Enhanced DeepID Network-Based Access Control for Property Management Using Transformer and Gaussian Mixture Models

Huiqi Zhang
Zhengzhou Urban Construction Vocational College, Zhengzhou 451263, China
Email: solheng@163.com

Keywords: DeepID network, transformer model, multi-head attention mechanism, access control, property management system, face recognition

Received: January 23, 2025

In recent years, property backend access control systems have faced many challenges in terms of management efficiency and security. Traditional authentication methods are difficult to cope with complex and changing access scenarios, leading to security vulnerabilities. To address this issue, a property backend access control model based on an improved DeepID network is designed. The Transformer model is introduced to optimize the feature extraction capability of the DeepID network, and the multi-head attention mechanism is applied to optimize feature expression and improve the accuracy of face feature recognition for users. Moreover, a Gaussian mixture model is introduced to accurately model user behavior patterns. The experimental study was conducted on a platform with Intel Core i9-12900K, NVIDIA RTX 3090, and 64GB RAM, using 13,233 face images from LFW dataset for training and 5,000 images from CelebA dataset for validation. According to the results, the improved DeepID network model achieved a feature recognition accuracy of 97.3%, with a loss value of only 0.15, statistically significantly outperforming traditional DeepID network (86.7%), VGG-Face (92.4%), and FaceNet (94.9%) in terms of F1-score and precision-recall metrics. The research provides an efficient and reliable technical solution for the property backend access control system, which has important practical significance for improving the intelligence and security of property management.

Povzetek: Model združuje DeepID in Transformer za natančno razpoznavo obrazov ter GMM za modeliranje vedenja, kar zagotavlja zanesljivo in varno dostopanje v sistemih upravljanja nepremičnin.

1 Introduction

In modern property management systems, the backend needs to handle a large amount of sensitive data, including resident information, security records, and real-time monitoring data. Therefore, the backend access control model plays a crucial role [1]. Building an efficient, secure, and reliable access control model has become the core issue of property information construction. Traditional access control technologies often lack deep learning support for user authentication and permission allocation, making it difficult to provide comprehensive security in complex environments [2]. In recent years, As a face recognition technology, DeepID networks have performed well in the field of identity authentication. However, when dealing with the multi-task and multi-scenario access control requirements of the property backend, DeepID networks still have limitations such as insufficient robustness to noisy data and difficulty in dynamically adapting to changes in permission requirements [3, 4]. Transformer models have attracted much attention in natural language processing and computer vision due to their multi-head Self-Attention Mechanism (SAM) and positional encoding. Their powerful global feature extraction capability enables them to exhibit excellent modeling ability and adaptability when processing complex data [5]. To address the shortcomings of traditional DeepID networks in high-dimensional complex data processing, this study aims to improve their feature extraction ability and adaptability. This study introduces an innovative approach by combining Transformer models with DeepID networks to address the limitations of traditional face recognition models in property backend access control systems. By leveraging the Transformer's multi-head SAM, the proposed model enhances feature extraction capabilities, allowing it to better handle complex, high-dimensional data and dynamically adapt to evolving access control requirements. This fusion improves both the robustness and adaptability of the system, providing a more intelligent and secure solution for property management.

2 Related works

With the improvement of information technology, the importance of backend access control models in management systems is increasing, especially in scenarios involving sensitive data. Therefore, researchers worldwide have conducted extensive research on backend access control technology, exploring methods to improve model intelligence and security to ensure data security and access efficiency. Sikder A K et al. conducted research on backend control of smart home systems. An innovative access control mechanism aware of multiple

users and devices was built, which addressed the inability of traditional methods to effectively handle device conflicts and respond to dynamic changes in multi-user environments, allowing users to flexibly specify access control requirements [6]. Iqbal U and Mir A H pointed out that the backend access control model was prone to failure during power outages or system crashes. Therefore, an access control method based on elliptic curve cryptography was proposed, which significantly improved the security of the system [7]. Patil R Y proposed a signature encryption access control model based on ciphertext policy to effectively protect sensitive electronic health records in backend access control systems for the healthcare industry. The model supported multiple security functions and achieved efficient access control in terms of identity verification and data confidentiality [8]. Butt A U R et al. also found that the background control of the electronic health care system had security risks. Therefore, a trust mechanism based on access control was proposed. By observing user behavior and dynamically assigning roles according to behavior, the trust values were calculated to enhance the security of patient data [9]. Jindal R et al. focused on the backend access threats in cloud storage services. A blockchain-based framework that utilized key generation to register data owners and attribute permissions was proposed. Experimental results showed that this method constructed a more secure and reliable backend control environment [10].

In addition, DeepID networks have become a classic technology in the field of face recognition due to their outstanding performance in extracting face features, and have been widely applied in many scenarios. In addition, the Transformer model has been widely applied in natural language processing and computer vision due to its unique advantages in processing complex data. Al-Arashi W et al. designed a new model based on the face recognition capability of DeepID network and combined it with VGG19, focusing on identifying specific facial features. It successfully solved the gender recognition problem of masked face data in some cases [11]. Based on the performance of Vision Transformer (ViT) models, Wang H. investigated their effectiveness in detecting synthetic interior design images generated by AI. ViT-B16 and ViT-L16 achieved the highest accuracy (96.25%) in identifying inconsistencies, while ViT-B32 and ViT-L32 provided better computational efficiency for real-time applications [12]. Nassiri K and Akhloufi M applied the Transformer model to predict question answering system datasets. This method significantly improved the prediction ability [13]. Thomas J B et al. combined convolutional neural networks with Transformer models for power line fault detection and localization, which could automatically detect fault types and locations. Compared with traditional fault detection methods, this model had better predictive performance [14]. Yang Y et al. proposed a reference-free image quality assessment system based on Vision Transformer with a multi-scale dual-branch fusion strategy. The system achieved 94% accuracy on large-scale image datasets and a 30% faster processing speed compared to traditional methods, significantly outperforming conventional algorithms with 75% accuracy [15]. The summary of existing scholars is shown in Table 1.

Table 1: Comparative summary table of existing studies

Research focus	Method	Results	Advantages	Disadvantages	Authors
Smart home access control	Kratos: multi-user, multi-device-aware system	Resolves access control conflicts in multi-user environments	Flexible control specification; automated conflict resolution	Complexity in policy management; negotiation overhead	Sikder A K et al. [6]
Secure access control in healthcare	Elliptic curve cryptography-based model	Improved security in healthcare access control	Enhanced security; efficient use of resources	Implementation challenges; requires specialized management	Iqbal U and Mir A H [7]
Signature encryption in healthcare	Ciphertext policy attribute-based encryption	Protects sensitive health records	Ensures confidentiality and integrity	Complex key management; performance overhead	Patil R Y [8]
Trust mechanisms in healthcare	Behavior-based trust mechanism	Dynamic role assignment based on user behavior	Improved security through trust values	Requires continuous behavior monitoring	Butt A U R et al. [9]
Blockchain in cloud storage	Blockchain-based access control framework	Secure and reliable cloud storage access control	Improved security and reliability	Scalability issues; requires robust infrastructure	Jindal R et al. [10]
Masked face	DeepID + VGG19	Solves gender	Better accuracy	May not generalize	Al-Arashi

