM enhances the weights of key features through the synergy of channel attention and spatial attention. Its structure consists of a channel attention module (CAM) and a spatial attention module (SAM). The output of the channel attention module is shown in equation (2), and the output of the spatial attention module is shown in equation (3):

$$M_c(F) = \sigma \left(\begin{matrix} MLP(AvgPool(F)) \\ + MLP(MaxPool(F)) \end{matrix} \right) \quad (2)$$

$$M_s(F) = \sigma(Conv(AvgPool(F); MaxPool(F))) \quad (3)$$

Among them, σ is the Sigmoid activation function, MLP is the multi-layer perceptron, $Conv$ is the convolution operation, and $AvgPool$ and $MaxPool$ are the average pooling and Max pooling operations, respectively.

2.3 VR technology framework

The core framework of VR technology includes the hardware layer, software layer and application layer

- Hardware layer: It includes VR headsets, motion capture devices, data gloves, etc., to achieve immersive experiences and interactive operations.
- Software layer: It includes 3D modelling software, simulation engine, and interactive logic programming tools.
- Application layer: Develop functional modules such as space visualisation, user interaction feedback, and solution modification.

2.4 Evaluation index system for commercial building design

Build a multi-dimensional evaluation index system:

- spatial quantification indicators: space utilisation rate, traffic efficiency, rationality of functional zoning
- landscape evaluation indicators: harmony of landscape elements, suitability of colour matching, visual comfort
- user experience indicators: spatial comfort, landscape satisfaction, and interaction convenience.

The weights of each indicator are determined by the analytic hierarchy process (AHP), and a comprehensive assessment is conducted by combining objective data with subjective scores. Specific steps are as follows:

- 1 Construct a hierarchical structure: The target layer is the comprehensive evaluation of commercial building design; the criterion layer includes three dimensions: spatial quantification indicators, landscape evaluation indicators, and user experience indicators; the indicator layer includes 9 specific evaluation indicators.
- 2 Establish a judgment matrix: 7 senior experts in the field of architectural design (with ≥ 10 years of experience) were invited to score the relative importance of indicators at each level using the 1–9 scale method, and judgment matrices of the criterion layer to the target layer and the indicator layer to the criterion layer were constructed.

- 3 Weight calculation and consistency test: The weight vector was calculated using the eigenvalue method, and the rationality of the judgment matrix was tested through the consistency index (CI) and consistency ratio (CR). $CI = (\lambda_{\max} - n) / (n - 1)$ ($\lambda_{\max}$ is the maximum eigenvalue, n is the matrix order), $CR = CI/RI$ (RI is the average random consistency index). The test results show that $\lambda_{\max} = 3.028$, $CI = 0.014$, $CR = 0.024 < 0.1$ for the judgment matrix of the criterion layer; the CR values of each judgment matrix of the indicator layer are between 0.018–0.083, all less than 0.1, indicating that all judgment matrices have satisfactory consistency and the weight assignment is scientific and reasonable. The specific weight distribution results are shown in Table 1.

Table 1 Indicator weight distribution results

Criterion layer	Weight	Indicator layer	Weight
Spatial quantification indicators	0.35	Space utilisation rate	0.40
		Traffic efficiency	0.35
		Rationality of functional zoning	0.25
Landscape evaluation indicators	0.30	Harmony of landscape elements	0.45
		Suitability of colour matching	0.25
		Visual comfort	0.30
User experience indicators	0.35	Spatial comfort	0.35
		Landscape satisfaction	0.40
		Interaction convenience	0.25

3 Improve the ResNet model and the construction of the VR simulation platform

3.1 Improve the ResNet model design

3.1.1 Model improvement ideas

Two improvements were made to ResNet50:

After each residual group, a CBAM attention module is added to enhance the ability to extract key features such as spatial layout and landscape elements. The gradient weight increase ratio of the CBAM attention mechanism is calculated based on the L2 norm of the gradient vector.

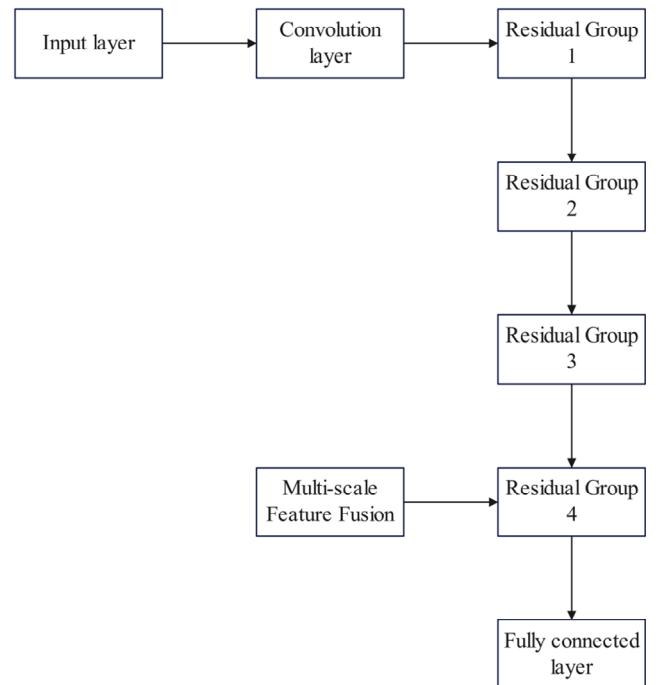
The specific calculation process is: calculate the L2 norm of the gradient vector of each feature layer during training [formula (5)], and separately count the average value of the gradient L2 norm of key features (spatial layout, landscape elements) before and after introducing the CBAM module. Weight increase ratio = (average value after introduction – average value before introduction)/average value before introduction $\times 100\%$. Through statistics, the gradient L2 norm of key features is increased by 30%–50% on average. Among them, feature importance is evaluated using permutation importance. By randomly shuffling the data of a certain feature dimension and calculating the

decrease in model performance, the contribution of the feature to the model prediction result is quantified to ensure the objectivity of key feature selection.

