Behavioral Biometrics-Based Continuous Authentication Using a Lightweight Latent Representation Masked One-Class Autoencoder

Mingming Hu, Ding Wang, *Member, IEEE*, Chen Li, Yang Xu, and Bibo Tu

***Abstract*—Behavioral biometrics-based continuous authentication has proven to be an excellent supplement to one-time authentication schemes that applies behavioral biometrics to authenticate smartphone users' identities throughout the session. However, it still has two key issues that need to be addressed: 1) Due to behavioral biometrics from attackers are not available a priori, continuous authentication models should be trained only with normal samples in an unsupervised manner rather than treated as binary or multi-class classification tasks; 2) Differences in behavioral biometrics between attackers and legitimate users are fine-grained, it is challenging to extract rich semantic information to model users' behavioral patterns. To fill this gap, we propose a lightweight latent representation masked one-class autoencoder, which is trained only with legitimate users' behavioral biometrics. It consists of two parts: masked latent representation generator (MLRG) and Reconstructor. First, we apply the MLRG to generate discriminative latent representation with low dimension and then as a mask to cover important parts of the latent representation generated from the Reconstructor. Second, we apply the Reconstructor to reconstruct input based on masked latent representation. Experimental results demonstrate that our approach achieves superior authentication performance of 0.68% EER, 0.94% EER, 2.14% EER, and 1.87% EER on four datasets, respectively.**

***Index Terms*—Behavioral biometrics, continuous authentication, latent representation learning, masked one-class autoencoder.**

I. Introduction

NOWADAYS, smartphone has become an essential tool for accessing online services in our daily life, and smartphone users always store a large amount of personal privacy information on smartphones (accounts, passwords, call records, pictures, etc.). However, while smartphones bring us convenience, our personal privacy information is also at risk of being leaked. Once an attacker gains access to the smartphone, private personal information stored on the smartphone will be available to the attacker. To prevent attackers from accessing the user's smartphone and stealing the smartphone user's personal privacy data, an access control scheme that can authenticate smartphone users' identity is urgent.

Current popular one-time authentication schemes can be divided into two categories: Knowledge-based authentication mechanisms and physiological-based authentication mechanisms. However, knowledge-based authentication mechanisms (e.g., passwords, PINs, and graphical passwords) are vulnerable to smudge attack [1], side channel attack [2], and shoulder surfing attack [3]. And the physiological-based authentication mechanisms (e.g., fingerprint recognition, face recognition, iris recognition) may lead to the leakage of users' physiological characteristics information, such as fingerprint information, face images, and so on. Besides, both knowledge-based and physiological-based authentication mechanisms only perform one-time authentication during the initial login process. Once the smartphone user passes the authentication, these one-time authentication mechanisms no longer re-authenticate the user's identity throughout the session.

Therefore, an authentication scheme that can continuously authenticate an access user's identity throughout the session is necessary. Continuous authentication based on behavioral biometrics is an excellent supplement to traditional one-time authentication schemes that applies behavioral biometrics collected from the embedded sensors (accelerometer, gyroscope, magnetometer, touchscreen sensor, gravity, etc.) to authenticate a smartphone user's identity throughout the session. Compared with traditional one-time authentication schemes, it has three advantages: 1) Behavioral biometrics-based continuous authentication scheme authenticates a smartphone user's identity continuously throughout the session rather than only authenticating the smartphone user once when accessing the smartphone [4]; 2) During the continuous authentication process, smartphone users do not need to participate in the authentication process and can perform their tasks without interruption [5]; 3) No additional hardware support is required, behavioral biometrics used to model smartphone users' behavioral patterns are collected from embedded sensors [6].

Received 18 July 2023; revised 9 July 2024; accepted 29 September 2024. Date of publication 3 October 2024; date of current version 15 May 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFB3106303, and in part by the Natural Science Foundation of Tianjin, China, under Grant 21JCZXJC00100. *(Corresponding authors: Bibo Tu; Ding Wang.)*

Mingming Hu is with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100045, China, also with the School of Cyber Security, University of Chinese Academy of Sciences, Beijing 101408, China, and also with Yellow River Henan Bureau, Zhengzhou 450004, China (e-mail: humingming0319@iie.ac.cn).

Ding Wang is with the College of Cyber Science, Nankai University, Tianjin 300350, China (e-mail: wangding@nankai.edu.cn).

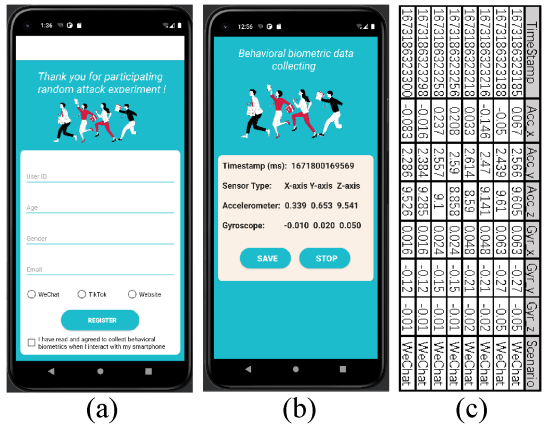
Chen Li, Yang Xu, and Bibo Tu are with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100045, China, and also with the School of Cyber Security, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: lichen@iie.ac.cn; xuyang@iie.ac.cn; tubibo@iie.ac.cn).

Digital Object Identifier 10.1109/TDSC.2024.3472631

Traditional machine learning and deep learning techniques have been widely applied in continuous authentication tasks. We categorize behavioral biometrics-based continuous authentication schemes that apply classical machine learning or deep learning techniques into three categories: 1) Classical machine learning-based continuous authentication: Classical machine learning algorithms are applied to learn users' behavioral patterns, such as Support Vector Machine (SVM) [6], [7], [8], [9], Random Forest [10], [11], [12], [13], [14], Hidden Markov Model (HMM) [15], [16], [17], Manhattan method [4], [18]; 2) Continuous authentication combining classical machine learning and deep learning: Deep learning models are used to learn rich semantic representations from behavioral biometrics in the first stage, and then rich representations are fed to the classical machine learning algorithms to distinguish the legitimate user and attacker in the second stage [19], [20], [21], [22], [23], [24], [25], [26]; 3) End-to-end deep neural network-based continuous authentication: End-to-end deep neural network-based continuous authentication schemes, in contrast, apply the latent representation learning directly for continuous authentication task [27], [28], [29], [30], [31], [32], [33].