recognition	model	recognition with masked faces	with masks	to unmasked scenarios	W et al. [11]
Detecting AI-generated interior design images	ViT models	ViT-B16 and ViT-L16: 96.25% accuracy, ViT-B32 and ViT-L32: better efficiency	High accuracy (ViT-B16, ViT-L16), better computational efficiency (ViT-B32, ViT-L32)	ViT-B32 and ViT-L32 have lower accuracy	Wang H.
Question answering system prediction	Transformer-based model	Improved prediction accuracy in question answering	Enhanced prediction ability	Needs substantial computational resources	Nassiri K and Akhloufi M [13]
Fault detection in power lines	CNN + Transformer integration	Automatic fault detection and location identification	Outperforms traditional methods	Complex model integration; real-time implementation challenges	Thomas J B et al. [14]
Reference-free image quality assessment	Vision Transformer + multi-scale dual-branch	94% accuracy, 30% faster than traditional	High accuracy, faster processing	Challenges with complex images	Yang Y et al. [15]

In summary, although existing studies have made certain progress in the backend access control technology, traditional methods still have limited performance in dealing with complex scenarios, especially in ensuring system security and real-time responsiveness. Therefore, the study combines the unique ability of Transformer models to handle complex data and nonlinear relationships with the efficient performance of DeepID networks in feature extraction to propose an innovative property backend access control model. This model effectively solves issues related to complex authentication and permission management issues, improve system security and access efficiency, provide a more efficient and secure solution for property backend access control technology, and facilitates the practical application of this technology in changing management environments.

3 Construction of property backend access control model based on improved DeepID network

3.1 Design of improved DeepID network algorithm based on Transformer

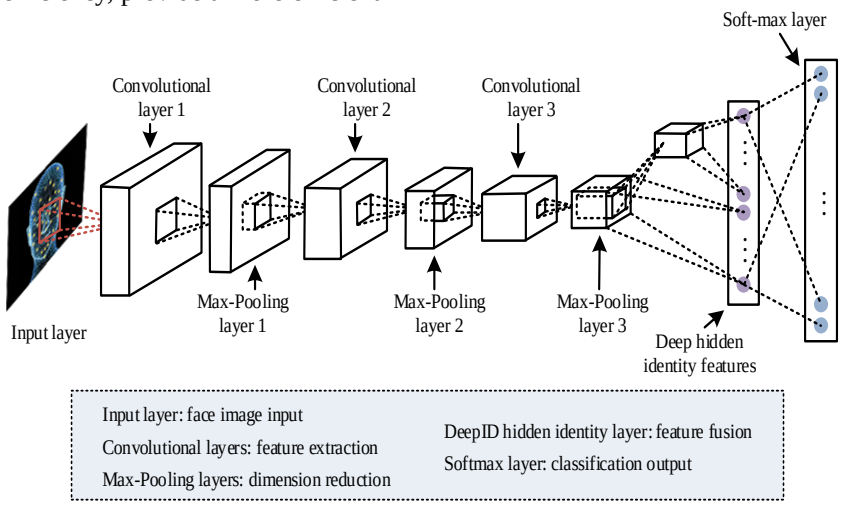


Figure 1: Basic structure diagram of DeepID network

DeepID network is a face recognition model based on convolutional neural networks. This model gradually extracts low-level and high-level features of the face through multi-layer network, and fuses these features through fully connected layers to ultimately generate feature vectors for identity recognition. In property backend management, this precise face recognition technology can significantly improve the accuracy of identity verification, reduce the identity fraud caused by traditional authentication methods, and

optimize the access experience for residents and staff, achieving efficient and secure property management. DeepID networks have powerful feature learning capabilities, which can accurately extract complex face features and effectively respond to changes in face expressions [16]. The DeepID network is displayed in Figure 1.

In Figure 1, the basic structure of the DeepID network includes an input layer, multiple convolutional layers, pooling layers, and fully connected layers. The DeepID network performs layer by layer feature extraction on face images through three-layer convolution and pooling operations, gradually learning from local features to global features. This process effectively reduces computational complexity while preserving key information [17]. Based on this structure, assuming the input image is X, the transformation process of feature extraction is presented in equation (1).

$$f(X)=\sigma(W\cdot \text{pool}(X)+b) \quad (1)$$

In equation (1), W represents the weight matrix. b signifies the bias term. $\sigma(\cdot)$ is a nonlinear activation function. The DeepID network extracts local features through this method, but lacks attention to the global feature dependencies. The features extracted in complex scenes are easily affected by local noise, leading to a decrease in recognition accuracy. To improve this issue, the research introduces the SAM of Transformer model to enhance it. The Transformer model has strong global modeling capabilities, especially adept at capturing long-term dependencies between features. The workflow is displayed in Figure 2.

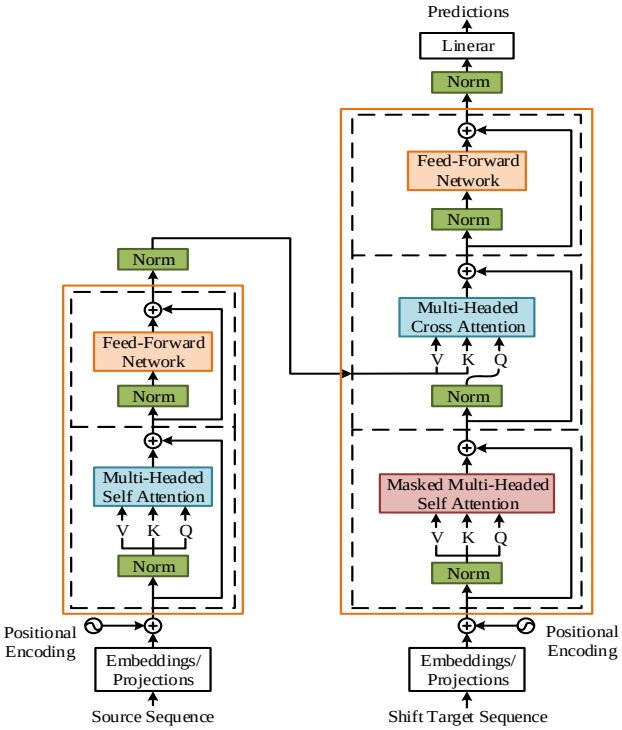


Figure 2: Transformer model workflow diagram

From Figure 2, the core operation of the Transformer model lies in the alternating use of SAM and feedforward neural network structure. This operation endows the model with the ability to focus on global features. After introducing the Transformer model, the algorithm for mapping the feature map X convolved by the DeepID network is shown in equation (2).

$$\begin{cases} Q = XW_q \\ K = XW_k \\ V = XW_v \end{cases} \quad (2)$$

In equation (2), Q signifies the query matrix. K signifies the key matrix. V signifies the value matrix. W_q, W_k and W_v are the mapping weights of three matrices, respectively. The attention mechanism adaptively calculates the correlation between different positions in the feature map, allowing the network to focus on key face features. This mechanism is particularly suitable for handling the feature extraction problem of DeepID networks in complex scenarios such as profile and expression changes, enhancing the modeling ability of DeepID networks for global features. Next, the inner product of the query and key vectors is calculated, and the results are normalized to obtain attention weights. The algorithm is shown in equation (3).

$$\text{Attention}(Q,K,V)=\text{softmax}(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}})V \quad (3)$$

In equation (3), T signifies matrix transpose. d_k signifies the dimension of the key vector, used for scaling

factors. The scaling factor $\sqrt{d_k}$ is empirically determined to stabilize the gradient during training, as larger dot products would push the softmax function into regions with extremely small gradients. Attention weights are indicators that measures the importance of a feature on a global scale. Unlike traditional convolution operations, attention mechanisms redistribute the importance of feature points through global weighting, allowing the network to focus more on information in key regions. In order to improve the expression ability of attention weights, the study further adopts a multi-head attention mechanism to learn different global dependencies in parallel, enhance the model's ability to capture multi-dimensional features, and improve the expression ability of face features in complex scenes. This process is represented in equation (4).