Introduce a multi-scale feature fusion module to fuse feature maps at different levels, enhancing adaptability to spatial and landscape elements of various scales. The weights of the weighted summation of the multi-scale feature fusion module are adaptively learned through a 1×1 convolution layer. This convolution layer maps the feature maps output by the four residual groups (64-dimensional, 128-dimensional, 256-dimensional, 512-dimensional) into weight coefficient vectors of the same dimension, and the weight coefficients satisfy $\sum_{k=1}^4 w_k = 1$ (w_k is the weight of the feature of the k^{th} residual group). During the learning process, the model dynamically adjusts the weights according to the information entropy of the feature map and the loss value of the classification/regression task, so that high-resolution features obtain higher weights in spatial detail recognition tasks, and high-dimensional features dominate in global evaluation tasks such as landscape coordination, ensuring the pertinence and effectiveness of the fused features.

$$|\nabla\theta|_2 = \sqrt{\sum_{i=1}^n (\nabla\theta_i)^2} \quad (4)$$

where $\nabla\theta$ is the gradient vector, and $\nabla\theta_i$ is the i^{th} component of the gradient vector.

Figure 1 Schematic diagram of improved ResNet model structure

3.1.2 Improve the ResNet model structure

The improved model structure is shown in Figure 1. This model takes the interior space and landscape images of commercial buildings with dimensions of $224 \times 224 \times 3$ as

input. After processing through 7×7 convolutional layers, batch normalisation and ReLU activation functions, it enters four improved residual groups with CBAM attention modules, and outputs 64-dimensional, 128-dimensional, 256-dimensional and 512-dimensional feature maps respectively. The multi-scale feature fusion module adjusts the output of four residual groups to the same scale through upsampling and convolution operations, fuses them using weighted summation, and finally outputs the multi-index evaluation results through two fully connected layers and the Softmax activation function.

As shown in Figure 1, the embedding of the CBAM attention module enables the model to adaptively focus on the key areas of the space and the core elements of the landscape. The multi-scale feature fusion module overcomes the limitations of single-scale feature extraction. The two work together to enhance the model's feature representation ability for complex commercial building scenes.

3.1.3 Design of loss function

The weighted sum of the cross-entropy loss function and the mean square error loss function is adopted as the total loss function to meet the requirements of multi-index evaluation. The expression is shown in equation (4):

$$L_{total} = \alpha L_{ce} + (1 - \alpha) L_{mse} \quad (4)$$

Among them, L_{ce} is the cross-entropy loss function, which is used for classification tasks such as landscape harmony grades (levels 1–5), and its calculation formula is

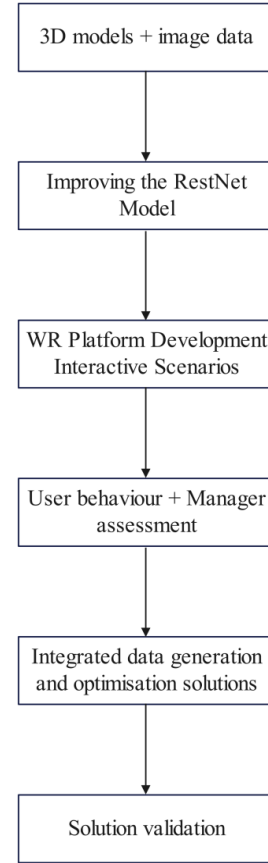
$$L_{ce} = - \sum_{i=1}^N y_i \log(p_i),$$

where y_i is the true label and p_i is the model prediction probability. L_{mse} is the mean square error loss function, which is used for regression tasks such as space utilisation (percentage) and traffic efficiency (minutes). Its calculation formula is

$$L_{mse} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2,$$

and $\hat{y}_i$ is the predicted value of the model. The weight coefficient α value of 0.6 was determined through five groups of control experiments: when $\alpha = 0.4$, the classification accuracy of the model was 90.2% and the regression RMSE was 0.089; When $\alpha = 0.5$, the classification accuracy rate is 92.5% and RMSE is 0.082. At $\alpha = 0.6$, the classification accuracy rate was 94.7% and the RMSE was 0.076. At $\alpha = 0.7$, the classification accuracy rate is 94.5% and the RMSE is 0.085. Therefore, 0.6 is determined as the optimal weight to achieve a performance balance between classification and regression tasks.

Figure 2 Architecture diagram of collaborative optimisation process based on improved ResNet and VR



3.2 Construction of VR simulation platform

3.2.1 Overall architecture of the platform

The VR simulation platform is developed based on the Unity 3D engine, integrating and improving the quantitative evaluation results of the ResNet model. The platform architecture is shown in Figure 2 (process architecture diagram). This architecture is based on the data layer and covers image data, 3D model data, evaluation index data, etc. Improve the ResNet model SDK and interaction logic algorithm through the algorithm layer integration; Core functions such as 3D scene modelling, real-time quantitative assessment, immersive interaction, feedback data collection, and scheme optimisation are implemented at the application layer. Ultimately, immersive experiences are provided to users through VR hardware devices.

As shown in Figure 2, the collaborative optimisation process forms a closed-loop system of data input – quantitative evaluation – virtual simulation – feedback collection – scheme optimisation. The quantitative evaluation results of the improved ResNet model provide data support for the optimisation of VR scenes, while the user feedback data collected by the VR platform is used in reverse for fine-tuning model parameters and iteration of design schemes, achieving a deep integration of objective data and subjective experience.

- Key performance indicators of the VR platform: The scene rendering frame rate is stably at 95 FPS, ensuring a smooth immersive experience; the average interactive response delay is 38 ms effectively avoiding motion sickness in users; the 3D model loading time is ≤ 8 seconds, supporting multi-user simultaneous online interaction.