Although continuous authentication schemes based on traditional machine learning and deep learning techniques have made remarkable progress, they still face the following limitations and challenges: 1) Complex feature engineering is required to process raw behavioral biometrics data to improve the ability of classical machine learning algorithms to learn smartphone users' behavioral patterns. However, complex feature engineering is time-consuming and is not scalable in practice; 2) Continuous authentication combining classical machine learning and deep learning techniques need to train two models to achieve satisfactory authentication performance. A deep learning model extracts discriminative deep features from behavioral biometric data, and a traditional machine learning model is trained to detect attackers. They have more computational overhead and time cost in the authentication phase; 3) End-to-end deep learning techniques based continuous authentication schemes are trained in a supervised manner for binary classification or multi-classification tasks, which face the problem of not knowing the distribution of anomalous samples in the real-world environment.

To solve these limitations existing in current continuous authentication methods. Inspired by general denoising autoencoders [34], we propose a lightweight latent representation masked one-class autoencoder for the continuous authentication task, which is trained only with normal samples in an unsupervised manner leveraging the behavioral biometrics collected from two embedded motion sensors (accelerometer and gyroscope). Anomaly scores based on reconstruction error using autoencoders have been widely applied as a base for the one-class classification task [35], [36], [37], [38], [39], [40], [41]. The autoencoders are always trained with normal samples, aiming to learn the distribution of normal samples. It is assumed that the trained autoencoder with normal samples will have a lower reconstruction error for normal samples and a higher reconstruction error for outlier samples. The proposed lightweight latent representation masked one-class autoencoder



TimeStamp	Acc_x	Acc_y	Acc_z	Gyr_x	Gyr_y	Gyr_z	Scenario
1671863233188	-0.005	2.439	9.61	0.063	-0.027	-0.005	WeChat
1671863233218	0.0033	2.614	8.59	0.048	-0.021	-0.002	WeChat
1671863233248	-0.146	2.47	9.141	0.048	-0.021	-0.002	WeChat
1671863233278	0.0216	2.59	8.858	0.024	-0.015	-0.001	WeChat
1671863233308	0.0237	2.557	9.1	0.024	-0.015	-0.001	WeChat
1671863233338	-0.016	2.384	9.285	0.016	-0.012	-0.001	WeChat
1671863233368	-0.013	2.386	9.528	0.016	-0.012	-0.001	WeChat

Fig. 1. The developed behavioral biometric data collection App (Android) for anti-attack analysis. a) user registration page; b) behavioral biometric data collection, display, and storage page; c) example of generated behavioral biometric data file (.csv) at /data/data/com.example.BehavioralBiometricsCollection/files/ in Android system.

consists of a masked latent representation generator (MLRG) and a Reconstructor. It only contains two convolutional layers and two transpose convolutional layers, which is lightweight enough for deploying on mobile platforms. In the training phase, we apply the masked latent representation generator (MLRG) to generate the low dimensional latent representation with discriminative semantic information. Then it is used as a mask to cover the optimal part of the latent representation generated from the encoder of the Reconstructor, rather than masking the input with a random masked proportion [42], [43], [44], [45], [46], [47], [48], [49]. And the Reconstructor is used to reconstruct the input based on the masked latent representation. In the test phase, instead of using the reconstruction error of the input to calculate the anomaly score, we apply the hidden reconstruction error [41] as a metric to detect anomalous samples from attackers.

We construct a continuous authentication system to evaluate the effectiveness of the proposed latent representation masked one-class autoencoder for the continuous authentication task. The experimental results show that the proposed continuous authentication system can achieve excellent performance on four public datasets. Additionally, to evaluate the ability of our continuous authentication system to resist various attacks in the real-world environment, we develop a behavioral biometric data collection App based on the Android system.[1] As shown in Fig. 1(a), participants who are selected as attackers use three application scenarios (WeChat, TikTok, and Website) to interact with legitimate user's smartphones, and our developed App can collect the behavioral biometric data from the accelerometer and gyroscope (shown in Fig. 1(b)) when attackers interact with the legitimate users' smartphones. Fig. 1(c) shows the example of stored behavioral biometric data (.csv) in Android system.

The contributions are summarized as follows:

- We propose a lightweight latent representation masked one-class autoencoder for continuous authentication task,

[1]https://github.com/IIEHU/Behavioral-Biometrics-Collection

which is trained only with the legitimate user's behavioral biometrics in an unsupervised manner. We apply the masked latent representation generator (MLRG) to generate the optimal semantic information of the input, and it is used as a mask to cover the optimal part of the latent representation generated from the encoder of the Reconstructor. Then the Reconstructor can learn richer fine-grained semantic information from the behavioral biometrics based on the masked latent representation.
- We design a continuous authentication system to evaluate the effectiveness of the proposed lightweight latent representation masked one-class autoencoder for continuous authentication tasks. It consists of the data preprocessing module, representation learning from normal samples module, and continuous authentication module.
- As shown in Fig. 1, we develop a behavioral biometric data collection App based on the Android system for random attack analysis in the real-world environment. It can capture behavioral biometric data through the built-in sensors (accelerometer and gyroscope) of smartphones during user interactions. We believe that the behavioral biometric data collection App can provide convenience for researchers in the field of continuous authentication based on behavioral biometrics or gait recognition.

