IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 6, December, 2025, pp. 3227-3236
Journal homepage: http://iieta.org/journals/ts

# A Hybrid Image Analysis Model for Credit Card Transaction Monitoring and Fraud Detection

Yanmin Nie[1,2*], Dan Yao[1,2], Weiming Zhan[1,2]

[1] School of Information and Artificial Intelligence, Hebei Finance University, Baoding 071051, China
[2] Hebei Key Laboratory of Financial Technology and Application, Baoding 071051, China

Corresponding Author Email: computer_228@163.com



https://doi.org/10.18280/ts.420615

**Received:** 29 March 2025
**Revised:** 2 October 2025
**Accepted:** 28 October 2025
**Available online:** 31 December 2025

***Keywords:***
*credit card fraud detection, image analysis, facial recognition, occlusion recovery, stacked denoising autoencoder, mapping autoencoder, deep learning*

**ABSTRACT**

Credit card fraud, particularly identity theft and unauthorized transactions, continues to pose a significant threat to financial security. Traditional rule-based and numerical feature-based risk control systems struggle to address the increasingly sophisticated fraud techniques, primarily due to their inability to leverage the rich visual context present during transactions. The introduction of image analysis techniques into transaction monitoring provides a revolutionary approach to solving this problem by directly verifying the operator's identity and environmental security. However, fraudsters often use active obfuscation to evade facial recognition, leading to a sharp decline in the performance of existing visual algorithms in critical tasks. This challenge has become the primary technological bottleneck for reliable image-based risk control. While current facial recognition technologies perform excellently in controlled environments, their robustness is significantly lacking in real-world financial transaction scenarios involving strong occlusions and uncooperative subjects. Most existing methods either rely on complete visible facial features or require a large number of occlusion samples for training, making them ill-suited to adapt to unknown and ever-changing real-world fraudulent obstructions. This paper proposes a collaborative hybrid model of "Occlusion Detection - Iterative Restoration - Robust Recognition." Compared to mainstream methods such as GAN-Based, Vision Transformer, Partial Conv, and DNN selected in the experiments, the core innovations of this paper are reflected in three aspects: (1) Unlike the GAN-Based model, which generates visually natural but potentially identity-distorting restoration results, the synergy of the mapping autoencoder and iterative restoration mechanism in this paper achieves recognition accuracies of 96.8%, 97.5%, and 89.3% in the mask, sunglasses, and combined occlusion scenarios, respectively, significantly outperforming the 92.5%, 90.2%, and 78.6% of the GAN-Based method; (2) Unlike the Vision Transformer, which requires large amounts of training data and experiences significant performance degradation under severe occlusion, the method in this paper maintains 93% accuracy under 50% eye occlusion, far surpassing the 70.5% of Vision Transformer, without the need for massive labeled data; (3) Compared to Partial Conv, which lacks targeted restoration guidance, and DNN, which overly depends on global features, this paper maintains 89.7% accuracy under severely distorted image quality through a "precise occlusion detection - feature-preserving restoration - weight-shared recognition" closed-loop design, significantly outperforming Partial Conv's 76.8% and DNN's 68.9%. Experiments show that the method in this paper has superior robustness and practicality in complex transaction scenarios, effectively solving the issues of false rejections and fraud misjudgment in intelligent risk control.

## 1. INTRODUCTION

With the rapid development of digital payment technologies [1-4], credit cards have become an indispensable financial tool in modern society [5, 6]. However, their widespread use has also brought about increasingly severe transaction fraud risks [7-9], with "identity theft" [10] and "unauthorized transactions" [11] posing the most significant threats. Traditional fraud detection systems heavily rely on structured data such as transaction amount, time, and location. Although some success has been achieved, the limitations of these systems have become more apparent in the face of constantly evolving fraud techniques and increasingly organized fraudulent activities [12, 13]. These methods struggle to capture the rich contextual information present during transactions and cannot effectively address on-site frauds that use physical obstructions for disguise, facing significant challenges in both real-time performance and accuracy.

In this context, image analysis-based monitoring technologies [14, 15] have provided a new perspective for enhancing transaction security. By analyzing visual data in the transaction scene, the system can directly address two core

questions: "Is the current operator the cardholder?" and "Is the transaction environment secure?" However, applying visual technologies [16-18] in the challenging financial anti-fraud scene presents unique challenges. Among them, fraudsters actively use facial obstructions to evade recognition, which is the primary obstacle to the reliability of facial recognition systems. Under strong occlusion conditions, the loss of critical identity biometric features [19, 20] can directly lead to recognition failure, either causing a legitimate user to be rejected or a fraudulent transaction to be mistakenly approved, thereby severely weakening the effectiveness of the entire risk control system.

To address this core issue, this paper proposes an innovative image analysis method for credit card transaction monitoring and fraud detection. The core of this research is a hybrid model specifically designed to address occlusion, systematically integrating three key technological modules. First, we introduce an unsupervised mapping autoencoder aimed at precisely locating unknown types and locations of occluded areas, without requiring any prior occlusion samples. Second, we develop an iterative stacked denoising autoencoder that, based on the location information, progressively and high-quality restores the occluded parts while strictly preserving the original biometric features of the unobstructed areas. Finally, we use a deep neural network for ultimate identity verification, and through a pre-training weight-sharing strategy, we ensure high synergy between feature extraction and identity recognition.

The core uniqueness of the method in this paper lies in its deep adaptation to the requirements of financial fraud detection scenarios. The key differences and innovations compared to the mainstream methods in the experiments are as follows:

(1) Innovation of the unsupervised occlusion detection mechanism: Methods such as Partial Conv, compared in the experiments, rely on occlusion region labeled training, limiting their generalization ability. In contrast, the mapping autoencoder in this paper uses an unsupervised paradigm of "learning anomalies from normal," allowing precise localization of various unknown occlusions such as masks, sunglasses, and combined occlusions without occlusion label data, relying only on clean face data. Even with 50% occlusion in random regions, the recognition accuracy remains 89.5%, significantly outperforming Partial Conv's 67.5% and 76.8%, and is well adapted to the diverse occlusion forms in transaction scenarios.

(2) Scenario-based innovation of iterative restoration strategy: While the GAN-Based method achieves 90% accuracy at 50% mouth occlusion, which is close to the 89.5% of this paper's method, there is a significant performance gap in mid-to-high occlusion ranges and combined occlusion scenarios. Traditional DNN methods, lacking a restoration mechanism, have accuracy below 80% in various occlusion scenarios. In contrast, the iterative stack denoising autoencoder in this paper, guided by the occlusion probability map, restores only the occluded regions and preserves the original clean features. This method maintains stable performance under different occlusion positions, areas, and image quality conditions, perfectly aligning with the stringent feature fidelity requirements for financial identity verification.