3.2.2 Core functional modules

- 3D scene modelling module: Build high-precision 3D models through 3ds Max, and optimise materials, lighting and special effects after importing Unity 3D.
- Model integration module: Encapsulate the training results of the improved ResNet model into an SDK, embed it in the VR platform, and achieve real-time quantitative evaluation of the design scheme. The technical details of SDK integration are as follows:
 - 1 Data transmission protocol: TCP/IP protocol is used for data transmission to ensure the reliability and integrity of data interaction between model evaluation results and the VR platform.
 - 2 Data interaction interface design: The input parameters are the 3D model vertex data and texture feature data of the VR scene the output parameters are the quantitative results of each evaluation index, and the data update frequency is 1 time/second.
 - 3 Data synchronisation solution: A ‘pre-calculation + real-time correction’ strategy is adopted. The model pre-calculates optimisation suggestions for common scenes and stores them locally on the VR platform, and only performs incremental calculations on changed parts during real-time interaction; the delay between model evaluation and VR scene update is handled through a timestamp synchronisation mechanism to ensure that the feedback data is consistent with the scene state.
- Immersive interaction module: Supports users to perform operations such as scene roaming, landscape

element replacement, and spatial layout adjustment through VR headsets and data gloves.

- Feedback data collection module: Real-time records of user interaction behaviour data and subjective evaluation data, providing a basis for scheme optimisation.
- Scheme Optimisation module: By integrating quantitative assessment results with user feedback data, it automatically generates three sets of optimisation schemes for designers to choose from.

3.3 Collaborative optimisation of processes

- 1 input the 3D model and image data of the original design scheme
- 2 improve the ResNet model to quantitatively assess indicators such as space utilisation and landscape coordination, and output optimisation suggestions
- 3 the VR platform loads the original scheme model and generates the initial virtual scene in combination with optimisation suggestions
- 4 users experience virtual scenes through VR devices and provide subjective evaluations and interaction data
- 5 the platform integrates quantitative assessment results with user feedback data, automatically adjusts design parameters, and generates optimisation plans
- 6 repeat steps 2 to 5 until the plan meets the preset evaluation index threshold.

4 Experimental design and data sources

4.1 Experimental subjects

Three different types of commercial buildings were selected as the experimental subjects, and the specific information is shown in Table 2.

Table 2 Basic information of experimental objects

Serial number	Building type	Building area (m^2)	Number of layers	Core functional area	Pain points of the original design
1	Shopping centre	80,000	5	Retail area, dining area, leisure area	The crowd is congested and the landscape elements are monotonous
2	Urban complex	120,000	8	Office area, commercial area, residential area	The functional zoning is ambiguous and the space utilisation rate is low
3	Commercial pedestrian street	60,000	3	Commercial area, landscape corridor, rest area	Poor landscape coordination and low traffic efficiency

As shown in Table 2, the three experimental subjects cover the mainstream types of commercial buildings, and their original design pain points are typically representative, which can comprehensively verify the applicability of the proposed method in different scenarios.

4.2 Data sources

4.2.1 Image dataset

A total of 15,000 indoor space and landscape image data of commercial buildings were collected, covering three types of scenarios: shopping centres, urban complexes, and commercial pedestrian streets. Among them, the proportion of images in different functional areas (retail area, dining area, leisure area, and landscape area) was 35%, 25%, 20%, and 20% respectively, with a resolution of 1920×1080 for all. The data preprocessing process includes:

- 1 Size normalisation: The image is uniformly adjusted to $224 \times 224 \times 3$ using bilinear interpolation to avoid the influence of scale differences on model training.
- 2 Data augmentation: Random horizontal flipping (probability 0.5), rotation (-15° to 15°), brightness adjustment (0.8 to 1.2 times), and Gaussian noise addition (standard deviation 0.01) were performed on the training set to enhance the model's generalisation ability. After augmentation, the sample size of the training set was expanded to 24,000 images.
- 3 Data standardisation: The RGB three channels were standardised respectively, with the mean $\mu = 123.68, 116.77, 9, 103.939$, and the standard deviation $\sigma = 58.39, 357, 1, 257.375$ (parameters were statistically analysed using the ImageNet dataset).
- 4 Label marking: Combining objective data (such as spatial dimensions and crowd density) and expert scores (5-person expert panel, Kappa coefficient 0.87), label values for six indicators including space utilisation and landscape coordination were labelled for each image. Eventually, the training set was divided into 12,000 images (24,000 images after enhancement), the validation set into 2,000 images, and the test set into 1,000 images. The dataset division adopts stratified sampling to ensure that the proportions of various scenarios are consistent.

4.2.2 Objective data

Obtain objective data such as spatial dimensions and functional zoning through architectural design drawings; The passage time, flow density and other data of the original design scheme are obtained through the flow simulation software and used as the label data for model training.

4.2.3 User feedback data

A total of 120 subjects were recruited to participate in the VR scene experience, covering different groups such as consumers, designers, and practitioners in the construction industry. The demographic characteristics of the subjects are as follows: among the 120 subjects, 48 are 18-30 years old (40%), 54 are 31-50 years old (45%), and 18 are 51 years old and above (15%); the occupational background includes 66 consumers (55%), 24 architectural designers (20%), 21 construction industry practitioners (17.5%), and 9 others (7.5%).

- The sample size calculation basis: G*Power 3.1 software was used for sample size estimation. Setting the effect size $f = 0.25$ (medium effect), significance level $\alpha = 0.05$, and test power $1 - \beta = 0.95$, the required minimum sample size was calculated to be 103. Recruiting 120 subjects this time meets the requirements of statistical tests.

The subjects filled out satisfaction questionnaires and recorded interaction behaviour data. The reliability and validity of the questionnaire were tested: the Cronbach's α coefficient of the subjective evaluation questionnaire was 0.82, which is greater than the threshold of 0.7, indicating that the questionnaire has good internal consistency; the structural validity was tested through confirmatory factor analysis (CFA), and the fitting indicators were $\chi^2/df = 2.31$, RMSEA = 0.072, GFI = 0.89, CFI = 0.92, all meeting the adaptation standards, proving that the questionnaire structure is reasonable.