The rest of this paper is organized as follows: Section II reviews the related work. We introduce the threat model in Section III. We describe the proposed lightweight latent representation masked one-class autoencoder in Section IV. Section V introduces the designed continuous authentication system. Comprehensive experiments in detail are presented in Section VI. We discuss our work in Section VII. Finally, we conclude this work in Section VIII.

II. Related Work

A. Behavioral Biometrics-Based Continuous Authentication

Many sensors (accelerometer, gyroscope, magnetometer, touchscreen sensor, and gravity, etc.) are integrated into smartphones, and behavioral biometrics data collected from these sensors can be used to measure smartphone users' behavioral habits when interacting with smartphones.

Continuous authentication based on motion sensors (accelerometer, gyroscope, magnetometer, etc.) has attracted many researchers and achieved remarkable progress [6], [11], [15], [20], [21], [22], [27], [29]. Zou et al. [29] developed a hybrid deep neural network to extract time and space domain features from the gait biometrics collected by the accelerometer and gyroscope. Shen et al. [15] applied a Hidden Markov Model-based one-class classifier to learn the smartphone user's behavioral patterns based on the motion sensor data. Li et al. [20] developed a convolutional neural network architecture to extract deep features from the sensory data collected by the accelerometer and gyroscope. Lee et al. [11] leveraged the accelerometer and gyroscope to collect the behavioral characteristics of smartphone users for the continuous authentication task. Li et al. [22] and Li et al. [6] applied the deep learning method and traditional data augmentation methods to enrich the diversity of behavioral biometrics collected from motion sensors, respectively.

Touch behavior-based continuous authentication has also made significant development, which uses touch behavioral biometrics collected from touchscreen sensor [10], [50], [51], [52], [53]. Frank et al. [50] collected touch biometrics data when smartphone users interact with smartphones in different sessions and tasks. Shen et al. [10] explored the use of touch-interaction behavior to learn users' behavioral patterns when interacting with smartphones, and then four classifiers were trained to classify the legitimate user and attacker. Shen et al. [51] attempted to learn discriminative features that characterize smartphone users' touch habits from four types of touch-interaction behaviors (sliding upwards, sliding downwards, sliding leftwards, and sliding rightwards). Teh et al. [53] proposed a touch dynamics biometrics-based continuous authentication scheme, and they evaluated the authentication performance of different groups of classification algorithms. Meng et al. [52] collected the touch behavioral biometrics data for the continuous authentication task when smartphone users browse the web.

Multi-modality-based continuous authentication fuses multiple behavioral biometrics to improve authentication performance [4], [13], [28], [33], [54], [55]. Sitová et al. [4] fused the behavioral features of hand movement, orientation, and grasp to authenticate smartphone a user's identity continuously. Shen et al. [54] constructed a behavioral biometrics dataset containing multiple modalities (motion movement biometrics, touch biometrics, and usage context) in the wild. Abuhamad et al. [28] collected smartphone users' behavioral biometrics from five embedded sensors and then fused these biometrics data to learn users' behavioral patterns. Buriro et al. [55] recorded the micro-movement signals of the smartphone and the movement signals of the user's finger on the touchscreen when interacting with the smartphone.

In this paper, we only leverage the behavioral biometrics data collected from two motion sensors (accelerometer and gyroscope) to model smartphone users' behavioral patterns, which do not need system privileges to read sensory data. Unlike these behavioral biometrics-based continuous authentication schemes, our proposed continuous authentication model utilizes only the legitimate user's behavioral biometrics to learn the user's behavioral patterns during the training process. It means that our proposed continuous authentication scheme does not require prior learning from the diverse and unknown distribution of samples from attackers to enhance authentication performance.

B. One-Class Classification

One-class classification (OCC) is widely applied in the domain of novelty detection, outlier detection, and anomaly detection. The one-class classification model is trained only with the normal samples from the target class and attempts to learn the distribution of the target class. The objective of OCC is to detect whether the input query sample belongs to the target class. The research of one-class classification can be represented into two strategies: 1) Computing the anomaly score of the query sample based on the encoder-decoder architecture [35], [36],

[37], [38], [39]; 2) Capturing the distribution of normal behavior from the target class based on the latent space [40], [41], [56], [57], [58], [59].

In the former strategy, Xia et al. [37] employed an autoencoder to learn the reconstruction of the input, and they supposed that the inliers and the outliers could be effectively distinguished according to the reconstruction errors. Sabokrou et al. [35] proposed an end-to-end network for one-class classification tasks based on generative adversarial networks (GANs). In the test phase, the discriminator acts as the one-class classifier and benefits from the support of the reconstructor. Perera et al. [38] trained a denoising-autoencoder with two discriminators and a classifier that the low-dimensional latent space can better model the distribution of the target class. Jewell et al. [39] designed a framework with two practical network modules, a mask module, and a reconstructor to learn richer contextual semantic information. Zhang et al. [36] presented an end-to-end adversarial autoencoder for one-class novelty detection, but they apply confidence estimates rather than the probability to detect novelties without additional calculations.

Regarding the latter strategy, one-class SVM [56], and SVDD [57] are usually applied to learn the distribution of inlier class. With the rapid development of deep learning, the current one-class classification approaches often use the deep learning model to automatically learn features from the inlier class. Pidhorskyi et al. [40] employed generative adversarial networks to learn the latent representation from the inlier class. And respect to the local coordinates of the manifold tangent space, the novelty probability is calculated. Ruff et al. [58] presented a Deep Support Vector Data Description (DSVDD) for anomaly detection, which learns valuable latent representations and minimizes the variance of inlier class. Unlike some novelty detection approaches, which only apply the reconstruction error of the input. Kim et al. [41] extended the reconstruction error concept in the input space to the hidden space. Li et al. [59] applied a self-supervised learning approach to learn latent representation from normal data and a one-class classifier is used to detect anomalous data based on the latent representation.