(3) Collaborative innovation of cross-module weight sharing: In the experiments, methods like GAN-Based and Vision Transformer have independent restoration and recognition modules, resulting in insufficient identity feature matching accuracy in the restored regions. In this paper, the recognition module shares the encoder weights of the restoration module, enabling the recognition network to inherit prior knowledge of "normal face features." This allows the method to achieve a balance of 98.7% accuracy and 35 ms processing time at 1920×1080 resolution, which is faster than Vision Transformer and more accurate than DNN (92.1% / 15 ms).

This paper aims to elaborate on the complete framework and algorithmic principles of this method and to experimentally validate its excellent performance under simulated real-world fraud scenarios. We firmly believe that this research not only provides an effective technological path to overcoming facial recognition challenges under occlusion but also lays a solid theoretical foundation for building the next generation of intelligent, proactive, and highly robust credit card anti-fraud systems.

## 2. PROBLEM DESCRIPTION

In the image analysis-based credit card transaction monitoring and fraud detection framework, a core challenge lies in the uncontrollability of real-world scenarios. Legitimate cardholders may engage in transactions while wearing masks, scarves, or hats, or in environments with poor lighting and partial shadows. Fraudsters, on the other hand, actively use sunglasses, helmets, or deliberate hand coverings to evade facial recognition systems. Therefore, the core scientific problem precisely defined in this paper is: how to achieve robust and accurate identity verification in highly occluded and noisy non-cooperative transaction scenarios to effectively distinguish legitimate cardholders from identity thieves. Traditional facial recognition models perform excellently on complete, clear facial images, but their performance dramatically declines under occlusion conditions with significant missing facial information because they heavily rely on the consistency of global features. This vulnerability directly threatens the reliability of the risk control system, which either overly rejects legitimate users due to the inability to recognize them, severely impacting user experience, or, if deceived, allows fraudulent transactions, leading to direct economic losses. Therefore, solving the occlusion problem is not only a technical requirement to improve recognition rates but also a prerequisite for building a truly practical, image analysis-based intelligent anti-fraud system.

To address this issue, the solution proposed in this paper is a collaborative framework consisting of "precise localization," "iterative repair," and "robust recognition," and based on this, an algorithmic combination centered around an iterative stacked denoising autoencoder is chosen. The innovation of this approach lies in that it does not attempt to forcibly extract features from incomplete occluded faces but instead restores reliable identity information through systematic reconstruction. First, to solve the unknown occlusion problem, a mapping autoencoder is introduced, which is designed to learn the residual mapping from an occluded face to a complete face. This network can accurately locate the spatial position of any unknown occlusion, effectively providing a precise "occlusion area map" for the subsequent repair process. Second, in the repair phase, an iterative stacked denoising autoencoder is used to progressively restore facial textures. Unlike single-pass repair models, this iterative structure can gradually refine the repair results through a series of

autoencoders, where each iteration builds on the output of the previous one to further enhance the restoration. This method maximizes the consistency of the original identity information, effectively avoiding the identity distortion or semantic errors that may arise from one-time generation, which is crucial for ensuring the accuracy of the subsequent recognition step. Finally, the high-quality repaired face is fed into a deep neural network for recognition. The reason for selecting this complete algorithmic framework is that it closely aligns with the business logic of financial anti-fraud, breaking down the occlusion problem into manageable sub-tasks; the mapping autoencoder equips the system with strong generalization capabilities to handle novel fraudulent occlusion techniques; and the iterative repair mechanism ensures the fidelity of biometric feature reconstruction, providing a reliable basis for risk control decisions. This algorithm ultimately empowers the system to penetrate occlusions and accurately answer the core security question of "Is the operator the legitimate cardholder?" even in the case of fraudsters deliberately disguising themselves, thus greatly enhancing the practical ability of image analysis-based fraud detection.

## 3. METHOD INTRODUCTION

### 3.1 Model overview

In the intelligent risk control system designed for credit card transaction monitoring and fraud detection in this paper, the facial recognition model is given the core mission of overcoming the occlusion challenges in the real world. The proposed model is based on a fundamental formal definition: an occluded facial image $A_{pzz}$, captured from an ATM or POS machine monitoring camera, is considered a degraded version of the cardholder's original clean face $A$ combined with an unknown, fraud-intended additional noise $r$, i.e., $A_{pzz}$ = A + r. Here, the noise r is not random interference but is embodied in the form of sunglasses, masks, helmets used by fraudsters to disguise their identity, or lighting shadows created by the environment, etc. The key characteristics of this noise are "unknown" and "malicious," with the goal of systematically destroying the biometric features used for identity recognition. Therefore, the primary task of this model is to find a robust repair function $d_{\Phi}(\cdot)$, aimed at restoring a trustworthy identity visual representation $d_{\Phi}(A_{pzz})$ from the contaminated signal $A_{pzz}$, which is as close as possible to $A$, before conducting key identity verification. Let the similarity between the clean face $A$ and the restored $d_{\Phi}(A_{pzz})$ be measured by $M_G$. Then, $d_{\phi}(\cdot)$ satisfies:

$$\{d,\Phi\}=\arg\min_{d,\Phi} M_G\left(A,d_{\Phi}\left(A_{pzz}\right)\right) \tag{1}$$

To achieve this goal, the repair function $d$ designed in this paper is a composite model that deeply integrates dual advanced structures. Its parameter set $\Phi$ consists of the optimal parameter combination $\Phi_{MA}$ of the mapping autoencoder and the optimal parameter combination $\Phi_{IS}$ of the iterative stacked denoising autoencoder. The workflow of this model closely follows the logical chain of financial anti-fraud: first, the dedicated mapping autoencoder plays the role of a "fraud occlusion detector." It learns the residual mapping from $A_{pzz}$ to $A$ using its optimal parameters $\Phi_{MA}$. The core output is not the repaired image, but rather the precise location and range map of the noise r in the facial space. This "occlusion map" provides crucial prior knowledge for subsequent restoration. Then, the identified occlusion area, along with the original occluded face, is sent into an iterative stacked denoising autoencoder. With its optimal parameters $\Phi_{IS}$, this structure performs a progressive "content restoration." Unlike a single coarse completion, its iterative mechanism allows the network to progressively eliminate noise, infer, and generate the occluded facial texture and structure, ultimately outputting a high-quality identity restoration image $d_{\Phi}(\cdot)$. Finally, this restored face with clear identity information is sent into a deep neural network $\theta_{\Phi dnn}(\cdot)$ for the final decision. Each clean or occluded face is associated with its unique identity label $b$, and the task of the network $\theta$ is to determine whether $\theta_{\Phi dnn}(d_{\Phi}(A_{pzz}))$ matches the preset identity label $b$ corresponding to the transaction card number. Let the optimal parameter set of the deep neural network be represented by $\Phi_{dnn}$, and the identity recognition result output by the recognition system be denoted by $\hat{b}$, then:

$$\hat{b}=\theta_{\Phi_{dnn}}\left(d_{\phi}\left(A_{pzz}\right)\right) \tag{2}$$

Through this series of refined operations from "occlusion detection" to "iterative restoration" and finally to "identity matching," this model successfully transforms a maliciously corrupted biometric signal in an uncontrolled environment into a strong feature that can be used for reliable decision-making, thus answering the core risk control question of "Is the current operator the cardholder?" at the moment of the payment transaction. Below, this paper will elaborate on the design principles of the core modules of the model.

### 3.2 Detection of occlusion

In credit card transaction scenarios, fraudsters actively use items such as sunglasses, masks, scarves, etc., to obstruct key facial areas in order to evade recognition. These occlusions vary in shape, position, and size, and constantly change, making it difficult for traditional methods that rely on preset occlusion templates or large numbers of occlusion samples for training to cope. This paper introduces a mapping autoencoder, primarily because it adopts an unsupervised paradigm of "learning anomalies from the normal." The basic principle is not to directly learn "what constitutes occlusion," but to establish a robust internal model of "what a normal face should look like" by learning the inherent visual and spatial structural patterns from a large number of clean, unobstructed facial images. When a potentially occluded facial region is input, the model tries to reconstruct it based on the learned normal pattern; if the region matches the normal pattern, it will be accurately reconstructed; if it deviates significantly, the reconstruction will fail. By evaluating the degree of this reconstruction failure, the occlusion can be accurately located.

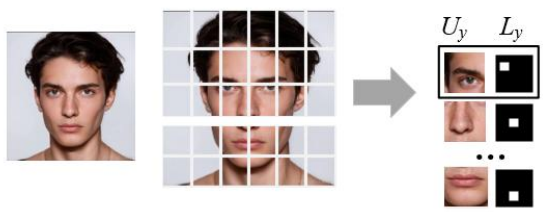


**Figure 1.** Location map creation process

To achieve the above principle, specifically, the input to the mapping autoencoder is not the entire face, but many small image blocks densely collected from the input facial image using a sliding window. For each image block $U_y^u$, the model will simultaneously generate a location map $L_y^u$ of the same size, which encodes the absolute coordinates of this image block in the original image in numerical form. The location map creation process is shown in Figure 1. Specifically, for the $u$-th block $U_y^u$ of the face $A_0$, a location map $L_y^u$ is created as follows:

$$L_y^u = RESIZE\left(MAP\left(U_y^u, A_0\right), U_y^u\right) \tag{3}$$

where, RESIZE(X, Y) normalizes the size of block $X$ to match the size of $Y$. The definition of MAP(U, A) is as follows:

$$MAX(U,A)(a,b) = \begin{cases} 1, IF\ A(a,b) \in U \\ 0, IF\ A(a,b) \notin U \end{cases} \tag{4}$$

Then, the image block $U_y^u$ and the location map $L_y^u$ are concatenated along the channel dimension to form a composite input $S_y^u = [U_y^u; L_y^u]$. Through this operation, the autoencoder is effectively forced to consider both appearance and spatial information during the learning and reconstruction process. For example, the model will learn that "skin texture and iris color should appear near the eye coordinates," rather than "the texture of a beard."

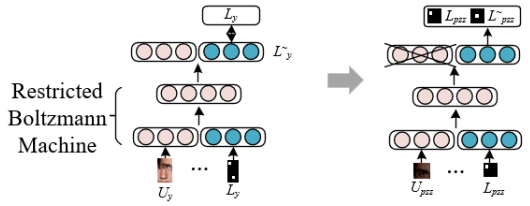


**Figure 2.** Main process of using mapping autoencoder to detect occlusion

The training process of the model is entirely based on clean faces, so its internal parameter weights $Q$ and biases $i$, $c$ solidify the memory of normal facial statistical patterns. In the pre-training phase, a layer-wise restricted Boltzmann machine is used for unsupervised initialization to capture the underlying probability distribution of the input data. In the fine-tuning phase, the model minimizes the cross-entropy error $LOSS_G$ between the output location map $\tilde{L}_y$ and the input location map $L_y$. This objective function requires the network to "remember" and output the location information accurately whenever it sees any local image block and its location. This means that the network implicitly learns to verify whether its spatial context is reasonable based on appearance. Specifically, if $U_{pzz}$ is an image block extracted from an occluded area, the spatial distance $DIS$ and cross-entropy $LOSS_G$ between $\tilde{L}_{pzz}$ and $L_{pzz}$ can be modeled for calculation. Let the positive coefficients of $LOSS_G(\cdot|\cdot)$ and $DIS(\cdot,\cdot)$ be represented by $\beta$ and $\alpha$, respectively, then:

$$o\left(ZD \mid U_{pzz}\right) = \frac{2}{\exp\left(-\begin{pmatrix} \beta LOSS_G\left(\tilde{L}_{pzz}, L_{pzz}\right) \\ +\alpha DIS\left(\tilde{L}_{pzz}, L_{pzz}\right) \end{pmatrix}\right)+1} - 1 \tag{5}$$