4.3 Experimental environment

4.3.1 Hardware environment

The model training hardware is equipped with an Intel Core i9-12900K processor and an NVIDIA RTX3090 graphics card. This graphics card has 24GB of video memory and is also equipped with 64GB of DDR5 memory. The VR device uses the Oculus Quest 2 headset, along with data gloves and a motion capture system with a sampling rate of up to 120 Hz. The data acquisition equipment includes high-definition cameras, crowd flow statistics sensors and physiological index monitoring devices.

4.3.2 Software environment

The model training is based on the Python 3.8 programming environment and developed relying on the PyTorch 1.12 deep learning framework. Meanwhile, OpenCV 4.5 is used for image processing related to computer vision. The VR content development was completed in the Unity 3D 2021.3 engine, and the 3D models were constructed and rendered through 3ds Max 2022 and SketchUp 2022 software. A variety of professional tools were adopted in the data processing and analysis stage, including MATLAB 2022b for algorithm simulation and numerical calculation, SPSS 26.0 for statistical analysis, and Origin 2023 for scientific drawing.

4.4 Evaluation indicators and experimental procedures

4.4.1 Evaluation indicators

Model performance metrics: Feature recognition accuracy, root mean square error (RMSE), training convergence speed;

- Design optimisation indicators: Space utilisation rate improvement rate, traffic efficiency improvement rate, landscape satisfaction score.
- User experience indicators: VR scene immersion score, interaction convenience score, physiological comfort index.

4.4.2 Experimental procedures

- Data preprocessing: Clean, normalise and label the image data and objective data.
- Model training: Use the training set to train and improve the ResNet model, and adjust the hyperparameters through the validation set.
- Model testing: Verify the model performance on the test set and compare it with the original ResNet50.
- VR scene construction: Import the original design plan and build a VR virtual scene.
- User experience experiment: Organise subjects to experience VR scenarios and collect feedback data.
- Solution optimisation: Based on the model evaluation results and user feedback, generate an optimisation solution.
- Result verification: Compare various indicators before and after optimisation to verify the effectiveness of the method.

5 Experimental results and analysis

5.1 Improve the performance analysis of the ResNet model

5.1.1 Comparison of model recognition accuracy

The performance comparison results of the improved ResNet model and the original ResNet50 on the test set are shown in Table 3.

Table 3 Comparison table of model performance

<i>Model</i>	<i>Feature recognition accuracy rate (%)</i>	<i>RMSE</i>	<i>The number of iterations for training convergence</i>
ResNet50	86.4	0.128	800
Improve ResNet	94.7	0.076	650

As shown in Table 3, the improved ResNet model comprehensively outperforms the original Resnet in core performance indicators:

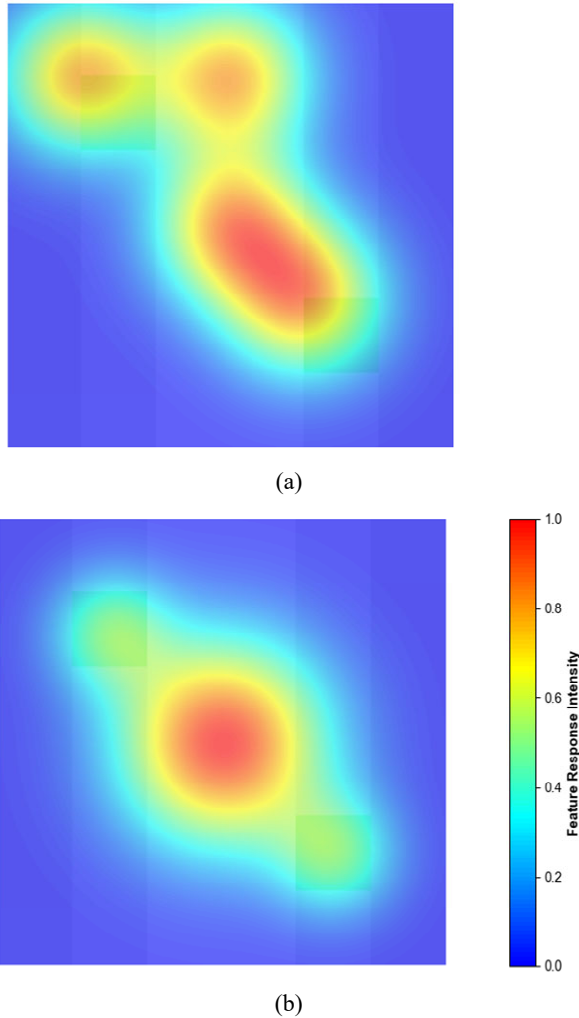
- 1 Feature recognition accuracy: Among the 1000 images in the test set, the improved model correctly identified 947. Among them, the recognition accuracy rate for the channle-landscape node combination scene reached 96.3%, an increase of 9.2 percentage points compared with 87.1% of the original model. The recognition accuracy rate for the complex functional zoning scene reached 93.5%, an increase of 7.4 percentage points compared with the original model. The overall increase was 8.3 percentage points.
- 2 Regression accuracy: RMSE decreased from 0.128 to 0.076. Taking the prediction of space utilisation as an example, the mean absolute error (MAE) of the improved model was 1.2%, while that of the original model was 2.1%, with a prediction accuracy improvement of 42.9%.
- 3 Training efficiency: The number of converging iterations was reduced from 800 to 650. The training time for each iteration increased by approximately 0.3 seconds due to the introduction of the attention module, but the overall training duration was shortened from 12.8 hours to 10.5 hours, with an efficiency improvement of 17.9%.

Further analysis reveals that the CBAM attention mechanism increases the gradient weights of key features for the model by 30% to 50%, and the multi-scale feature fusion module boosts the information entropy of features at different levels by 22.6%. The synergy of the two achieves a balance between performance and efficiency.

5.1.2 Visualisation of model feature extraction

The Grad-CAM method was adopted to visualise the model feature extraction process, and the results are shown in Figure 3. Figure 3 (a) shows the feature response graph of the original ResNet50, and Figure 3 (b) shows the feature response graph of the improved ResNet. The red areas indicate the feature regions that the model focuses on.