In this paper, to learn the richer fine-grained semantic information from the behavioral biometrics of smartphone users, we propose a latent representation masked one-class autoencoder for the continuous authentication task. Unlike [42], [43], [44], [45], [46], [47], [48], [49], which randomly mask a high proportion of the input. However, random masking may fail to fully cover important structures of the input, leading to suboptimal representations, which is particularly problematic for continuous authentication tasks based on behavioral biometrics. In this paper, we mask the optimal part of the low dimensional latent representation generated by the Encoder. Compared with the original input, the generated latent representation contains more abstract and discriminative features, and then the Reconstructor can learn richer fine-grained semantic information in reconstructing the masked latent representation.

III. Threat Model

We assume that the attacker successfully gains access to the user's smartphone in the following two ways: 1) Crack the one-time authentication mechanism of the smartphone through some attack methods, such as smudge attack [1], side channel attack [2], and shoulder surfing attack [3]; 2) Successfully circumvent the one-time authentication mechanism of the smartphone by some methods, for example, the user left the smartphone in some public places after passing login authentication, such as toilets, restaurants, buses, etc. After gaining access to the smartphone, personal privacy information on the smartphone are available for attackers throughout the session, and the random attackers attempt to use their own behavioral biometrics or behavioral biometrics from the public dataset to attack the continuous authentication system. Before the attacks, the random attackers have no knowledge of the legitimate users' behavioral patterns when interacting with their smartphones.

IV. Latent Representation Masked One-Class Autoencoder

Anomaly score based on reconstruction error has been widely applied as a base for one-class classification task. In test phase, the reconstruction error of the normal samples is low, and the reconstruction error of the anomalous samples will be high because the trained autoencoder does not know the distribution of anomalous samples. Some variants of autoencoder are proposed to learn richer semantic information from normal samples [38], [39], [42], [43], [60], [61].

As shown in Fig. 2, the proposed latent representation masked one-class autoencoder consists of two parts: the masked latent representation generator (MLRG) and the Reconstructor. The MLRG generates a mask to maximize the reconstruction error, and the Reconstructor attempts to minimize the reconstruction error with the masked latent representation. The MLRG and Reconstructor are both constructed with a convolutional autoencoder. The standard autoencoder usually consists of an Encoder and a Decoder. The Encoder encodes high-dimensional inputs into low-dimensional hidden variables, forcing the neural network to learn the most informative features. The Decoder is used to restore the hidden variables in the hidden layer to the initial dimension and try to perfectly or approximately restore the original input. To clearly distinguish Encoder and Decoder in MLRG and Reconstructor, we use Encoder_M and Decoder_M to represent Encoder and Decoder in the MLRG, and the Encoder_R and Decoder_R to define Encoder and Decoder in the Reconstructor.

A. Masked Latent Representation Generator

As shown in the left part of Fig. 2, the masked latent representation generator (MLRG) is constructed with an autoencoder network. The MLRG is trained with the loss function as follow:

$$L_{\text{mlrg}} = \left\| X - \tilde{X} \right\|^2 \tag{1}$$

where the X is the input and the $\tilde{X}$ is the reconstruction of input with the MLRG.

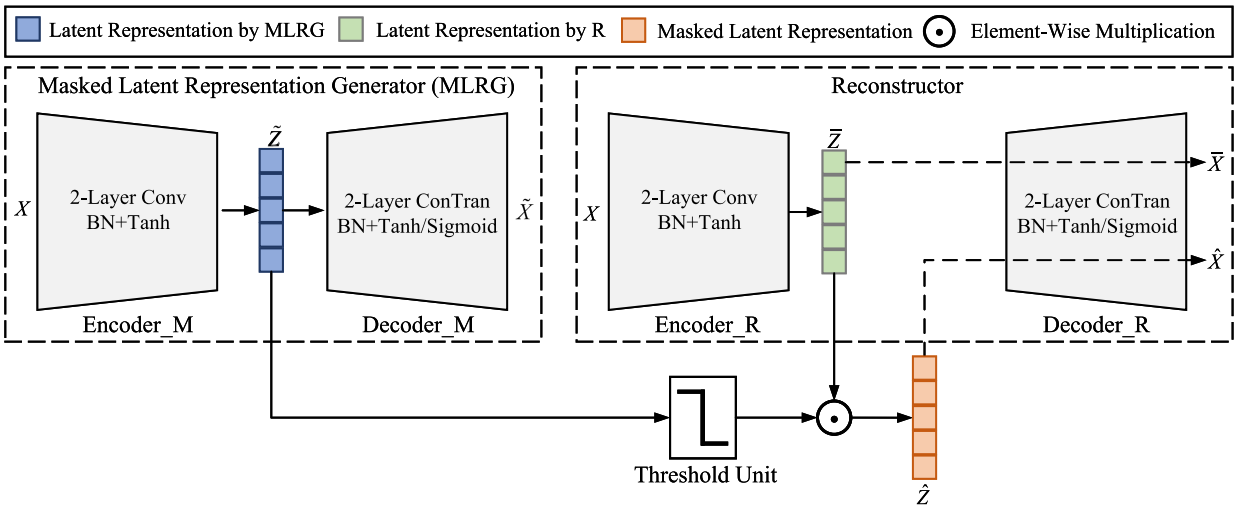


Fig. 2. The architecture of the proposed latent representation masked one-class autoencoder. The masked latent representation generator (MLRG) is used to generate a mask, which can mask the important part of the latent representation of the input. The Reconstructor is used to reconstruct the input based on the masked latent representation. The standard autoencoder consists of an Encoder and a Decoder. To clearly distinguish the Encoder and Decoder in MLRG and Reconstructor, we use Encoder_M and Decoder_M to represent Encoder and Decoder in the MLRG, and the Encoder_R and Decoder_R to represent Encoder and Decoder in the Reconstructor.