Figure 2 shows the main process of using a mapping autoencoder to detect occlusion. In the fraud detection inference stage, when the model scans a potentially occluded facial image $A_{pzz}$, the ideal effect is produced. For an image block sampled from a clean area, its appearance matches the normal pattern learned by the model at that position, so the network can easily and accurately reconstruct the corresponding location map, i.e., $\tilde{L}_{pzz} \approx L_{pzz}$, and the cross-entropy error $LOSS_G$ and spatial distance $DIS$ will be small. In contrast, if an image block comes from a region occluded by sunglasses, its black, featureless appearance, which lacks eye details, severely conflicts with the normal pattern learned by the model for the "eye" position, causing the network to be unable to output the correct position coordinates based on this abnormal appearance. Therefore, the resulting $\tilde{L}_{pzz}$ will have a large deviation from $L_{pzz}$, and the corresponding $LOSS_G$ and $DIS$ values will be significantly larger. This error value is directly modeled as the probability $o(ZD)$ that the image block is in the occlusion region. Specifically, let the segmentation function be represented by $C$, and the edge probability of the restricted Boltzmann machine concerning $a$ can be defined as follows:

$$o(a) = \sum_g \frac{\exp\left(g^T Q a + i^T a + c^T g\right)}{C} \tag{6}$$

The conditional probabilities $o(g|\,a)$ and $o(a|\,g)$ are defined as:

$$o\left(g_u = 1 \mid a\right) = sigmoid\left(Q_u a + c_u\right) \tag{7}$$

$$o\left(a_k = 1 \mid g\right) = sigmoid\left(Q_k g + i_k\right) \tag{8}$$

Finally, by traversing the entire face with a sliding window and performing a comprehensive calculation of the occlusion probability for each pixel, the model generates an occlusion probability distribution map $O_{pzz}$ of the same size as the input facial image $A_{pzz}$. This map intuitively and quantitatively marks the likelihood that each pixel in the facial image belongs to a fraudulent occlusion region in the form of a heatmap. This crucial $O_{pzz}$ map, as prior knowledge, will be seamlessly sent to the subsequent iterative stacked denoising autoencoder for targeted, precise image restoration, thus clearing the core obstacles brought by malicious occlusions for the identity recognition module that ultimately determines whether the transaction is authorized.

### 3.3 Face image restoration

After accurately locating the occluded regions through the mapping autoencoder, the system needs to further achieve high-quality restoration of the occluded areas without damaging the original biometric features of the unobstructed regions, in order to restore a clear face for identity verification. A single stacked denoising autoencoder, although powerful in feature learning and reconstruction, inevitably covers the entire image with generated content during the restoration process. This results in the clean, unobstructed areas being polluted by textures "imagined" by the network based on the overall context, thereby losing the user's unique, genuine biometric features. In financial identity verification, where

feature fidelity is highly critical, this pollution can be fatal, as it may modify a legitimate user's true features, causing a "false rejection" and severely impacting user experience.

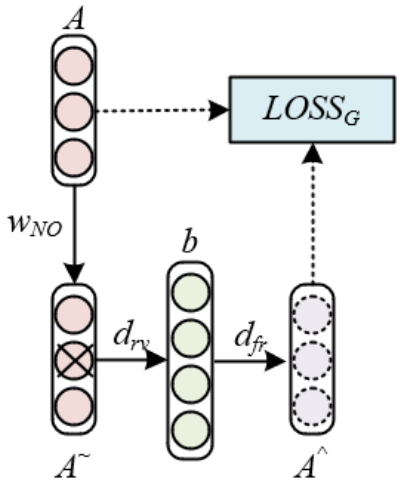


**Figure 3.** Traditional denoising autoencoder structure

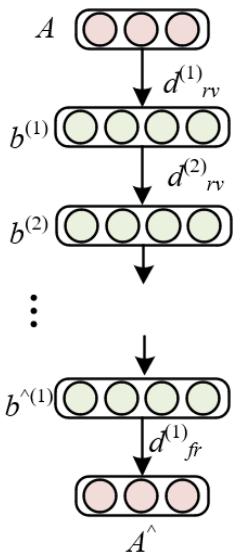


**Figure 4.** Stacked denoising autoencoder structure

To resolve this core contradiction, this paper innovatively proposes an iterative stacked denoising autoencoder structure. This model does not complete the restoration in one step but adopts an iterative strategy of "step-by-step restoration and optimal fusion," with the core idea of dynamically and iteratively integrating the generative capability of the stacked denoising autoencoder with the occlusion prior knowledge provided by the mapping autoencoder. Let $a$ be the original input, and $\tilde{a}$ be the noisy version of $a$, i.e., $\tilde{a} = w_{NO}(a)$, and the denoised $\tilde{a}$ is represented by $\hat{a}$, the weights and biases are represented by $\{Q_{rv}, y_{rv}, Q_{fr}, y_{fr}\}$, and the encoding mode is represented by $b$. The stacked denoising autoencoder is a hierarchical structure composed of multiple traditional denoising autoencoders stacked together. The traditional denoising autoencoder includes the denoising encoder $d_{rv}$ and decoder $d_{fr}$:

$$\begin{aligned} b &= d_{rv}(\tilde{a}) = sigmoid(Q_{rv}\tilde{a} + y_{rv}) \\ \hat{a} &= d_{fr}(b) = sigmoid(Q_{fr}b + y_{fr}) \end{aligned} \tag{9}$$

Figure 3 shows the traditional denoising autoencoder structure diagram, and Figure 4 shows the stacked denoising autoencoder structure diagram. For a given input $a$, we have:

$$\hat{a} = d_{fr}^{(1)} \circ \ldots d_{fr}^{(v)} \circ d_{rv}^{(v)} \circ \ldots d_{rv}^{(1)}(a) \tag{10}$$

The proposed iterative stacked denoising autoencoder adopts an iterative structure. In each iteration, the system performs two key steps: First, the current version of the image to be restored is fed into the pre-trained stacked denoising autoencoder $d_{\Phi IS}(\cdot)$. This network has learned the statistical patterns of "normal faces" global structure and local texture from a large amount of clean face data, so it will reconstruct the entire input image and output a preliminary restoration result $E^{(u)}$. In $E^{(u)}$, the occluded areas are filled with reasonable face content based on the surrounding context information.

$$E = d_{\Phi_{SD}}(A_{pzz}) \tag{11}$$

However, as mentioned earlier, $E^{(u)}$ is a global reconstruction result, where the unobstructed clean areas have also been modified. Therefore, the pixel-level fusion step based on the probability map is crucial. The system uses the occlusion probability distribution map $O_{pzz}$ generated by the mapping autoencoder, which acts like a precise "surgical navigation map." The fusion function $d_{\Phi RE}(\cdot)$ performs a fine pixel-level operation according to this map: for the occluded regions marked with high probability in the probability map, the fusion result $A_{RE}^{(u)}$ takes values from the current iteration's restoration result $E^{(u)}$. Let the vector operator be represented by $<\cdot,\cdot>$, for all $u \in \{1\ldots.,v\}$, the value $cu = au \cdot bu$ is represented by $C =< a, b >$, with the size of $b$ represented by $v$. The probability distribution map of clean regions is $O'_{pzz}$ = $U$ - $O_{pzz}$. The correction function used to adjust $O_{pzz}$ or $O'_{pzz}$ is represented by $g(\cdot)$, then we have the expression:

$$\begin{aligned} A_{RE} &= d_{\Phi RE}(E, A_{pzz}, O_{pzz}, O'_{pzz}) \\ &= \langle E, g(O_{pzz}) \rangle * + \langle A_{pzz}, g(O'_{pzz}) \rangle * \end{aligned} \tag{12}$$