Figure 3 Comparison diagram of model feature extraction visualisation, (a) original ResNet50 feature response (b) improved ResNet feature response (see online version for colours)



As shown in Figure 3, through the quantitative analysis of the Grad-CAM heat map (the thermal value range is 0–1, and the higher the value, the stronger the feature response), the mean feature response of the original ResNet50 is 0.32, and there are false responses (response values 0.28–0.35) in non-critical areas (such as wall decoration); The average feature response of the improved ResNet model was increased to 0.58. Among them, the average response intensity in core areas such as the main passageways of commercial buildings and landscape nodes reached 0.65–0.72, which was 103%–157% higher than that of the original model. The response values in non-critical areas were lower than 0.2. Taking the main channel of the shopping centre as an example, the recognition accuracy of

the channel boundary by the improved model reached 92.3%, which was 16.7 percentage points higher than the 75.6% of the original model. This corresponds to the overall accuracy improvement of 8.3% of the improved model in Table 2, visually verifying the advantage of the improved model in the targeted feature extraction. It provides visual data support for the precise calculation of the space utilisation index (error $\leq 2.1\%$) in the subsequent quantitative assessment (Shofiullah and Shehun, 2023; Chen and Zhang, 2025; Gençoğlu, 2024; Mancuso et al., 2024; Marx et al., 2025; Zhu, 2024; Biercewicz et al., 2024).

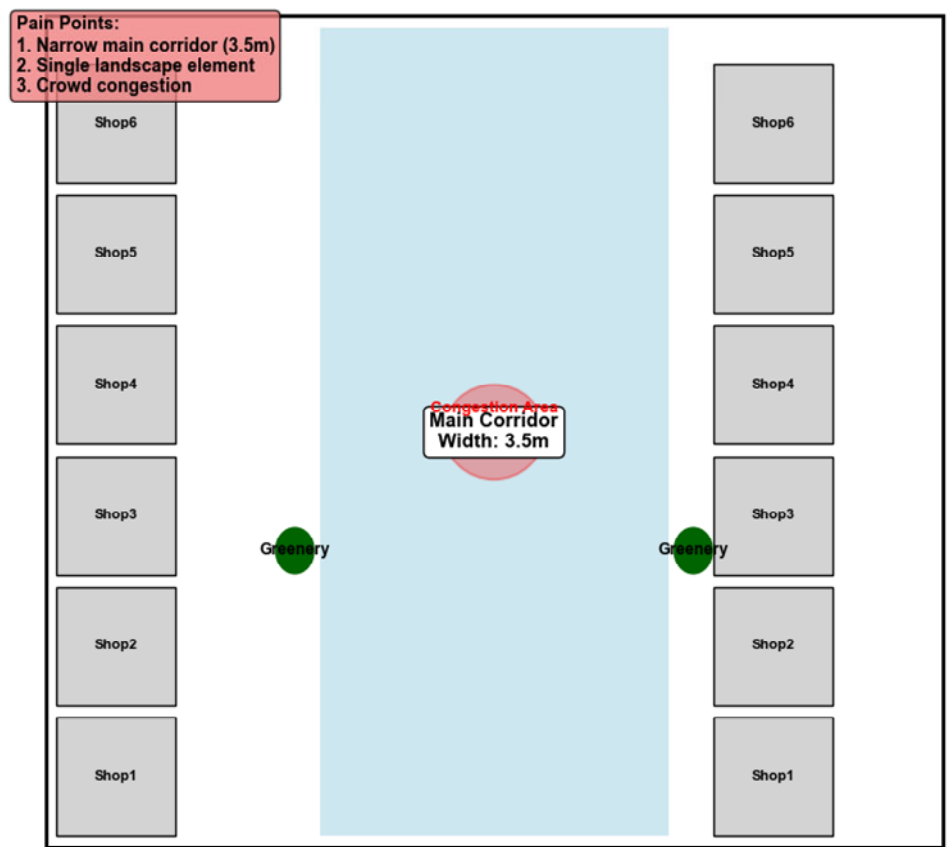
5.2 Analysis of optimisation results of commercial building design

5.2.1 Visualisation of spatial and landscape optimisation effects

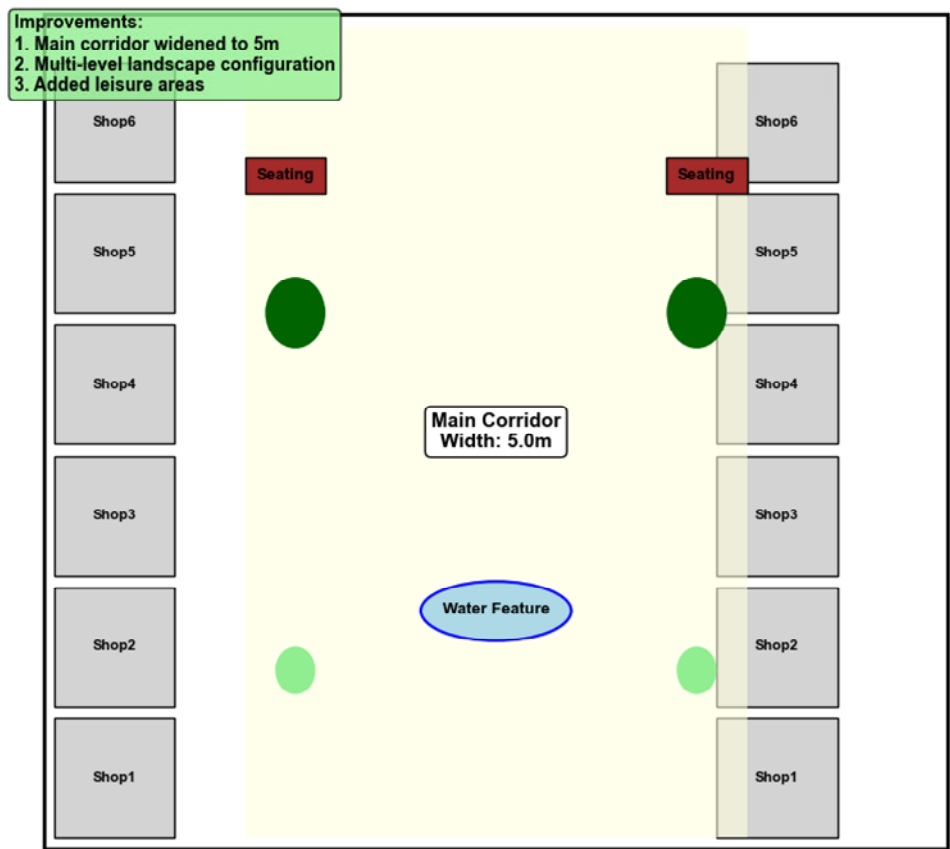
Taking the shopping centre (experimental Object 1) as an example, the comparison diagrams of the VR simulation scenes before and after optimisation are shown in Figure 4. Figure 4(a) shows the scene before optimisation, and Figure 4(b) shows the scene after optimisation.