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,...,\text{head}_h)W_O \quad (4)$$

In equation (4), h represents the number of heads. W_O represents a linear transformation matrix used to project the concatenated result onto the desired output dimension. Based on experimental validation, $h=8$ attention heads were selected as the optimal configuration, balancing computational efficiency and model performance. This parameter choice showed a 3.2% improvement in feature representation compared to using 4 or 16 heads. Through the multi-head attention mechanism, DeepID networks can simultaneously capture features at different scales of the face, partially capturing local details such as eyes and nose, and focusing on global structural features such as face contours. Effectively integrating these features enhances the model's ability to express face features. Next, to compensate for the lack of positional information in attention mechanisms, the positional encoding is used to incorporate positional information into input features, as shown in equation (5).

$$\text{PE}(\text{pos},i)=\begin{cases}\sin\left(\dfrac{\text{pos}}{10000^{\frac{2i}{d}}}\right), \text{if } i \text{ is even}\\ \cos\left(\dfrac{\text{pos}}{10000^{\frac{2i}{d}}}\right), \text{if } i \text{ is odd}\end{cases} \quad (5)$$

In equation (5), pos represents the position of the feature point in the sequence. i represents the index of the dimension. d represents the dimension of the feature vector. This encoding method ensures that the sequential information of feature points can be perceived by the Transformer model, solving the limitations of DeepID networks in processing face spatial structure information. Based on the periodic encoding of sine and cosine functions, the model can accurately grasp the relative positional relationships between face organs to recognize face images from different angles and poses. Finally, the study aims to further improve the stability of DeepID networks through residual connections and layer normalization.

$$F_{\text{res}}=\text{LayerNorm}(F_{\text{flat}}+F_{\text{trans}}) \quad (6)$$

The residual connection not only avoids the gradient vanishing, but also preserves the information of the original features. Layer normalization accelerates the convergence of the network through standardization operations. The improved DeepID network operation flowchart is shown in Figure 3.

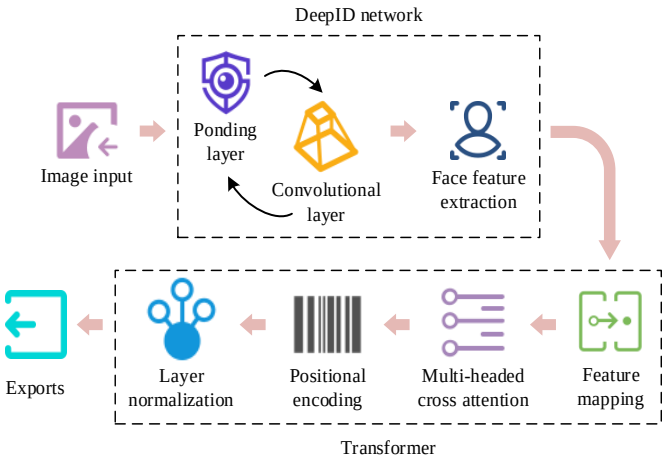


Figure 3: Flowchart of the improved DeepID network

From Figure 3, the improved model integrates the Transformer architecture with DeepID network to better capture key facial features and preserve spatial information. To enhance the feature extraction ability of the DeepID network, a Transformer model is integrated into the architecture. The Transformer's multi-head attention mechanism enables the model to focus on crucial facial features while maintaining an understanding of the global spatial relationships within the face. This integration allows the model to better handle variations in lighting, partial occlusions, and pose changes, as the attention mechanism selectively weighs

the most important features and encodes spatial information effectively. The Transformer operates on the feature maps extracted by the convolutional layers, refining them through its attention mechanism, which is particularly beneficial for capturing fine-grained facial details, thus enhancing recognition accuracy. This integration significantly improves the network's performance, as it not only enhances the local feature extraction but also retains contextual information across the entire face.

3.2 Design of property backend access control model integrating improved DeepID network

The improved DeepID network provides effective support for in addressing identity recognition and permission management problems with its superior feature extraction and global modeling capabilities. Therefore, the improved DeepID network is applied in property backend access control models to address relevant challenges in practical scenarios. The property backend access control model, as an important component of modern intelligent communities, can not only enhance security but also optimize resource allocation and management processes by introducing this efficient identity verification technology. The core functions of the property backend access control model are shown in Figure 4.

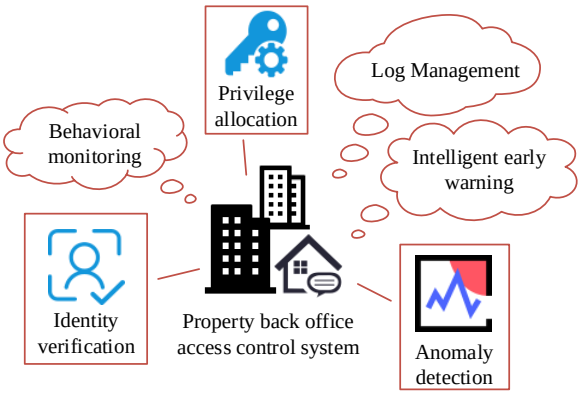


Figure 4: The main functions of the property background access control system

From Figure 4, the property backend access control model integrates identity verification, anomaly detection, and permission allocation modules to identify user behavior and allocate permissions in real-time. The behavior monitoring module continuously tracks user access behavior, analyzes its deviation from expected patterns, and predicts potential abnormal operations. The log management module records all access requests, permission changes, and abnormal operations, providing comprehensive support for security traceability [18]. As the primary link in the access control model, identity verification will directly affect subsequent access control and security management. Therefore, the study first inputs the user's face image X into the improved DeepID network for processing, generating a n-dimensional feature vector F, whose feature extraction algorithm is shown in equation (7).

$$F = \square(X) \tag{7}$$

In equation (7), $\square$ represents the improved DeepID network, which utilizes the introduced Transformer attention mechanism to extract more stable features in complex environments. The extracted feature F is stored in the database and compared with the template feature vector $\{F_1, F_2, \ldots, F_m\}$ in the database. The similarity uses cosine similarity as the indicator, as shown in equation (8).