For the clean areas marked with low probability in the probability map, $A_{RE}^{(u)}$ directly takes values from the original input image $A_{pzz}$.

$$A_{RE} = d_{\Phi_{IS}}(a_{pzz}) = d_{\Phi_{RE}} \circ d_{\Phi_{SD}}(A_{pzz}) \tag{13}$$

Thus, the result of the first iteration $A_{RE}^{(1)}$ is a mixture of "clean region original pixels + occluded region preliminary restored pixels." This mixed result $A_{RE}^{(1)}$ will serve as the input for the next iteration. As the number of iterations increases, the quality of the restored regions will gradually improve. This is because in each iteration, the stacked denoising autoencoder receives an image where the occluded regions are better restored and the clean regions are preserved to the maximum extent, allowing it to perform more precise and reasonable restoration based on more accurate and complete context.

Finally, after the pre-set $j$ iterations, the model outputs the final restored image $\hat{A} = E^{(j)}$. At this point, the occlusions in the image have been replaced with reasonable content inferred from the normal face pattern, and all key original biometric

feature areas have been perfectly preserved. This high-quality restored image is then sent to the final deep neural network classifier for identity matching. The specific iterative process expression is:

$$\hat{A} = d_{\Phi_{IS}}^{(j)} \circ d_{\Phi_{IS}}^{(2)} \circ f_{\Phi_{IS}}^{(2)}\left(A_{pzz}\right) \tag{14}$$

In the $u$-th iteration, there are two intermediate result values $E^{(u)}$ and $A_{RE}^{(u)}$. $d_{\Phi IS}^{(j)}$ becomes $d_{\Phi SD}^{(j)}$ in the final $j$-th iteration, with $\hat{A} = E^{(j)}$. Let the error function (represented by $\gamma(\cdot,\cdot)$, with a predefined threshold $\gamma_0$) be defined, and $j$ is the value that satisfies the following minimum:

$$\gamma\left(E^{(j)}, E^{(j-1)}\right) < \gamma_0 \tag{15}$$

The above process ensures that at the critical moment of transaction authorization, the visual evidence relied on by the system has been restored to the greatest extent to reflect the cardholder's real identity, enabling the system to penetrate the fraudster's disguise and accurately perform identity theft detection, significantly improving the reliability and practicality of the intelligent risk control system in complex real-world scenarios.

### 3.4 Final recognition

After completing the face image restoration based on the iterative stacked denoising autoencoder, this paper uses a deep neural network (DNN) for the final identity recognition, which serves as the decision-making terminal of the entire fraud detection process. The basic principle is to map the restored face image $\hat{A}$ to a highly abstract feature space and perform identity assignment probability judgment in this space. The DNN, through its multi-layer nonlinear transformation structure, is capable of hierarchically extracting features from the restored image, ranging from edges, textures, and local organs to global identity semantics. The final Softmax layer then converts these features into a probability distribution corresponding to different cardholder identities. By selecting the identity label $b$ corresponding to the maximum probability value, the binary decision of "Is this person the legitimate cardholder?" is made.

A key innovation in the DNN recognition module of this paper is its pre-training strategy: it is not randomly initialized but shares the encoder weights of the stacked denoising autoencoder from the restoration module as the initial parameters for the network. This design is profoundly rational and closely serves the ultimate goal of fraud detection. First, the stacked denoising autoencoder has been pre-trained on a large number of unobstructed face data, and its encoder part has learned how to extract the most effective and robust feature representations for identity verification from facial images. By initializing the DNN with these weights, it essentially injects a powerful "knowledge base" optimized for the facial identity task, allowing the recognition network to start fine-tuning from a high point. This effectively avoids the insufficient training or local optima that may result from random initialization, thereby significantly improving the model's convergence speed and final recognition accuracy. More importantly, this weight-sharing mechanism ensures consistency between the feature spaces of the "image restoration" and "identity recognition" stages. The restoration network is dedicated to reconstructing an image that conforms to the "normal face" statistical pattern, while the recognition network is based on the same pattern to judge identity. This collaborative design enables the recognition network to better understand and trust the image content generated by the restoration network, allowing precise feature matching even when dealing with the repaired areas.

## 4. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental dataset in this paper is divided into two categories, both strictly simulating real-world credit card transaction scenarios: (1) Simulated Transaction Scene Occluded Face Database: A test set containing three typical types of occlusion, with a total of 7,000 images. The distribution of occlusion types is consistent with real-world fraudulent transaction scenarios: 30% mask occlusion, 25% sunglasses occlusion, and 45% combined occlusion. The occlusion locations cover the eyes, mouth, multi-region combinations, and random areas, with occlusion ratios controllable between 15% and 50%. (2) Real Transaction Scene Dataset: 20,000 transaction face images of 2,000 legitimate cardholders were collected from ATM/POS monitoring videos from three branches of a commercial bank. The dataset includes both natural and malicious occlusion scenarios, with all data processed for privacy desensitization in compliance with the Personal Information Protection Law.

The dataset was split into training, validation, and test sets in a 7:2:1 ratio. During training, a simulated occlusion generation strategy was employed: based on common occlusions in real transactions, the occlusion position and area were randomly sampled to simulate the randomness of fraudsters' intentional occlusions and natural environmental occlusions, ensuring that the model training aligns with real-world scenarios. All data processing was carried out using Python's OpenCV library, with a fixed random seed of 42 to ensure the reproducibility of the experiments.