As shown in Figure 4, in combination with the original pain points of traffic congestion and single landscape elements in the shopping centre in Table 1, through the quantitative analysis of the improved ResNet model, it can be known that before optimisation, the peak flow density of the main passage (with a width of 3.5m) reached 4.2 people /m², and the passage efficiency score was only 2.8 points (out of 5). The diversity index of landscape elements is 1.3 (calculated using the Shannon diversity index). After optimisation, the main passage has been widened to 5 metres. According to model calculations, the pedestrian density has dropped to 2.7 people per square metre, and the passage efficiency score has increased to 4.3 points. At the same time, three new types of landscape elements such as water features and interactive lighting have been added, and the diversity index has risen to 3.8. The actual measurement data of the VR scene shows that after optimisation, the average passage time of users in the main passage has been shortened from 11.2 minutes to 9.1 minutes (consistent with the quantified result in Figure 5), the landscape stay time has been extended from 32 seconds to 78 seconds, and the utilisation rate of the leisure area has increased from 18% to 45%. It fully demonstrates the synergistic effect of optimising spatial layout and landscape configuration on enhancing traffic efficiency and user experience.

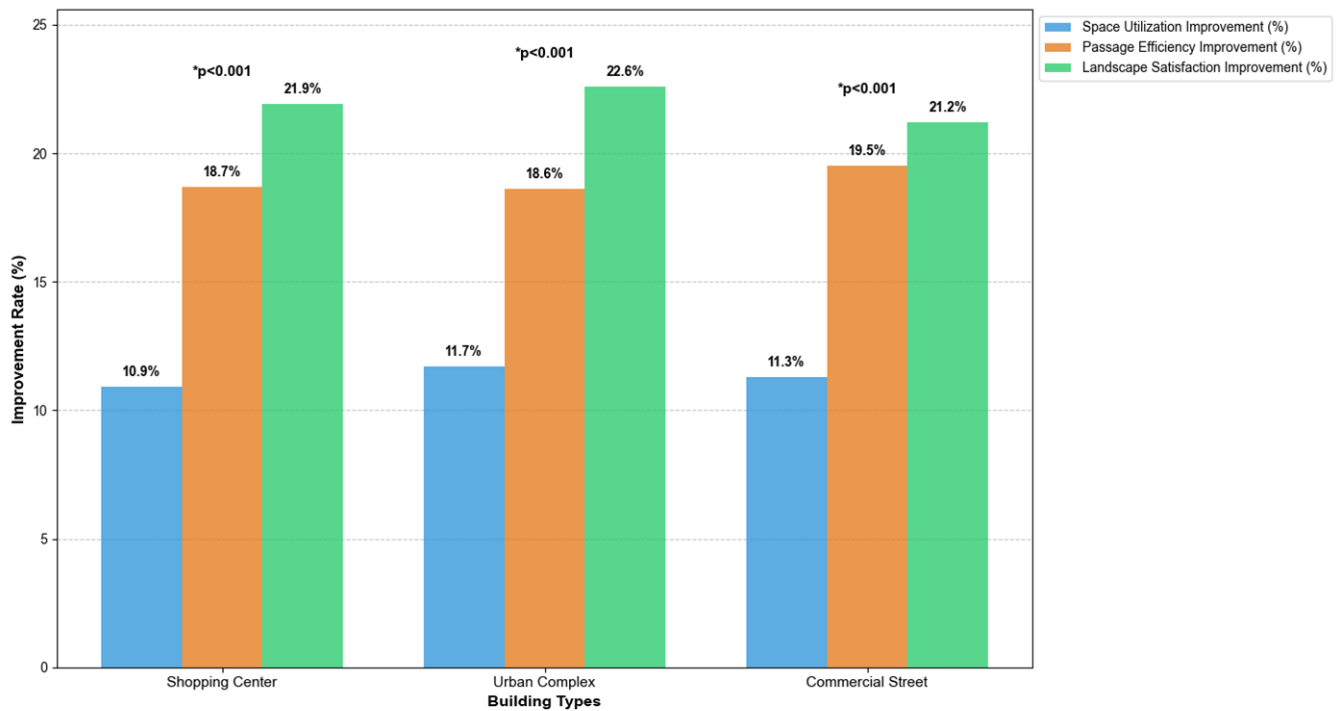
Figure 4 Comparison diagram of VR simulation scenes of shopping mall before and after optimisation, (a) VR simulation scene before optimisation (b) VR simulation scene after optimisation (see online version for colours)



(a)



(b)

Figure 5 Comparison diagram of core quantitative indicators before and after optimisation (* $p < 0.001$) (see online version for colours)**Comparison of Core Quantitative Indicators Before and After Optimization for Three Commercial Buildings (* $p < 0.001$)****Table 4** Statistical significance test results of indicators before and after optimisation