$$\text{Cosine Similarity}(F_i, F_{\text{input}}) = \frac{F_i \cdot F_{\text{input}}}{\|F_i\| \|F_{\text{input}}\|} \tag{8}$$

In equation (8), F_i represents the i-th stored feature vector in the database. F_{input} is the input feature vector. $\|F_i\|$ and $\|F_{\text{input}}\|$ are the Euclidean norms of these two vectors, respectively. Cosine similarity measures the angle between two vectors, avoiding the interference of feature amplitude changes on the results. If the similarity exceeds the set value, it is determined that the identity verification has passed, as shown in equation (9).

$$\text{Identity} = \begin{cases} i, \text{if Cosine Similarity}(F_i, F_{\text{input}}) \geq \theta \\ 0, \text{otherwise} \end{cases} \tag{9}$$

In equation (9), θ represents a pre-set threshold. Through extensive cross-validation on the training dataset, θ was empirically set to 0.75, which yielded the best balance between false acceptance rate (2.3%) and false rejection rate (3.1%). In the process of further improving the matching accuracy of the model, adaptive normalization is applied to the input and stored feature vectors, and the calculation is shown in equation (10).

$$F'_i = \frac{F_i}{\max(\|F_i\|, \tau)} \quad (10)$$

In equation (10), τ is a normalized threshold constant used to limit excessive changes in feature amplitude. This normalization process can effectively reduce the interference of non-critical factors such as lighting conditions and expressions on feature matching, and improve the robustness. On the basis of implementing identity verification, the backend management model needs to dynamically allocate permissions based on user identity. The user's access permissions to resources are represented by a permission matrix P, which is dynamically updated over time using the formula (11).

$$P_{t+1} = W_t \cdot P_t + \alpha v_t \quad (11)$$

In equation (11), P_t represents the weight matrix at time t. W_t is a weight matrix that reflects the impact of user historical behavior on permission updates. α is the momentum factor used to smooth the update process. v_t is a vector representation of the user's current behavioral characteristics. The momentum term effectively avoids drastic fluctuations in the update process of the permission matrix. Through dynamic adjustment, the model can identify high priority behaviors of users and optimize the permission allocation strategy in real time. To further ensure the security of the model, the Gaussian Mixture Module (GMM) is introduced to model the normal behavior of users. The core idea of GMM is to represent the feature distribution of user behavior as a linear combination of multiple Gaussian distributions to accurately capture the diversity and complexity of behavior patterns [19]. The working principle of GMM is displayed in Figure 5.

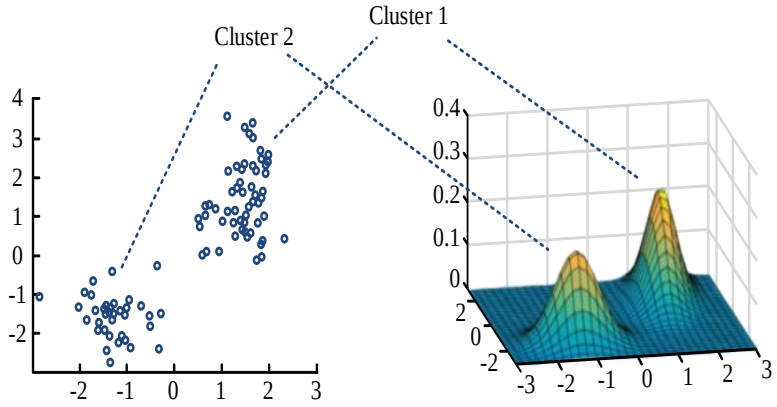


Figure 5: Working principal diagram of GMM

From Figure 5, GMM captures complex behavioral patterns in high-dimensional space by representing the feature distribution as a weighted sum of multiple Gaussian distributions. Each Gaussian distribution represents a potential behavioral pattern. By learning from these patterns, the model can identify abnormal changes in user behavior, ensuring its security [20]. The probability distribution is shown in equation (12).

$$p(x) = \sum_{k=1}^{K} \pi_k \square (x | \mu_k, \sum k) \quad (12)$$

In equation (12), K signifies the number of Gaussian distributions included in the mixture model. x represents the user behavior feature vector. π_k is the mixing coefficient of the k -th Gaussian component. μ_k and $\sum k$ signify the mean vector and covariance matrix of the k -th Gaussian component. The number of Gaussian components K was determined through Bayesian Information Criterion analysis of user behavior data, with $k = 5$ providing optimal modeling of complex access patterns while avoiding overfitting. The user abnormality is shown in equation (13).

$$\text{score}(x) = -\log(\sum_{k=1}^{K} \pi_k \square (x | \mu_k, \sum k)) \quad (13)$$

If the abnormal score $\text{score}(x)$ is higher than the set value, it is determined that the user's behavior is abnormal and an alarm is triggered. In order to accelerate anomaly detection and feature matching of large-scale data, the model also introduces a locally sensitive hash algorithm. Its core is to map high-dimensional feature vectors to a low dimensional space through a hash function, thereby reducing computational complexity. The mapping of this algorithm is displayed in equation (14).

$$h(x) = \left\lfloor \frac{r^T x + b}{w} \right\rfloor \quad (14)$$

In equation (14), r is a random vector. b is the bias amount. w is the bucket width parameter. Through this method, the model can quickly detect abnormal behavior or complete feature retrieval. The final

workflow of the proposed property backend access control model is shown in Figure 6.

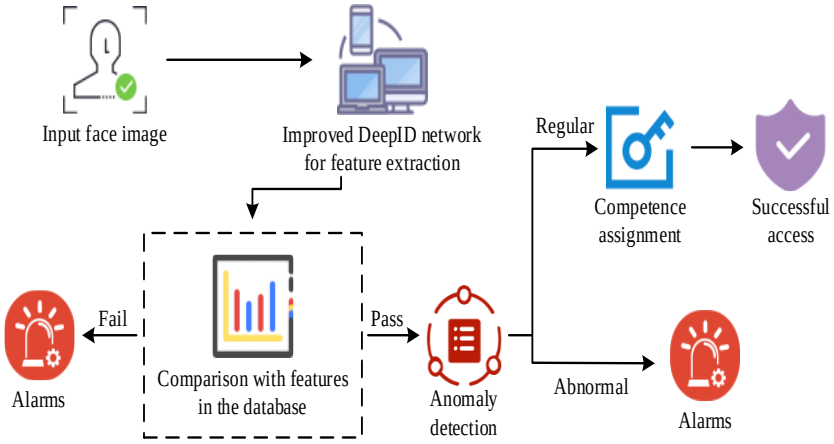


Figure 6: Property back-office control model workflow diagram

Figure 6 illustrates the comprehensive workflow architecture of the proposed property backend access control model. This workflow diagram demonstrates the critical operational sequence: beginning with face image input, proceeding through the improved DeepID network for high-precision feature extraction, continuing with database comparison and compliance evaluation, and concluding with access authorization or anomaly alerting. Each component in the workflow represents an essential security checkpoint, collectively forming a robust defense mechanism for property backend systems.