The MLRG aims to generate the low dimensional latent representation $\tilde{Z} = Encoder\_M(X)$ with discriminative semantic information of the input. And then the $\tilde{Z}$ as activation map is input to the threshold unit to generate a binary mask:

$$M_i = \begin{cases} 0, & \text{if } \tilde{Z}_i \geq \tau \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

where the M is the mask matrix, and the τ is threshold. Based on the learned latent representation from the MLRG, we apply an appropriate percentile (87.5%) value as the threshold τ to compute and mask the optimal part. We get the optimal percentile through many experiments.

Based on the generated mask matrix M, we perform a masking operation on the low dimensional representation $\bar{Z} = Encoder\_R(X)$ generated by the Reconstructor, and then generate the masked latent representation $\hat{Z}$:

$$\hat{Z} = M \odot \bar{Z} \tag{3}$$

where the $\odot$ is the element-wise multiplication operation. During the masking process, the element that is 0in the mask matrix M makes the corresponding element in latent representation matrix $\bar{Z}$ become 0; otherwise, the element in corresponding position in $\bar{Z}$ will remain unchanged.

B. Reconstructor

As shown in the right part of Fig. 2, the Reconstructor is also an autoencoder network aiming to reconstruct high-quality input. Unlike the common autoencoder, which reconstructs the input only with the reconstruction loss as follow:

$$L_{\text{rec}} = \left\| X - \bar{X} \right\|^2 \tag{4}$$

where the $\bar{X}$ is the reconstruction of the input with the Reconstructor.

In this paper, we also reconstruct the input with masked latent representation $\hat{Z}$, the loss function can be formulated as follow:

$$\begin{aligned} \hat{X} &= Decoder\_R(\hat{Z}), \\ L_{\text{mask}} &= \left\| X - \hat{X} \right\|^2 \end{aligned} \tag{5}$$

The Reconstructor attempts to minimize the reconstruction loss based on the masked latent representation $\hat{Z}$. Since the masked latent representation is masked out of the important features, the Reconstructor needs to learn richer semantic information from the input to minimize the reconstruction loss.

C. Implementation Details

The lightweight latent representation masked one-class autoencoder consists of a masked latent representation generator (MLRG) and a Reconstructor, and both of them are constructed with a convolutional autoencoder. They share the same autoencoder architecture but do not share training parameters. The detailed architecture of the shared autoencoder network is shown in Fig. 3, which is lightweight enough and effective. The Encoder contains two convolutional layers, and the batch normalization and the tanh activation operations follow each convolutional layer. The Decoder has two transposed convolutional layers, each transposed convolutional layer followed by the batch normalization and the tanh activation operations, except for the last transposed convolutional layer. The last layer applies the sigmoid activation operation instead of the tanh activation operation. During the training phase, the overall objective loss function can be denoted as:

$$L = L_{mlrg} + L_{rec} + L_{mask} \tag{6}$$

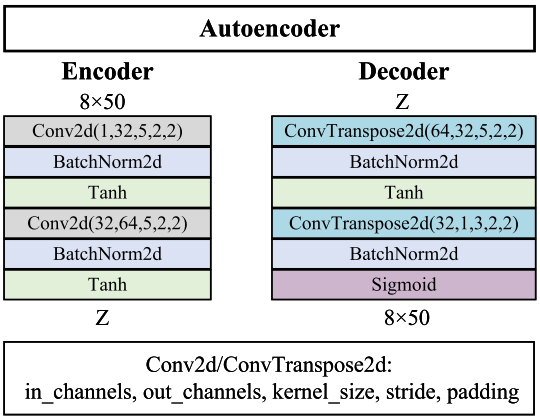


Fig. 3. The detail architecture of the shared autoencoder. Different layers are represented with different colors.

V. Continuous Authentication System

To evaluate the effectiveness of the proposed lightweight latent representation masked one-class autoencoder for continuous authentication tasks, we design and construct a continuous authentication system based on the proposed one-class autoencoder. As shown in Fig. 4, the proposed continuous authentication system consists of three parts: data preprocessing module, representation learning from normal samples module, and continuous authentication module. We will describe each module of the proposed continuous authentication system in detail below.

A. Data Preprocessing

The original behavioral biometrics data collected from the motion sensors (accelerometer and gyroscope) cannot be used directly to learn rich semantic representations based on the proposed latent representation masked one-class autoencoder. We should perform data normalization and sample generation operations on the original behavioral biometrics data.

Data normalization: The sensory data collected from different motion sensors have considerable differences in value. If we directly apply the original behavioral biometrics data to train the proposed latent representation masked one-class autoencoder, it may highlight the role of sensory data with a more significant value, while relatively weakening the role of sensory data with a smaller value. This situation may cause the proposed latent representation masked one-class autoencoder to be unable to extract discriminative semantic representation from behavioral biometric data. Therefore, we need to keep the data from different motion sensors within the same range through the data normalization operation. Besides, the data normalization operation can improve the convergence speed of the proposed one-class autoencoder network during the training phase.

In this paper, we apply the min-max normalization approach to perform a linear transformation with original sensory data. The collected sensory data from each motion sensor can be denoted as $(x, y, z) \in \mathbb{R}^3$ for each time point, where x, y, and z are the three-axis of the sensory data. Besides, to reduce the influence of noise in motion sensor data, we also add the fourth axis (Magnitude) of sensory data for each time point, which can be denoted as $ma = \sqrt{x^2 + y^2 + z^2}$. For each time point, the sensory data collected from two motion sensors (accelerometer and gyroscope) can be represented as $(x_{acc}, y_{acc}, z_{acc}, ma_{acc}, x_{gyr}, y_{gyr}, z_{gyr}, ma_{gyr}) \in \mathbb{R}^8$. The acc and gyr indicate the accelerometer and gyroscope, respectively. Then the min-max normalization operation can be denoted as:

$$b_i = \frac{a_i - \min\limits_{1 \le j \le n} \{a_j\}}{\max\limits_{1 \le j \le n} \{a_j\} - \min\limits_{1 \le j \le n} \{a_j\}} \tag{7}$$

then sensory data after normalization can be represented as $(\bar{x}_{acc}, \bar{y}_{acc}, \bar{z}_{acc}, \overline{ma}_{acc}, \bar{x}_{gyr}, \bar{y}_{gyr}, \bar{z}_{gyr}, \overline{ma}_{gyr}) \in [0, 1]$.