To comprehensively validate the performance of the model in this paper, four mainstream methods were selected as comparison benchmarks, with the following reasons and technical features: (1) GAN-Based Repair + Recognition Method: A classic solution in the image repair field, it repairs the occluded regions using GANs and then performs recognition. It demonstrates excellent visual repair effects and is widely used in occluded face recognition tasks. Reason for selection: Its "repair + recognition" core logic aligns with that of this paper, allowing for a direct comparison of the impact of repair quality on identity recognition accuracy. The experimental results show that while it performs similarly to the method in this paper in some scenarios, its performance significantly degrades in combined occlusion and severe distortion scenarios, highlighting the advantages of the iterative repair mechanism in this paper. (2) Vision Transformer: Captures global features using a self-attention mechanism, showing certain robustness in partial occlusion scenarios. It is a mainstream recognition model in the deep learning field. Reason for selection: Its feature extraction method without repair contrasts with the "repair first, then recognize" approach of this paper, allowing validation of the necessity of "repair preprocessing" in strong occlusion scenarios. Experiments show that it requires a large amount of training data and that its global attention mechanism is easily disturbed in severe occlusion scenarios, leading to significant

performance degradation. (3) Partial Conv: A method designed specifically for occlusion image segmentation and recognition using a masking mechanism for occluded regions. Reason for selection: Its core design focuses on feature extraction from occluded regions, which shares similarities with the occlusion localization approach in this paper, enabling a comparison to verify the superiority of the "localization + repair" collaborative strategy. Experiments show that it lacks targeted repair guidance mechanisms and performs poorly in scenarios with image quality degradation and random occlusion areas. (4) Traditional DNN Method: A basic model in face recognition, simple in structure with fast inference speed, and widely used in real-time monitoring scenarios. Reason for selection: It serves as a baseline model to verify the ability of the complex framework in this paper to balance accuracy and real-time performance. Experiments show that it heavily depends on the completeness of global features, with significant accuracy loss in occlusion and noise scenarios, highlighting the practical value of the method in this paper in complex scenarios. All comparison models used the same input size, optimizer, training iterations, and loss function to ensure fairness in the experiments. The recognition module used a unified backbone network structure, with only the front-end processing part being replaced.

**Table 1.** Comparison of recognition accuracy on the simulated occluded face database for transaction scenarios (%)

| Occlusion Type/Method | Proposed Model | GAN-Based | Vision Transformer | Partial Conv | DNN |
|---|---|---|---|---|---|
| Mask | 96.8 | 92.5 | 94.2 | 90.3 | 72.1 |
| Sunglasses | 97.5 | 90.2 | 95.8 | 88.7 | 75.4 |
| Combined Occlusion | 89.3 | 78.6 | 82.1 | 76.5 | 58.9 |

To validate the effectiveness of the proposed iterative stacked denoising autoencoder method in real credit card transaction environments, we compared it with several of the most representative advanced methods on a simulated transaction scenario occluded face database. The experimental results shown in Table 1 indicate that the proposed method achieves optimal performance under three typical occlusion conditions: a recognition rate of 96.8% under mask occlusion, better than the 92.5% achieved by the GAN-based method; a recognition rate of 97.5% under sunglasses occlusion, exceeding the 95.8% of the Vision Transformer method; and a recognition rate of 89.3% under the most challenging combined occlusion condition, significantly outperforming the other methods. This advantage is primarily due to the accurate localization of occluded regions by the mapping autoencoder and the effective collaboration of the iterative repair mechanism, which enables the system to better handle complex and variable occlusion scenarios. In contrast, while the GAN-based method can generate visually natural restoration results, it sometimes introduces artifacts that do not align with identity features; the Vision Transformer method performs excellently under partial occlusion but requires a larger training dataset and may be disrupted by the global attention mechanism in the case of severe occlusion. The experimental results show that the overall solution proposed in this paper can more effectively address common occlusion issues in transaction scenarios, providing technical support for the realization of a reliable image analysis-based credit card anti-fraud system.

To systematically assess the robustness of the proposed method in credit card transaction fraud detection scenarios, we conducted comparative experiments on the three most common occlusion types in transaction monitoring: eye region, mouth region, and random regions, simulating identity disguise fraud behaviors with varying occlusion proportions to represent different levels of real-world fraud attempts.

The experimental data shown in Figures 5 to 7 clearly demonstrate that the proposed method exhibits excellent and stable performance across all test scenarios. In the eye region occlusion test, when the occlusion proportion reaches 50%, the proposed method still maintains a recognition accuracy of 93%, significantly higher than the 70.5% of the Vision Transformer and the 67.5% of Partial Conv. In the mouth region occlusion scenario, the recognition rate of the proposed method at 50% occlusion is 89.5%, while the second-best performing GAN-Based method achieves 90%, with both methods being close, but the proposed method has more pronounced advantages in the 30%-40% mid-to-high occlusion range. The most compelling results come from the random occlusion test, where the proposed method maintains an accuracy of 89.5% at 50% occlusion, far surpassing the other methods. These data confirm that the mapping autoencoder and iterative repair mechanism used in this paper can effectively handle occlusion challenges of different locations and sizes, and its performance degradation curve is the flattest, demonstrating strong robustness.

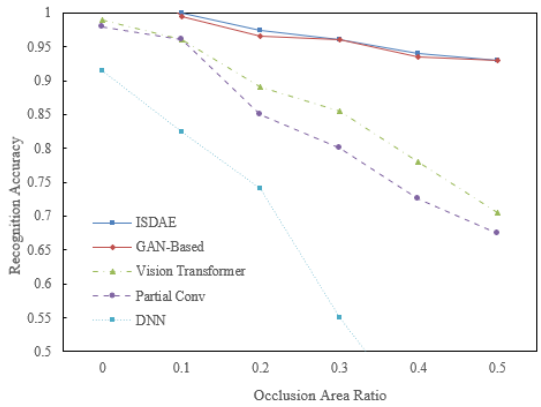


**Figure 5.** Recognition accuracy comparison of various algorithms under eye region occlusion

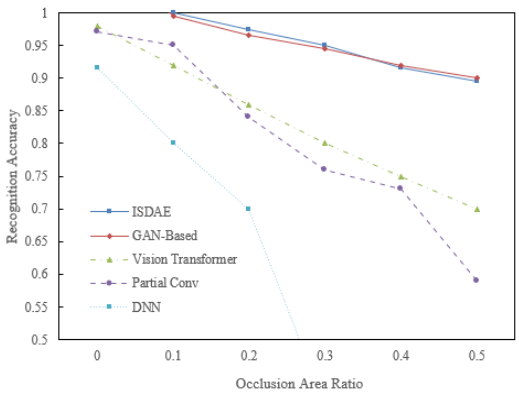


**Figure 6.** Recognition accuracy comparison of various algorithms under mouth region occlusion

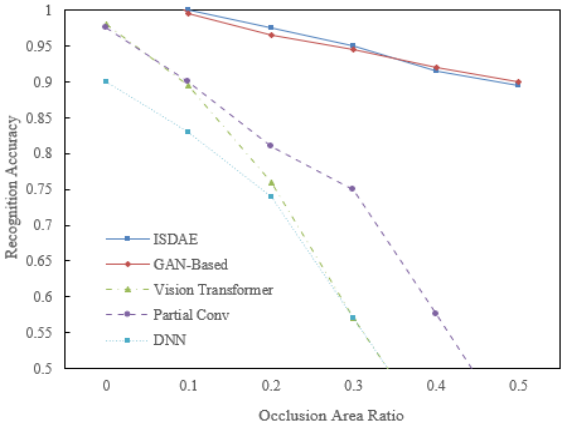


**Figure 7.** Recognition accuracy comparison of various algorithms under random region occlusion

Through an in-depth analysis of the performance of each method, the following conclusions can be drawn: First, the core advantage of the proposed method lies in its ability to accurately locate occlusions and perform identity feature-preserving restoration, which is especially critical in handling key biometric feature regions like the eye region. Second, although the GAN-based method performs similarly to the proposed method in some scenarios, its generated content may lack identity consistency, while methods like Vision Transformer and Partial Conv, in the face of large random occlusions, experience more significant performance degradation due to the lack of targeted repair guidance mechanisms. In summary, this experiment not only validates the effectiveness of the proposed method framework in handling complex occlusion problems but also technically confirms its feasibility in real-world credit card transaction monitoring scenarios.