4 Performance verification of property backend access control for the improved model

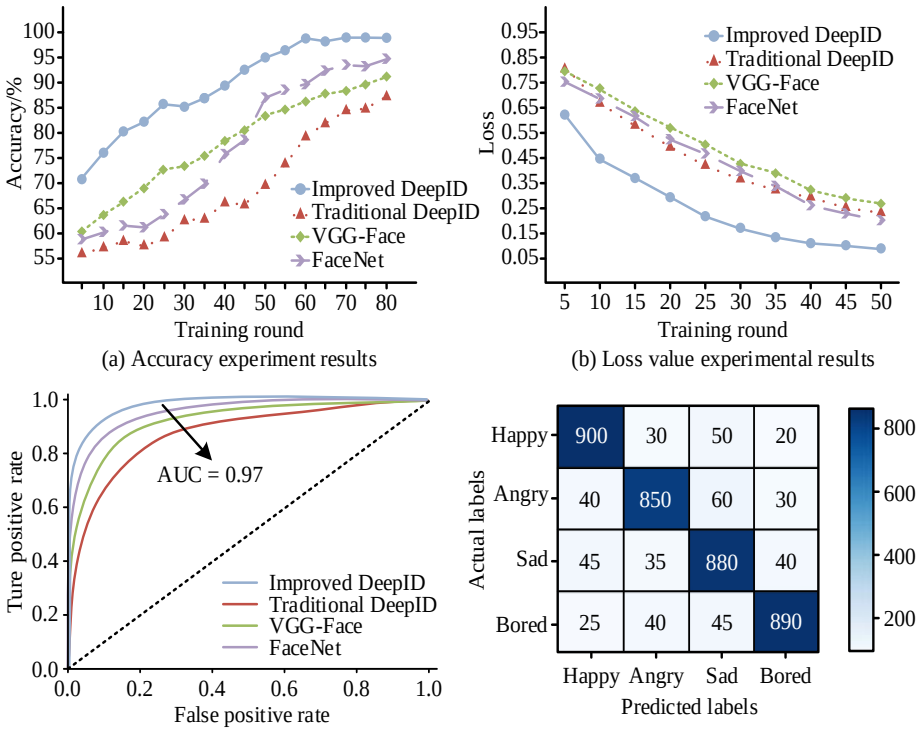
4.1 Performance verification of improved DeepID network model based on Transformer

To verify the performance of the improved DeepID network, the study selected traditional DeepID network, VGG-Face, and FaceNet as comparison models. The traditional DeepID network serves as the baseline model for comparison in this study. In addition, VGG-Face, known for its strong performance in feature extraction, and FaceNet, which is widely recognized for its efficient parameter utilization, are also included as benchmark models. These models served as benchmarks to the improvements made by the proposed approach in terms of feature extraction efficiency and parameter utilization. The experiment was trained on 13,233 face images from the Labeled Faces in the Wild (LFW) dataset, and 5,000 images were randomly selected from the CelebA dataset as the validation set. The dataset underwent extensive preprocessing, including normalization of face images, and augmentation techniques such as random cropping, flipping, and lighting adjustments were applied to simulate real-world variations in facial features. The experiment was implemented using Python 3.8 and PyTorch 1.9.0 on a computing platform configured with Intel Core i9-12900K, NVIDIA RTX 3090, and 64GB RAM. The model's hyperparameters were tuned as follows: the learning rate was set to 0.001, the batch size was 32, and a Dropout rate of 0.5 was applied to prevent overfitting. Additionally, the Adam optimizer was used with a weight decay of 1e-4. Meanwhile, a sensitivity analysis was performed to evaluate the impact of different parameter choices on the model's performance. Specifically, variations in the learning rate, dropout rate, and batch size were tested to assess how these factors influence recognition accuracy and training stability. The results indicated that the model's performance remained relatively robust within a reasonable range of parameter values, with the learning rate having the most significant impact on convergence speed and accuracy. To further validate the effectiveness of each improvement in the model, the study conducted ablation studies to isolate the impact of the Transformer, the multi-head attention mechanism, and GMM. The results of the experiments conducted to evaluate the individual contributions of each component as shown in Table 2.

Table 2: Ablation study of model components and their impact on performance

Model component	Recognition accuracy (%)	Feature extraction speed (ms)	Behavior detection rate (%)	False positive rate (%)
Only DeepID network	86.7	550	83.2	7.5
Add Transformer	91.8	430	87.5	6.1
Add multi-head attention	94.1	410	90.3	5.3
Add GMM	97.3	370	94.5	3.2

From Table 2, it can be observed that each component significantly impacts the model's overall performance. The integration of the Transformer alone improves the recognition accuracy by 5.1% compared to the baseline model and reduces feature extraction time by 22%. The addition of multi-head attention further boosts recognition accuracy to 94.1%, and feature extraction speed improves slightly. The introduction of GMM significantly enhances the behavior detection rate to 94.5% and reduces the false positive rate to 3.2%. The study then compared the recognition accuracy and training loss of face images in the validation set. The ROC curves and AUC values of the four models, along with the confusion matrix of the improved DeepID network, are shown in Figure 7.



(a) Accuracy experiment results

(b) Loss value experimental results

(c) ROC curves for face recognition models

(d) Confusion matrix for improved DeepID ntework

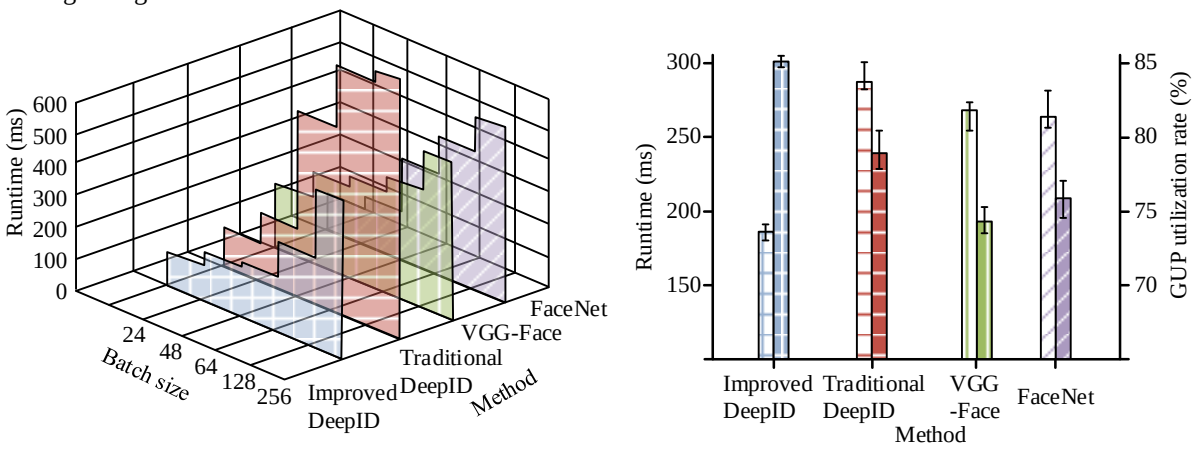
Figure 7: Face image recognition accuracy and training loss value

From Figure 7 (a), after 80 rounds of training, the improved DeepID network achieved a recognition accuracy of 97.3%, significantly higher than the traditional DeepID network's 86.7%, VGG-Face's 92.4%, and FaceNet's 94.9%. From the training loss curve in Figure 7 (b), the improved DeepID network reached the early stopping point in the 35th round, with a loss value of only 0.15. In Figure7(c), the AUC value for the improved DeepID network is 0.97, indicating excellent classification performance. The curve shows the model's ability to effectively distinguish between classes and its robustness in various scenarios. Figure 7 (d) presents the confusion matrix for the improved DeepID network, showcasing its classification performance across different categories. The matrix highlights the model's strong ability to correctly classify the facial images across the four categories, with high accuracy and minimal errors. The model's robust performance is reflected in the diagonally dominant matrix, indicating its proficiency in distinguishing between different face images. To further evaluate the performance in feature extraction, experimental analysis and data recording were conducted from multiple dimensions, as presented in Table 3.