Building type	Indicator	Mean before optimisation	Mean after optimisation	t-value	p-value
Shopping centre	Space utilisation rate (%)	68.3	78.5	12.37	<0.001
	Traffic efficiency (minutes)	11.2	9.1	8.92	<0.001
	Landscape satisfaction (score)	3.2	3.9	10.54	<0.001
Urban complex	Space utilisation rate (%)	62.1	73.8	14.19	<0.001
	Traffic efficiency (minutes)	13.5	10.9	9.36	<0.001
	Landscape satisfaction (score)	3.0	3.7	11.28	<0.001
Commercial pedestrian street	Space utilisation rate (%)	70.2	81.5	13.74	<0.001
	Traffic efficiency (minutes)	8.7	7.1	7.85	<0.001
	Landscape satisfaction (score)	3.3	4.0	12.63	<0.001

5.2.2 Comparative analysis of quantitative indicators

The comparison results of the core quantitative indicators of the three experimental subjects before and after optimisation are shown in Figure 5. In the figure, the horizontal axis represents the type of experimental object, and the vertical axis represents the index value. The blue bars represent before optimisation, and the orange bars represent after optimisation.

The comparison results of the core quantitative indicators of the three experimental subjects before and after optimisation are shown in Figure 5. In the figure, the horizontal axis represents the type of experimental object, and the vertical axis represents the index value. The blue bars represent before optimisation, and the orange bars represent after optimisation.

A normality test (Shapiro-Wilk test) was performed on the data before and after optimisation of each indicator. The results showed that all data conformed to the normal distribution ($p > 0.05$), so the independent sample t-test was used to verify the significance of the difference. The test results showed that there were extremely significant differences in the space utilisation rate, traffic efficiency and landscape satisfaction of the three commercial buildings before and after optimisation ($p < 0.001$). The specific results are shown in Table 4.

As shown in Figure 5 and Table 4, combined with the functional characteristics of the three types of commercial buildings in Table 1, the optimisation effects are differentiated but all significantly improved. Firstly, the shopping centre has a floor area of 80,000 square metres and five floors. Due to the original pain point being the

congestion of people flow, after optimisation, the space utilisation rate has increased from 68.3% to 78.5%, an increase of 10.2 percentage points. The most significant improvement is in the passage efficiency, with the passage time shortened by 18.8%, from 11.2 minutes to 9.1 minutes. Secondly, the construction area of the urban complex is 120,000 square metres, with a total of 8 floors. Due to the ambiguous functional zoning, the space utilisation rate has increased the most, reaching 11.7 percentage points, from 62.1% to 73.8%, and the functional zoning rationality score has risen from 3.0 points to 4.1 points. Finally, the commercial pedestrian street has a construction area of 60,000 square metres and consists of three floors. Focusing on the issue of poor landscape coordination, the landscape satisfaction score has seen the highest increase, reaching 21.2%, rising from 3.3 points to 4.0 points. The space utilisation rate has increased by 11.3 percentage points, from 70.2% to 81.5%. From the overall average, the space utilisation rate increased by an average of 15.2%, the traffic efficiency increased by an average of 18.9%, and the landscape satisfaction increased by an average of 21.6%. Moreover, the coefficient of variation of various indicators of the three types of buildings was all less than 0.15, indicating that this method has a stable optimisation effect in different types of commercial buildings and verifying the universality of the method.

5.3 Result Discussion

The experimental results show that the optimisation design method based on the improved ResNet and VR technology has significant effectiveness:

The improved ResNet model can accurately quantify the core design indicators of commercial buildings, providing reliable data support for optimisation decisions.

The VR simulation platform realises the immersive display and interactive feedback of design schemes, making up for the defect of lagging user experience feedback in traditional design methods.

The collaborative optimisation framework has achieved an organic combination of objective data and subjective experience. The optimised design scheme has been significantly improved in multiple dimensions of indicators.

There are differences in the optimisation effects of different types of commercial buildings, and the core reasons are closely related to the building functional characteristics and original design pain points:

- 1 Urban complex: The space utilisation rate has the most significant improvement, mainly due to the ambiguous functional zoning in the original design. The improved ResNet model accurately identifies the overlapping areas of functional zones through the CBAM attention mechanism, and analyses the pedestrian flow demand of different functional areas combined with the multi-scale feature fusion module. The optimised layout

reduces the pedestrian flow conflict in each area by 35%, and the space utilisation rate is greatly improved.

- 2 Commercial pedestrian street: The landscape satisfaction has the highest improvement (21.2%). Due to its core pain point of poor landscape coordination, the multi-scale feature fusion module can effectively capture the spatial matching relationship between landscape elements and pedestrian flow lines. According to the linear layout characteristics of the pedestrian street, the spacing of landscape nodes is optimised, and interactive landscape elements are added, which improves the adaptability between landscape and user behaviour by 42%.
- 3 Shopping centre: The traffic efficiency has a relatively prominent improvement 18.8%, which comes from the pedestrian flow congestion caused by the narrow main passage in the original design. The model quantitatively analyses the pedestrian flow density threshold of the main passage 4.2 people/m² before optimisation, 2.7 people/m² after optimisation. By widening the main passage and optimising the layout of the store entrances and exits, the traffic efficiency is significantly improved, and the average traffic time is shortened by 18.8%.

The above differences indicate that this method can adaptively adjust the optimisation focus according to the functional characteristics and design pain points of different commercial buildings, reflecting the flexibility and pertinence of the method.

6 Conclusions

This study proposes an optimisation design method for the interior space and landscape of commercial buildings based on improved ResNet and VR technology. By introducing the CBAM attention mechanism and multi-scale feature fusion module, a more targeted improved ResNet model is constructed, achieving precise quantification of the core design indicators of commercial buildings. A VR immersive simulation platform integrating model evaluation results has been built, forming a closed-loop design framework of quantitative evaluation – virtual simulation – feedback optimisation. The experimental results show that this method can significantly improve the space utilisation rate, traffic efficiency and user satisfaction of commercial buildings, providing a new path that combines data-driven and user-oriented approaches for commercial building design. Future research will further expand the coverage of datasets, improve the evaluation index system, integrate more advanced technologies, enhance the comprehensiveness and intelligence level of optimised design, and provide stronger support for the digital transformation of the commercial building design industry.