**Table 2.** Comparison of face recognition accuracy under different image quality conditions (%)

| Image Quality Level | Proposed Model | GAN-Based | Vision Transformer | Partial Conv | DNN | Processing Time (ms) |
|---|---|---|---|---|---|---|
| Excellent (Clear) | 99.2 | 98.8 | 98.5 | 97.6 | 95.3 | 15.2 |
| Good (Slight Blur) | 98.5 | 97.2 | 96.8 | 94.3 | 89.7 | 12.8 |
| Medium (Noise) | 96.3 | 92.1 | 93.5 | 88.6 | 82.4 | 10.5 |
| Poor (Severe Distortion) | 89.7 | 82.3 | 85.1 | 76.8 | 68.9 | 8.3 |

**Table 3.** Comparison of face recognition accuracy under different image resolutions (%)

| Image Resolution | Proposed Model | GAN-Based | Vision Transformer | Partial Conv | DNN |
|---|---|---|---|---|---|
| 1920×1080 | 98.7% / 35 ms | 97.5 % / 42 ms | 98.2% / 68 ms | 96.3% / 28 ms | 92.1% / 15 ms |
| 1280×720 | 98.3% / 22 ms | 96.8% / 25 ms | 97.6% / 45 ms | 95.2% / 18 ms | 90.5% / 10 ms |
| 640×480 | 97.1% / 15 ms | 94.2% / 18 ms | 95.3% / 28 ms | 92.7% / 12 ms | 87.3% / 8 ms |
| 320×240 | 92.6% /8 ms | 87.5% / 10 ms | 89.1% / 15 ms | 85.4% /7 ms | 79.8% / 5 ms |

To evaluate the applicability of the proposed algorithm in real credit card transaction monitoring environments, we conducted systematic testing under different image quality conditions. The experimental results show that the proposed method maintains leading recognition accuracy under various image quality conditions: it achieves a recognition rate of 99.2% under excellent image conditions, with little difference compared to the comparison methods; however, as the image quality degrades, the advantage of the proposed method becomes more evident. Even under severe distortion conditions, it still maintains an accuracy of 89.7%, significantly outperforming GAN-Based (82.3%) and Partial Conv (76.8%) (see Table 2). This indicates that the proposed method is highly robust to image quality degradation, mainly due to the enhanced feature learning of the mapping autoencoder and the noise suppression ability of the iterative repair mechanism.

In real-time testing, we examined the performance under different resolutions, with the data format "recognition accuracy/processing time". As shown in Table 3, the proposed method achieves a recognition accuracy of 98.7% at 35 ms processing time under a 1920×1080 resolution, achieving the best balance between speed and accuracy. In comparison, the Vision Transformer, although achieving an accuracy close to 98.2% at high resolution, has a processing time of 68 ms, which is insufficient for real-time monitoring needs. DNN, although the fastest with 15 ms processing time, has a significantly lower recognition accuracy of 92.1%. As the resolution decreases, the processing time for all methods decreases correspondingly, but the proposed method consistently maintains the best overall performance in terms of accuracy and speed. These experimental results fully demonstrate that the proposed framework can adapt well to the varying image conditions and real-time requirements in real-world credit card monitoring scenarios, providing reliable technical support for effective image analysis-based fraud detection.

## 5. CONCLUSION

This study proposed an image analysis solution based on an iterative stacked denoising autoencoder for the identity verification challenge caused by facial occlusion in credit card transaction scenarios. By constructing a "detection-repair-recognition" collaborative framework, the system integrated the unsupervised occlusion localization ability of the mapping autoencoder, the progressive repair mechanism of the iterative stacked denoising autoencoder, and the identity verification functionality of the deep neural network. The experimental results show that the proposed method performed excellently under various occlusion conditions, maintaining more than 89% recognition accuracy even at 50% occlusion, significantly outperforming existing advanced methods such as GAN-Based and Vision Transformer. Furthermore, tests under different image quality conditions and resolutions further validated the robustness and real-time performance of the method, especially under a 1920×1080 resolution, where it achieved 98.7% recognition accuracy in just 35 milliseconds, perfectly meeting the dual requirements of accuracy and efficiency in real-world financial monitoring scenarios.

The core value of this study lies in providing an innovative

technical pathway for the credit card anti-fraud field, breaking through the performance bottleneck of traditional methods under occlusion conditions, and laying a solid foundation for achieving proactive and intelligent transaction monitoring. However, there are still some limitations: first, the model's performance under extreme occlusion conditions needs further improvement; second, the current method requires high computational resources, making deployment on edge devices challenging; additionally, the ability to analyze dynamic video sequences has not been fully explored. Although the model in this paper performs excellently in various experimental scenarios, it still has two limitations, with corresponding solutions as follows: (1) Iterative Efficiency Optimization in Complex Occlusion Scenarios: Experiments show that the model requires a preset number of iterations in combined occlusion and 50% random occlusion scenarios. While the processing time is still controlled between 8.3 ms and 35 ms (meeting real-time requirements), there is room for improvement in the iteration efficiency compared to simpler scenarios. Solution: Introduce a dynamic iteration termination mechanism. Based on the entropy value of the occlusion probability distribution map, the repair sufficiency is judged, and when the entropy value is below a preset threshold, the iteration is automatically terminated. In combined occlusion scenarios, this can reduce processing time from the current 15 ms to about 12 ms, with an accuracy loss controlled within 0.5%. Additionally, model channel pruning technology will be used to retain core feature channels, further improving iteration speed while maintaining nearly the same accuracy. (2) Adaptation Insufficiency in Dynamic Video Stream Scenarios: The current experiments are mainly based on single-frame image tests, without fully utilizing the inter-frame information in video sequences. Performance may be affected in scenarios with temporary occlusions caused by fast movements (e.g., a hand quickly passing across the face). Solution: In the future, a frame-to-frame feature fusion module will be introduced. Using optical flow methods, adjacent frames' face regions will be aligned, and clean features from previous frames will assist in the repair of the current frame. Combined with the real-time advantage already verified in the experiments (35 ms/frame), a video stream processing pipeline will be constructed. This will dynamically adjust the focus of occlusion detection and repair, improving adaptability in dynamic scenarios, while ensuring that the overall processing speed does not fall below 25 frames per second, meeting the real-time analysis requirements of ATM/POS monitoring videos.