Sample generation: Time series sensory data collected from motion sensors cannot be directly used as input to train the latent representation masked one-class autoencoder network. We apply a time window t to segment the time series sensory data without time overlap. Sensory data after data normalization for each time point can be represented as: $(\bar{x}_{acc}, \bar{y}_{acc}, \bar{z}_{acc}, \overline{ma}_{acc}, \bar{x}_{gyr}, \bar{y}_{gyr}, \bar{z}_{gyr}, \overline{ma}_{gyr})^T \in \mathbb{R}^8$. Then for a time window t, the time series sensory data can be formulated as a $d \times m$ matrix:

$$S = \begin{bmatrix} \bar{x}^1_{acc} & \bar{x}^2_{acc} & \cdots & \bar{x}^{m-1}_{acc} & \bar{x}^m_{acc} \\ \bar{y}^1_{acc} & \bar{y}^2_{acc} & \cdots & \bar{y}^{m-1}_{acc} & \bar{y}^m_{acc} \\ \bar{z}^1_{acc} & \bar{z}^2_{acc} & \cdots & \bar{z}^{m-1}_{acc} & \bar{z}^m_{acc} \\ \overline{ma}^1_{acc} & \overline{ma}^2_{acc} & \cdots & \overline{ma}^{m-1}_{acc} & \overline{ma}^m_{acc} \\ \bar{x}^1_{gyr} & \bar{x}^2_{gyr} & \cdots & \bar{x}^{m-1}_{gyr} & \bar{x}^m_{gyr} \\ \bar{y}^1_{gyr} & \bar{y}^2_{gyr} & \cdots & \bar{y}^{m-1}_{gyr} & \bar{y}^m_{gyr} \\ \bar{z}^1_{gyr} & \bar{z}^2_{gyr} & \cdots & \bar{z}^{m-1}_{gyr} & \bar{z}^m_{gyr} \\ \overline{ma}^1_{gyr} & \overline{ma}^2_{gyr} & \cdots & \overline{ma}^{m-1}_{gyr} & \overline{ma}^m_{gyr} \end{bmatrix} \tag{8}$$

where $d = 8$ is the number of axes of sensory data at each time point, and $m = t \times f$. The sensory data is collected with a sample rate $f = 100$HZ, and we apply a time window $t = 0.5s$ to segment the sensory data. Then the training samples used to train the proposed one-class autoencoder have a shape of 8×50. Besides, we also evaluate continuous authentication performance with different time windows, which will be described in Section VI-D.

B. Representation Learning From Normal Samples

We apply the proposed lightweight latent representation masked one-class autoencoder to learn rich fine-grained semantic information from the legitimate user's behavioral biometrics, which is trained only with normal samples in an unsupervised manner. We have described the proposed lightweight latent representation masked one-class autoencoder in detail in Section IV. In the authentication phase, we calculate the hidden layer reconstruction error of the query sample as the anomaly score.

C. Continuous Authentication

Common anomaly score calculation methods based on the autoencoder only apply the query sample's reconstruction error as an anomaly score to classify anomaly samples and normal samples. In this paper, we also consider the hidden layers reconstruction error when calculating the anomaly score of the

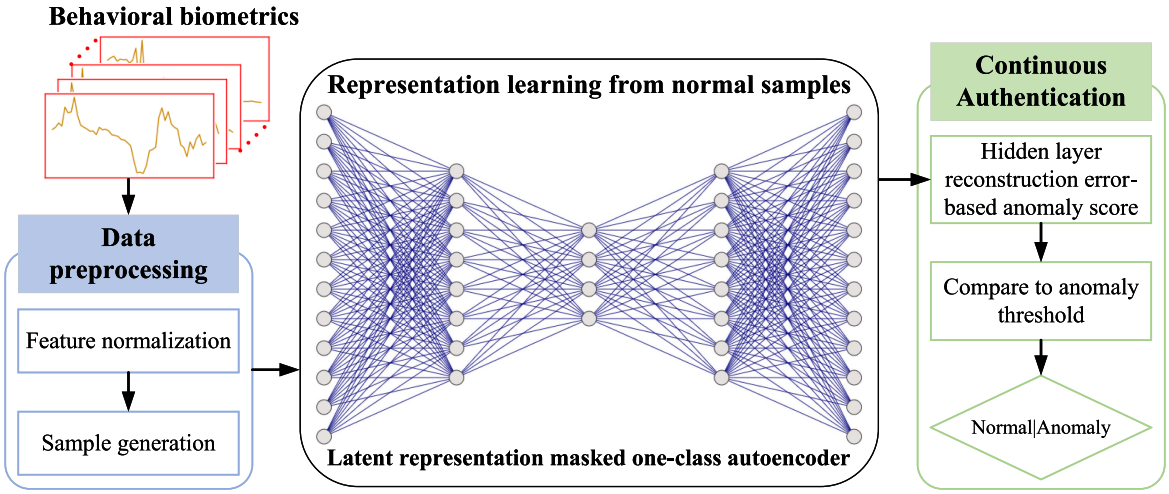


Fig. 4. The architecture of the proposed continuous authentication system consists of three parts: data preprocessing module, representation learning from normal samples module and continuous authentication module.