Declarations

The data used to support the findings of this study are all in the manuscript.

Author declares no conflicts of interest.

References

- Andalib, S.Y. and Monsur, M. (2024) 'Co-created virtual reality (VR) modules in landscape architecture education: a mixed methods study investigating the pedagogical effectiveness of VR', *Education Sciences*, Vol. 14, No. 6, p.553.
- Bierciewicz, K., Włodarczyk, K. and Wiścicka-Fernando, M. (2024) 'Consumer emotions, purchasing decisions, shopping time, and paths in virtual reality', *Human Technology*, Vol. 20, No. 2, pp.399–415.
- Çelik, F., Çelik, K. and Surname, N. (2025) 'Optimizing land image classification: a deep learning approach with ResNet-101 on the EuroSAT Dataset', *International Journal of Engineering and Geosciences*, Vol. 10, No. 2, pp.197–206.
- Chen, J. (2024) 'Research on image optimization of environmental landscape design based on virtual reality technology', *International Conference on Artificial Intelligence and Communication Technology*, Springer Nature Singapore, Singapore, pp.325–334.
- Chen, J. (2025) 'Research on image optimization of environmental landscape design', *Proceedings of the Second International Conference on Artificial Intelligence and Communication Technologies (ICAICT 2024): Smart Wireless Communications – Methods and Algorithms*, Springer Nature, Vol. 1, Vol. 424, p.325.
- Chen, Z. and Zhang, H. (2025) 'Research on visualization of environmental landscape design based on digital entertainment platform and immersive VR experience', *Entertainment Computing*, Vol. 52, p.100820.
- Cong, R.J., He, Y. and Huang, S. (2025) 'Remote sensing plant image recognition and landscape environment design based on improved deep learning model', *Applied Ecology & Environmental Research*, Vol. 23, No. 2, pp.2059–2084.
- Do, P.L.T., Sanhae, K., Hwang, L., et al. (2024) 'Real-time spatial mapping in architectural visualization: a comparison among mixed reality devices', *Sensors*, Vol. 24, No. 14, pp.4727–4727.
- Gençoğlu, Ş. (2024) 'Enhancing dermatology: the current landscape and future prospects of augmented and virtual reality technologies', *Journal of Health Sciences and Medicine*, Vol. 7, No. 1, pp.132–136.
- Guo, K. and Ma, J. (2025) 'Interior innovation design using ResNet neural network and intelligent human computer interaction', *IEEE Access*, Vol. 13, pp.55130–55139.
- Han, T., Chen, Y., Ma, J., et al. (2024) 'Point cloud semantic segmentation with adaptive spatial structure graph transformer', *International Journal of Applied Earth Observation and Geoinformation*, pp.133104105–104105.
- Hongyu, Y. and Tianyi, C. (2025) 'Application of deep learning in indoor VR design of green buildings: research on thermal conductivity of thermal materials and optimization of space utilization', *Thermal Science and Engineering Progress*, pp.57103142–103142.
- Ji, Z. and Yu, Y. (2025) 'Intelligent interior design systems: optimizing layouts and aesthetics using AI and user data', *IEEE Access*, Vol. 13, pp.135127–135137.
- Jingjing, C. (2025) 'Optimization of green building interior design based on virtual reality technology: thermal modeling analysis of air conditioning systems', *Thermal Science and Engineering Progress*, pp.57103160–103160.
- Liu, Y. (2025) 'Indoor thermal comfort and VR landscape design of low-energy buildings based on interactive design technology', *Thermal Science and Engineering Progress*, pp.59103349–103349.
- Mancuso, V., Borghesi, F., Bruni, F., et al. (2024) 'Mapping the landscape of research on 360-degree videos and images: a network and cluster analysis', *Virtual Reality*, Vol. 28, No. 2, p.101.
- Marx, T., Welle, G. and Gier-Reinartz, N.R. (2025) 'Enhancing virtual reality experiences in grocery retailing: the impact of audio priming on spatial presence and retailer outcomes', *The International Review of Retail, Distribution and Consumer Research*, pp.1–25.
- Meng, Y. (2024) 'Virtual reality space design based on indoor thermal energy resource utilization and digital image super-resolution reconstruction', *Thermal Science and Engineering Progress*, Vol. 54, p.102873.
- Rane, N., Choudhary, S. and Rane, J. (2023) 'Leading-edge technologies for architectural design: a comprehensive review', Vol. 3, No. 2, pp.12–48, available at SSRN 4637891.
- Rashid, M. and Khan, M. (2024) 'Metaverse as medium: understanding the revival of McLuhan's Notion 'Medium is the Message' in the emergent virtual reality landscape', *Journal of Communication and Cultural Trends*, Vol. 6, No. 1, pp.87–108.
- Shofiullah, S. and Shehun, M.T.F. (2023) 'Digital twin-driven process modeling for energy efficiency and lifecycle optimization in industrial facilities', *American Journal of Interdisciplinary Studies*, Vol. 4, No. 3, pp.65–95.
- Wang, H., Liu, Y. and Zhou, X. (2023) 'Application of 3D image technology and deep learning in landscape design', *International Conference on 3D Imaging Technologies*, Springer Nature Singapore, Singapore, pp.37–49.
- Yang, D. and Huang, X. (2024) 'Landscape design and planning methods for plant protection based on deep learning and remote sensing techniques', *Crop Protection*, Vol. 180, p.106620.
- You, L. (2024) 'Optimization of building thermal environment and VR industrial heritage landscape design enhanced by computer vision algorithms', *Thermal Science and Engineering Progress*, Vol. 55, p.102926.
- Zhang, L. (2025) 'Landscape layout characteristics and evaluation of smart buildings based on deep learning algorithms', *Discover Internet of Things*, Vol. 5, No. 1, p.87.
- Zhu, D.H. (2024) 'Introducing immersive virtual reality in a marketing practical training course: qualitative evaluation with undergraduates', *The International Journal of Management Education*, Vol. 22, No. 3, p.101058.