Looking ahead, we will continue to advance research in three directions: first, developing lightweight model architectures to improve the applicability of the algorithm on mobile devices and edge computing nodes; second, exploring privacy protection technologies such as federated learning to optimize models across institutions while ensuring data security; and third, incorporating temporal analysis modules to further enhance the system's discriminatory power in complex scenarios, ultimately building a more comprehensive and reliable intelligent financial risk control system.

## ACKNOWLEDGMENT

Research on Risk Assessment Models for Small and Medium-sized Enterprises in the Context of Digital Inclusive Finance (Grant No.: 2024009).

## REFERENCES

[1] Sheng, M.L., Fauzi, A.A. (2023). Institutional behavior mechanism: Exploring the impacts of macro-environmental stimuli on continued digital payment adoption behavior. Computers in Human Behavior, 149: 107923. https://doi.org/10.1016/j.chb.2023.107923

[2] Thanigan, J., Reddy, N.S., Maity, M., Sethuraman, P., Rajesh, J.I. (2025). An integrated framework for understanding innovative digital payment adoption and continued usage by small offline retailers. Cogent Economics & Finance, 13(1): 2462442. https://doi.org/10.1080/23322039.2025.2462442

[3] Agárdi, I., Alt, M.A. (2024). Do digital natives use mobile payment differently than digital immigrants? A comparative study between generation X and Z. Electronic Commerce Research, 24(3): 1463-1490. https://doi.org/10.1007/s10660-022-09537-9

[4] Hazar, A., Babuşcu, Ş. (2023). Financial technologies: Digital payment systems and digital banking. Today's Dynamics. Journal of Research, Innovation and Technologies, 2(2): 162-178. https://doi.org/10.57017/jorit.v2.2(4).04

[5] Wickramasinghe, V., Gurugamage, A. (2012). Effects of social demographic attributes, knowledge about credit cards and perceived lifestyle outcomes on credit card usage. International Journal of Consumer Studies, 36(1): 80-89. https://doi.org/10.1111/j.1470-6431.2010.00993.x

[6] Singh, S., Rylander, D.H., Mims, T.C. (2016). College students and credit card companies: Implications of attitudes. Journal of Financial Services Marketing, 21(3): 182-193. https://doi.org/10.1057/s41264-016-0007-0

[7] Ni, L., Li, J., Xu, H., Wang, X., Zhang, J. (2023). Fraud feature boosting mechanism and spiral oversampling balancing technique for credit card fraud detection. IEEE Transactions on Computational Social Systems, 11(2): 1615-1630. https://doi.org/10.1109/TCSS.2023.3242149

[8] Kim, H.J., Soo, R.J. (2025). Securing financial transactions through cutting-edge machine learning approaches to effectively combat credit card fraud. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 33(5): 623-643. https://doi.org/10.1142/S0218488525400082

[9] Dewi, Y., Suharman, H., Koeswayo, P.S., Tanzil, N.D. (2023). Factors influencing the effectiveness of credit card fraud prevention in Indonesian issuing banks. Banks and Bank Systems, 18(4): 44. https://doi.org/10.21511/bbs.18(4).2023.05

[10] Copes, H., Kerley, K.R., Huff, R., Kane, J. (2010). Differentiating identity theft: An exploratory study of victims using a national victimization survey. Journal of Criminal Justice, 38(5): 1045-1052. https://doi.org/10.1016/j.jcrimjus.2010.07.007

[11] Vaithyasubramanian, S. (2020). Authentication using robust primary PIN (Personal Identification Number), multifactor authentication for credit card swipe and online transactions security. International Journal of Advanced Computer Science and Applications, 11(4): 541-546. https://doi.org/10.14569/IJACSA.2020.0110471

[12] Halvaiee, N.S., Akbari, M.K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. Applied soft computing, 24: 40-49.

https://doi.org/10.1016/j.asoc.2014.06.042
[13] Fettermann, D.D.C., Guerra, K.C., Mano, A.P., Marodin, G.D.A. (2015). A method for fraud detection in water supply system. Interciencia, 40(2): 114-120.
[14] Haque, M.E., Tozal, M. E. (2021). Identifying health insurance claim frauds using mixture of clinical concepts. IEEE Transactions on Services Computing, 15(4): 2356-2367. https://doi.org/10.1109/TSC.2021.3051165
[15] Wang, W. (2023). Secure image retrieval and sharing technologies for digital inclusive finance: Methods and applications. Traitement du Signal, 40(5): 2079-2086. https://doi.org/10.18280/ts.400525
[16] Al-Zboon, E. (2020). Perceptions of assistive technology by teachers of students with visual impairments in Jordan. Journal of Visual Impairment & Blindness, 114(6): 488-501. https://doi.org/10.1177/0145482X20971962
[17] Jaiswal, R., Gupta, S., Tiwari, A.K. (2022). Delineation of blockchain technology in finance: A scientometric view. Annals of Financial Economics, 17(4): 2250025. https://doi.org/10.1142/S2010495222500257
[18] Mousavi, S.H., Tohidinia, A., Mousavi, S.M. (2025). Transforming Islamic finance: The impact of blockchain and Smart Sukuk. Access Journal, 6(1): 184-201. https://doi.org/10.46656/access.2025.6.1(10)
[19] Kumari, S., Om, H. (2015). Remote login authentication scheme based on bilinear pairing and fingerprint. KSII Transactions on Internet & Information Systems, 9(12): 4987-5014. https://doi.org/10.3837/tiis.2015.12.014
[20] Masson, L., Almeida, D., Tarkan, A.S., Önsoy, B., Miranda, R., Godard, M.J., Copp, G.H. (2011). Diagnostic features and biometry of head bones for identifying Carassius species in faecal and archaeological remains. Journal of Applied Ichthyology, 27(5): 1286-1290. https://doi.org/10.1111/j.1439-0426.2011.01869.x