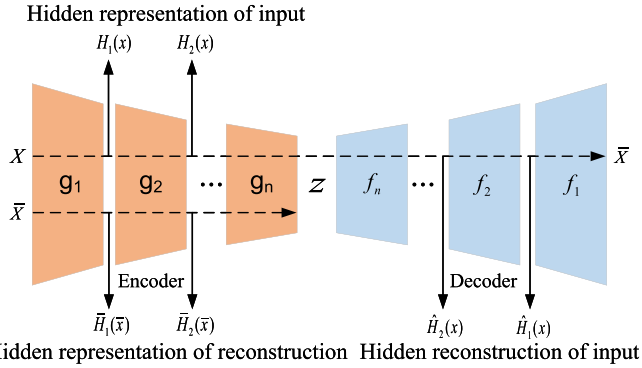


Fig. 5. Hidden space representation of the proposed latent representation masked one-class autoencoder.

query sample in the continuous authentication phase. As shown in Fig. 5, let $H_i(x)$ be the representation of the $i-th$ hidden layer when the input x is fed to the encoder g. Let $\bar{H}_i(\bar{x})$ be the representation of the $i-th$ hidden layer when the reconstruction $\bar{x}$ is fed to the encoder g. Let $\hat{H}_i(x)$ be the reconstruction of the $i-th$ hidden layer when the latent representation z is fed to the Decoder f. The f and g can also be computed as:

$$\begin{aligned} f &= f_1 \bullet f_2 \cdots \bullet f_n \\ g &= g_n \cdots \bullet g_2 \bullet g_1 \end{aligned} \tag{9}$$

where n is the number of hidden layers in the encoder and decoder. Then the $H_i(x)$, $\bar{H}_i(\bar{x})$, can be expressed as:

$$\begin{aligned} \mathrm{H}_i(x) &= g_{:i}(x) \\ \bar{H}_i(\bar{x}) &= g_{:i}(\bar{x}) \end{aligned} \tag{10}$$

and defining $\mathcal{M} = \{f(g(x)), x \in \mathbb{R}^m\}$, $M_i = \{g_{:i}(x)) : x \in \mathcal{M}\}$,

$$\begin{aligned} \forall x \in \mathcal{M}, x &= f(g(x)) \\ \forall k \in M_i, k &= (g_i \bullet f_i)(k) \end{aligned} \tag{11}$$

the g_i and f_i are inverse function of each other. Then the $\hat{H}_i(x)$ can be expressed as:

$$\hat{H}_i(x) = (f_{n:i+1} \bullet g_{i+1:})(H_{:i}(x)) \tag{12}$$

then for every $x \in \mathcal{M}$:

$$\begin{aligned} \hat{H}_i(x) &= (f_{n:i+1} \bullet g_{i+1:})(H_{:i}(x)) \\ &= (f_{n:i+1} \bullet g)(x) \\ &= (g_{:i} \bullet f_{n:i+1} \bullet f_{i:} \bullet g)(x) \\ &= (g_{:i} \bullet f \bullet g)(x) \\ &= g_{:i}(\bar{x}) = \bar{H}_i(\bar{x}) \end{aligned} \tag{13}$$

then the hidden layer space reconstruction error of the autoencoder can be formulated as:

$$L_{H-rec} = \sum_{i=0}^{n} \| H_i(x) - \bar{H}_i(\bar{x}) \|^2 \tag{14}$$

Instead of using the reconstruction error of a query sample, we apply (14) to calculate the anomaly score of the query sample. In the continuous authentication phase, the anomaly score of the normal samples from legitimate users is low, and the anomaly score of the anomalous samples from the attacker will be high because the trained latent representation masked one-class autoencoder does not know the distribution of anomalous samples.

VI. Experiments

The proposed continuous authentication system is evaluated on four public datasets. The main goals of the comprehensive experiments are as follows: 1) Experimental results in details on four public datasets; 2) Comparison with representative continuous authentication methods; 3) Comparison with representative autoencoder-based one-class classification methods for continuous authentication task; 4) Experimental results with different time windows; 5) Ablation studies are conducted to demonstrate the effectiveness of each module of our proposed latent representation masked one-class autoencoder architecture; 6) Time efficiency analysis; 7) Computational efficiency analysis; 8) Anti-attack analysis.

A. Experiment Setting

The proposed continuous authentication system is implemented using Python and PyTorch frameworks. The latent representation masked one-class autoencoder is trained on the Tesla T4 GPU, and the operating system version is Ubuntu 16.04.1. The learning rates of the encoder and decoder are 0.0001 and 0.001, respectively. The batch size is set to 32. The behavioral biometric data collection App is developed with *Android Studio* basic development environment.

B. Evaluation Metrics

We apply the area under the ROC curve (AUC), False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER) to evaluate the effectiveness of the proposed continuous authentication system. We calculate the FAR, FRR, and EER with a TPR (True Positive Rate) equal to or greater than 97.0% in our experiments.

False Acceptance Rate (FAR): FAR refers to the probability of the anomalous samples from attackers predicted to be the normal samples from legitimate users, which can be calculated as: $\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{TN+FP}}$.

False Rejection Rate (FRR): FRR refers to the probability of the normal samples from legitimate users predicted to be anomalous samples from attackers, which can be calculated as: $\mathrm{FRR} = \frac{\mathrm{FN}}{\mathrm{TP+FN}}$.

Equal Error Rate (EER): EER is defined as the rate when the FAR equals the FRR.

Besides, True Positive (TP) means the number of normal samples from legitimate users predicted to be normal samples. False Positive (FP) means the number of anomalous samples from attackers predicted to normal samples. True Negative (TN) means the number of anomalous samples from attackers predicted to anomalous samples. False Negative (FN) means the number of normal samples from legitimate users predicted to be anomalous samples.

C. Datasets

Comprehensive experiments are performed on four public datasets, which can be used for the continuous authentication task. In this paper, we only extract the behavioral biometric data from two motion sensors (accelerometer and gyroscope) based on four public datasets.

HMOG dataset: Sitová et al. [4] introduce the HMOG (Hand Movement, Orientation, and Grasp) behavioral biometrics for continuous authentication task, which contains 100 participants' (53 male and 47 female) behavioral biometric data. During behavioral biometrics data collection, they considered two scenarios (sitting and walking) with 24 sessions (8 reading sessions, 8 writing sessions, and 8 map navigation sessions).