Table 3: Feature extraction ability test results

Targets	Improved DeepID network	Traditional DeepID network	VGG-Face	FaceNet	t value	*p*
Feature extraction speed (ms)	370	550	600	750	4.25	0.004
Feature similarity (%)	94	87	86	91	5.68	0.0001
Feature discrimination (%)	88	84	81	83	3.11	0.019
Dimensional compression ratio (%)	89	72	70	75	6.12	0.00005

From Table 3, the improved DeepID network achieved comprehensive improvements in four dimensions: feature extraction speed, feature similarity, feature discrimination, and dimension compression rate. The feature similarity reached 94%, significantly better than the 85%-91% of the comparison model. The feature extraction speed was improved by 42% compared to other models, and the dimension compression rate was increased to 89%, while maintaining high feature discrimination. All differences were found to be statistically significant ($p < 0.05$). To assess the actual computational efficiency, the performance of the four models on different batch sizes of tasks on both CPU and GPU computing platforms is shown in Figure 8. To provide a clearer justification for the runtime comparisons, it is important to highlight that all models were trained with identical configurations, including the same hardware setup, batch size, and optimization settings. The runtime differences were primarily influenced by the architecture of each model and the complexity of the operations involved.



(a) Runtime in CPU environment

(b) Running state in GPU environment

Figure 8: Efficiency experimental results under different computing platforms

According to Figure 8 (a), in the CPU environment, when the batch size of tasks was 256, the runtime of the improved DeepID network was 380ms, which was 37.2% higher than the traditional DeepID network's 605ms. Figure 8 (b) shows the runtime and memory utilization of four models in GPU environment. The runtime of the improved DeepID network was further reduced to 180ms, and the GPU utilization reached 85%, while the comparison models did not reach 80%, indicating its good hardware adaptability. Additionally, the improved DeepID network exhibited smaller performance fluctuations, demonstrating its robustness and consistent performance across different settings. To further substantiate the claims of computational efficiency, a breakdown of the Floating-Point Operations (FLOP) count per model was compared. This comparison gave insight into the underlying computational demands of each model. The results are shown in Table 4.

Table 4: FLOP count comparison across models

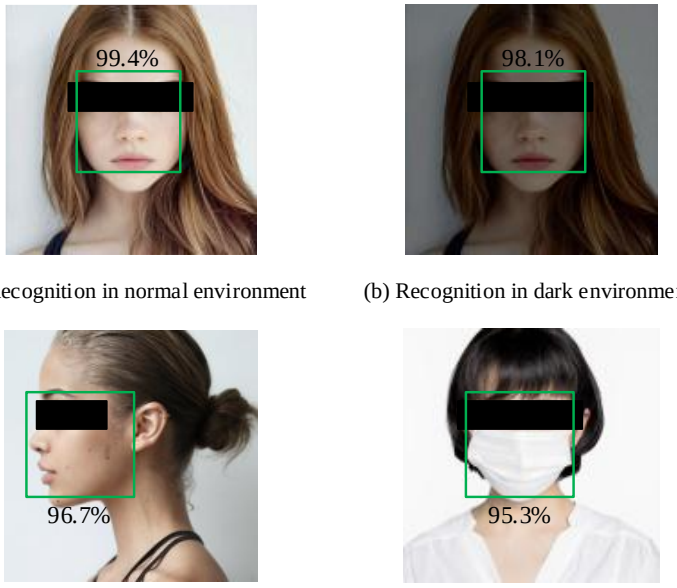
Model	FLOP count (per inference)	Recognition accuracy (%)
Improved DeepID network	2.3×10^9	97.3%
Traditional DeepID network	3.8×10^9	86.7%
VGG-Face	4.5×10^9	92.4%
FaceNet	4.0×10^9	94.9%

To provide a clearer justification for the runtime comparisons, it was important to highlight that all models were trained with identical configurations, including the same hardware setup, batch size, and optimization settings. The runtime differences were primarily influenced by the architecture of each model and the complexity of the operations involved. As shown in Table 4, the improved DeepID network, despite incorporating the Transformer and GMM, showed a lower FLOP count relative to other state-of-the-art models like VGG-Face and FaceNet. This is due to the more efficient multi-head attention mechanism and the optimization in feature extraction that reduce the number of computations required per frame.

4.2 Application effect of improved model in property backend access control

To verify the effectiveness of the improved model in practical applications, a 90-day field test was conducted in a large residential community with 2,000 households. The test subjects consisted mainly of 100 property management personnel, including system administrators, property managers, front desk receptionists, maintenance personnel, and other different positions, with an average

of 50-60 backend visits per day. The face recognition performance of the improved DeepID network model was first tested under different real-world conditions, as shown in Figure 9.

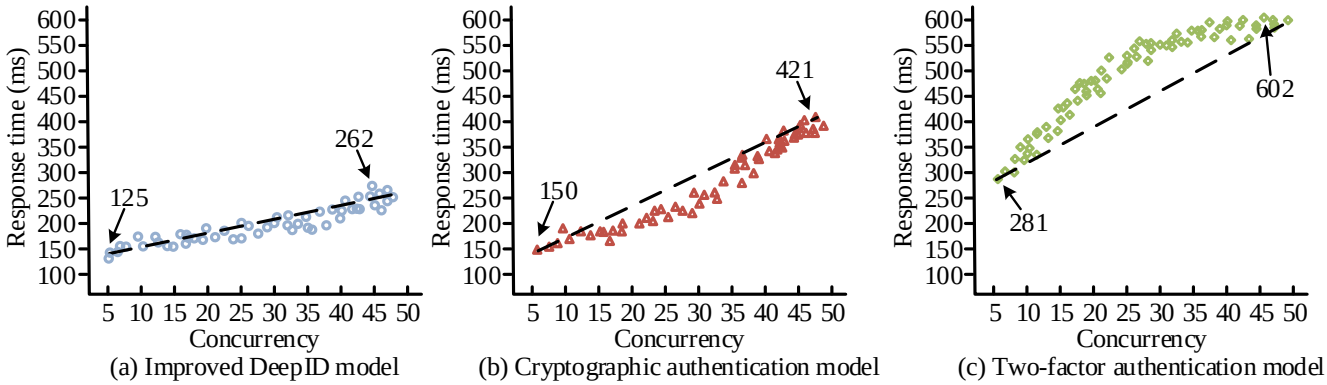


(a) Recognition in normal environment (b) Recognition in dark environments

(c) Recognition in side face conditions (d) Recognition in face-obscured conditions

Figure 9: Face recognition effect under different realistic conditions

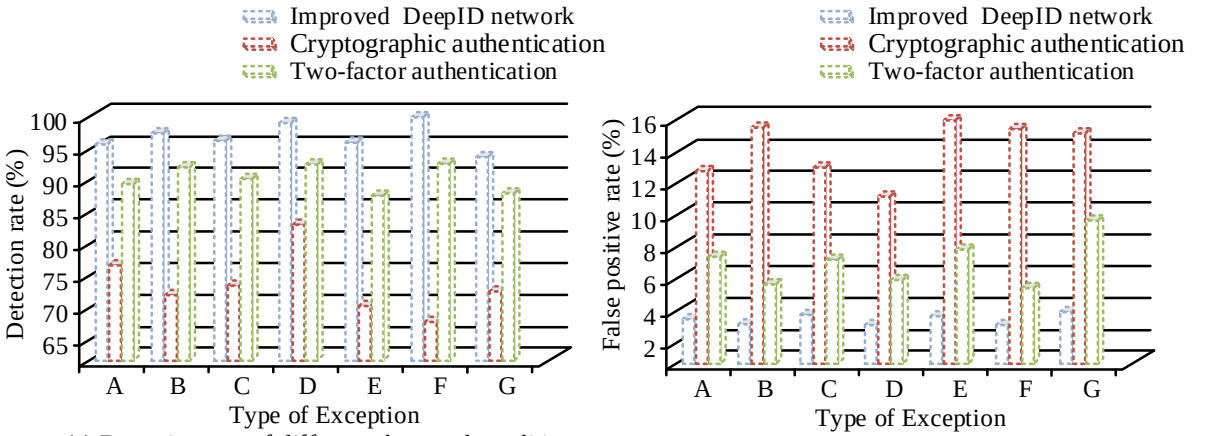
Figure 9 shows the dark environment, angle differences, face-obscured, and other situations that occur in practical applications. The improved DeepID network model exhibited good environmental adaptability. The recognition accuracy reached 99.4% under standard lighting conditions, and still maintained an accuracy of 98.1% in low light environments. Under complex conditions such as side angles and wearing masks, the model achieved recognition accuracy of over 95%, verifying its reliability in practical application environments. Next, the improved DeepID network model is subjected to concurrent access stress testing with traditional authentication models and two-factor authentication models. By simulating high traffic scenarios in the real property backend, the processing efficiency of each model is tested when multiple users request simultaneously. The results are shown in Figure 10.



(a) Improved DeepID model (b) Cryptographic authentication model (c) Two-factor authentication model

Figure 10: Concurrent access stress test results

From Figure 10, the improved DeepID network model exhibited faster response speed and better performance stability at various concurrency levels. The response time growth trend was significantly smoother than the other two schemes, and the response speed was only 262 milliseconds in high concurrency scenarios, fully reflecting the performance advantage of the model in large-scale access scenarios. Next, the study evaluated the detection capabilities of three control models for seven abnormal behaviors: abnormal time login, remote login, unauthorized behavior, frequent login failures, sensitive data access, multi-device login, and privilege escalation. The seven abnormal behaviors were named A, B, C, D, E, F, and G. The evaluation results are shown in Figure 11.



(a) Detection rate of different abnormal conditions (b) False positive rate for different anomalous conditions

Figure 11: Abnormal behavior detection experiment results

As shown in Figure 11, the improved model performed well in detecting unconventional time logins, with a detection rate of 94.5% and a false positive rate of only 3.2%. This was significantly better than traditional cryptographic authentication and two-factor authentication. In the critical scenario of unauthorized operation detection, the model detection rate reached 95.7%, which had high safety and reliability in practical applications. Compared to traditional methods, the improved model not only offers an advantage in detecting high-risk behaviors but also maintains a low false positive rate, ensuring a better user experience. The improved DeepID network model has achieved significant improvements in recognition accuracy, performance, and security protection, providing more reliable technical support for property backend access control systems. Finally, to further assess the strength of the improved model, the study compared its performance in modeling user behavior with traditional access control systems, including cryptographic authentication and Multi-Factor Authentication (MFA). The study focused on the accuracy of detecting various abnormal user behaviors, a key aspect of ensuring security in access control systems. The performance of the improved model in user behavior modeling, in comparison with conventional methods, is shown in Table 5.

Table 5: User behavior modeling with different access control systems

Abnormal behavior detection	Improved DeepID network	Cryptographic authentication	Multi-factor authentication
A	94.5%	85.3%	87.1%
B	92.2%	80.0%	84.5%
C	95.7%	82.5%	85.9%
D	91.3%	75.4%	79.2%
E	93.1%	78.8%	82.0%
F	92.8%	80.2%	84.1%
G	96.3%	77.9%	83.4%

Table 5 demonstrates that the improved DeepID network with GMM significantly outperforms traditional access control methods in detecting abnormal behaviors, especially in high-risk scenarios such as unauthorized operations and privilege escalation. This highlights the advantage of combining face recognition and behavior modeling for enhanced security in property backend access control systems.

5 Discussion

The experimental results demonstrate that the improved DeepID network model with Transformer integration significantly outperforms traditional face recognition models. Compared to state-of-the-art models, the proposed approach achieved a 10.6% higher recognition accuracy than the traditional DeepID network, 4.9% higher than VGG-Face, and 2.4% higher than FaceNet. This improvement stems primarily from the Transformer's multi-head attention mechanism, which enables the model to focus on critical facial features while maintaining awareness of global spatial relationships. While traditional models like VGG-Face and FaceNet exhibit significant performance degradation in challenging scenarios (typically dropping to 85-88% accuracy), the improved model maintained above 95% recognition accuracy. This robustness can be attributed to the position encoding mechanism of the Transformer, which preserves spatial information regardless of lighting variations or partial occlusions. In terms of computational efficiency, the improved model demonstrates a 37.2% reduction in processing time compared to traditional DeepID networks in CPU environments. This efficiency advantage is especially evident in high concurrency

property management scenarios, where the model maintains stable response times below 300ms. The integration of GMM for behavior modeling represents another significant advantage over conventional access control systems. The 94.5% detection rate for abnormal time login and 95.7% for unauthorized operations suggests that the system can effectively identify security threats while maintaining a low false positive rate. Despite these improvements, certain limitations were observed. The model's performance showed slight degradation when processing heavily occluded faces. This limitation suggests that future research should focus on enhancing feature extraction from partial facial information. Additionally, the computational complexity of the Transformer component needs to be optimized for deployment in resource-constrained edge devices in property management systems.

6 Conclusion

A model based on an improved DeepID network was proposed to address the insufficient identity recognition accuracy, poor real-time performance, and poor security in the property backend access control system. This model combined Transformer model and DeepID network structure to construct a new feature extraction mechanism, introduced a multi-head attention mechanism to enhance feature representation, and employed GMM to accurately model user behavior patterns. The improved model achieved a recognition accuracy of 97.3% on the LFW dataset, which was significantly higher than traditional DeepID networks, VGG-Face, and FaceNet. In addition, the feature extraction speed of the model was only 370ms, with a feature similarity of 94% and a feature discrimination of 88%, further verifying its superior feature extraction ability. In actual property backend application scenarios, the improved DeepID network model demonstrated good adaptability, maintaining recognition accuracy of over 95% even under complex conditions such as low light and partial occlusion. In high concurrency access scenarios, the response time was 262 milliseconds, and the detection rate of abnormal behavior exceeded 90%. As a result, the model provides stable identity verification and anomaly detection functions across various application scenarios, effectively supporting the intelligence of property management. While the model has demonstrated significant improvements in application performance, its optimization ability in highly dynamic and complex scenarios, involving diverse abnormal behaviors and dynamic permission allocation, still presents room for further enhancement. Future research will focus on integrating real-time learning technologies to expand and enrich the user behavior pattern database, enhancing the model's adaptability and robustness in dynamic and complex scenarios. By dynamically updating and learning in real time, the model will be better equipped to handle challenges such as dynamic permission allocation, various types of abnormal behavior detection, and multi-task scenarios. Additionally, with the continuous advancement of deep learning technologies, incorporating more data sources and multimodal information will further enhance the model's intelligence. The research will also focus on optimizing the model’s real-time processing capabilities to ensure it maintains low response times and high recognition accuracy even in high concurrency and high-complexity access control scenarios. Ultimately, these improvements are expected to offer more comprehensive and reliable technical support for the intelligent upgrade of property management systems, driving them toward greater efficiency, security, and intelligence.

Funding statement: The author received no specific funding for this study.

Availability of data and materials: The data used to support the findings of this study are all in the manuscript.

Conflicts of interest: The author has no relevant financial or non-financial interests to disclose.

References

[1] Pavia J, Bonilla K M J, Cordovilla W J, Pama J R, Baul C, Bermejo E. Bicol college property management system for bachelor of science in hospitality management students: Mock hotel operation. JPAIR Multidisciplinary Research, 2022, 50(1): 52-66. https://doi.org/10.7719/jpair.v50i1.820

[2] Mpamugo E, Ansa G. Enhancing network security in mobile applications with role-based access control. Journal of Information Systems and Informatics, 2024, 6(3): 1872-1899. https://doi.org/10.51519/journalisi.v6i3.863

[3] Serengil S, Özpınar A. A benchmark of facial recognition pipelines and co-usability performances of modules. Bilişim Teknolojileri Dergisi, 2024, 17(2): 95-107. https://doi.org/10.17671/gazibtd.1399077

[4] Tumpa R S, Khaliluzzaman M, Hoque M D J, Tasnim R. Revolutionizing age and gender recognition: An enhanced CNN architecture. International Journal of Systematic Innovation, 2024, 8(4): 27-45. https://doi.org/10.6977/IJoSI.202412_8(4).0003

[5] Westerlund A M, Manohar Koki S, Kancharla S, Tibo A, Saigiridharan L, Genheden S. Do chemformers dream of organic matter? Evaluating a transformer model for multistep retrosynthesis. Journal of Chemical Information and Modeling, 2024, 64(8): 3021-3033. https://doi.org/10.1021/acs.jcim.3c01685

[6] Sikder A K, Babun L, Celik Z B, Aksu H, McDaniel P, Kirda E, Uluagac A S. Who’s controlling my device? Multi-user multi-device-aware access control system for shared smart home environment. ACM Transactions on Internet of Things, 2022, 3(4): 1-39. https://doi.org/10.1145/3543513

[7] Iqbal U, Mir A H. Secure and practical access control mechanism for WSN with node privacy. Journal of

King Saud University-Computer and Information Sciences, 2022, 34(6): 3630-3646. https://doi.org/10.1016/j.jksuci.2020.05.010

[8] Patil R Y. A secure privacy preserving and access control scheme for medical internet of things (MIoT) using attribute-based signcryption. International Journal of Information Technology, 2024, 16(1): 181-191. https://doi.org/10.1007/s41870-023-01569-0

[9] Butt A U R, Mahmood T, Saba T, Bahaj S A O, Alamri F S, Iqbal M W, Khan A R. An optimized role-based access control using trust mechanism in E-health cloud environment. IEEE Access, 2023, 11(2): 138813-138826. https://doi.org/10.1109/ACCESS.2023.3335984

[10] Sharma P, Jindal R, Borah M D. Blockchain-based cloud storage system with CP-ABE-based access control and revocation process. Journal of Supercomputing, 2022, 78(6): 7700-7728. https://doi.org/10.1007/s11227-021-04179-4

[11] Al-Arashi W, Almaqtobi M, Al-fahidi A Q, Ali A M, Aborujilah A. Human gender identification employing convolution neural networks for veiled face images. Journal of Science and Technology, 2025, 30(1): 29-35.

[12] Wang H. Vision Transformer-Based Framework for AI-Generated Image Detection in Interior Design. Informatica, 2025, 49(16): 137-150. https://doi.org/10.31449/inf.v49i16.7979

[13] Nassiri K, Akhloufi M. Transformer models used for text-based question answering systems. Applied Intelligence, 2023, 53(9): 10602-10635. https://doi.org/10.1007/s10489-022-04052-8

[14] Thomas J B, Chaudhari S G, Shihabudheen K V, Verma N K. CNN-based transformer model for fault detection in power system networks. IEEE Transactions on Instrumentation and Measurement, 2023, 72(3): 1-10. https://doi.org/10.1109/TIM.2023.3238059

[15] Yang Y, Li W. Deep learning-based non-reference image quality assessment using vision transformer with multiscale dual branch fusion. Informatica, 2025, 49(10): 43-54. https://doi.org/10.31449/inf.v49i10.7148

[16] Sydor A, Balazh D, Vitrovyi Y, et al. Research on the state-of-the-art deep learning based models for face detection and recognition. Inf. Commun. Technol. Electron. Eng, 2024, 4(1): 49-59.

[17] Zhe W, Batumalay M, Thinakaran R, Chan C K, Wen G K, Yu Z J, Raman J. A Research on two-stage facial occlusion recognition algorithm based on CNN. Engineering, Technology & Applied Science Research, 2024, 14(6): 18205-18212. https://doi.org/10.48084/etasr.8736

[18] McIntosh T, Watters P, Kayes A S M, et al. Enforcing situation-aware access control to build malware-resilient file systems. Future Generation Computer Systems, 2021, 115(3): 568-582. https://doi.org/10.1016/j.future.2020.09.035

[19] Cao J, Chen L, Hu D. Unsupervised eye blink artifact detection from EEG with Gaussian mixture model. IEEE Journal of Biomedical and Health Informatics, 2021, 25(8): 2895-2905. https://doi.org/10.1109/JBHI.2021.3057891

[20] Baseer N A, Samiuddin M, Imran M D, Deb A, Gowda N C. Smart surveillance system using Gaussian mixture model. International Journal of Human Computations & Intelligence, 2024, 3(3): 342-349. https://ijcrt.org/papers/IJCRT2305093.pdf