BrainRun dataset: BrainRun [62] consists of three parts: 1) Touch biometrics collected from the registered users; 2) Information regarding registered users, information regarding registered devices, and information regarding all the games played; 3) Behavioral biometric data collected from embedded motion sensors (accelerometer, gyroscope, and magnetometer, etc.).

IDNet dataset: IDNet [63] contains 50 participants' gait biometric data, which are collected from embedded sensors (accelerometer, gyroscope, magnetometer, etc.) using Android smartphones. The participants were asked to walk as they felt comfortable when smartphones were worn in the right front pocket of their trousers.

WISDM dataset: WISDM [64] collects the behavioral biometrics data of 51 participants from motion sensors (accelerometer and gyroscope). The participants are asked to perform 18 diverse activities of daily living during the data collection phase.

Some smartphone users' behavioral biometric data are unavailable because the sensory data of two motion sensors (accelerometer and gyroscope) are not simultaneously collected at a certain time point or period during the data collection process. In addition, less behavioral biometric data collected from some smartphone users resulted in insufficient samples to train the proposed latent representation masked one-class autoencoder. Table I lists the ID of participants used for continuous authentication experiments. As shown in Table I, the number of participants from each dataset is 90, 29, 42, and 35, respectively. We perform the experiments to evaluate the effectiveness of the proposed continuous authentication system with 196 smartphone users in total.

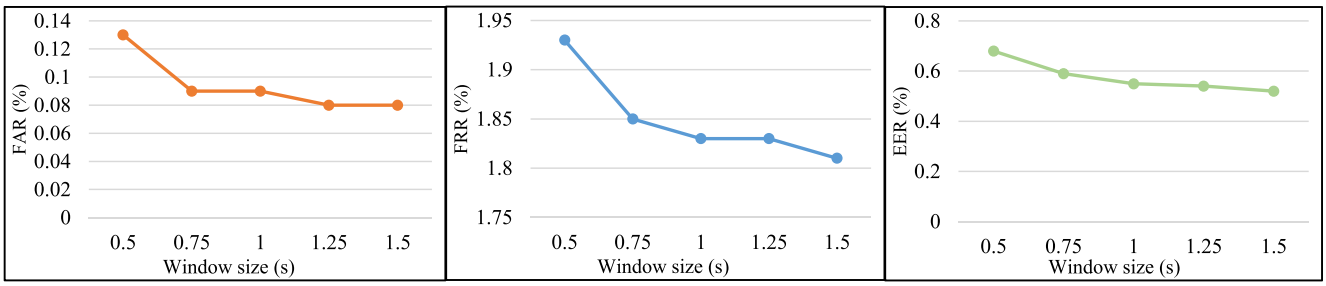
D. Experimental Results With Different Time Windows

We evaluate the authentication performance of the proposed authentication system with the time window changing from 0.5s to 1.5s. We first apply time windows of different sizes (0.5s, 0.75s, 1s, 1.25s, and 1.5s) to segment the sensory readings to generate training samples with different dimensions. Then based on the generated training samples with different dimensions (8×50, 8×75, 8×100, 8×125 and 8×150), we train the continuous authentication system individually and report the average experimental results of multiple smartphone users on the HMOG dataset. For each smartphone user, 80% of behavioral biometrics is carried out as normal samples to train the continuous authentication system, and the remaining 20% of behavioral biometrics is used for testing. Anomalous samples of attackers are selected from the remaining smartphone users.

Fig. 6 shows the trend of FAR, FRR, and EER with the increase of time window sizes. As shown in Fig. 6, the proposed authentication system achieves excellent experimental results of FAR (less than 0.15%), FRR (less than 2%), and EER (less than 1%) with different time window size. The FAR decreases slightly with the increase of time window size from 0.5s to 0.75s and then oscillates in a small area with the increase of time window size from 0.75s to 1.5s. The FRR and EER show the same trend as the FAR. Table II lists the experimental results of FAR, FRR, and EER in detail on different time window size. During the authentication phase, we can achieve better authentication performance with the increase in time window sizes. However, the time overhead of each authentication will also increase with the increase in time window size. After weighing the authentication performance and time overhead of each authentication, we

TABLE I
ID OF PARTICIPANTS SELECTED FROM EACH DATASET

HMOG									
663153	923862	973891	962159	207696	803262	579284	151985	815316	862649
100669	220962	398248	578526	264325	892687	966655	776328	553321	395129
893198	621276	501973	368258	366286	785899	540641	657486	893255	278135
998757	675397	872895	554303	489146	733568	879155	808022	865501	180679
863985	342329	980953	897652	326223	186676	984799	918136	336172	622852
556357	876011	865881	352716	472761	698266	913228	527796	431312	937904
785873	588087	525584	693572	396697	717868	218719	257279	990622	751131
771782	248252	201848	763813	745224	720193	827212	277905	261313	538363
240168	856302	171538	594887	777078	710707	561993	799296	405035	841866
BrainRun									
l1doqao	frl4fzq	9gykp6g	sxvkh3b	9gx7uks	n0g0zsv	b0602fr	6jtbpdh	677l8bq	ta0ko40
y2opfq8	r13q05q	w764hxe	vo441ru	d99p79w	7ksck80	z3u2vz	gzx7rv	lm8ujtt	vk43uke
lvrishm	ioxyr9y	68n9ll	w8f2wrs	pxipu7n	uui53he	x8rbf3x	8xjh8a	fit126a	
IDNet									
u023	u018	u028	u009	u037	u024	u030	u011	u043	u047
u001	u008	u026	u003	u036	u029	u019	u006	u040	u033
u022	u016	u050	u031	u002	u012	u025	u034	u013	u038
u020	u048	u004	u042	u015	u041	u017	u014	u032	u046
u049	u027								
WISDM									
1647	1649	1609	1624	1645	1629	1614	1622	1600	1637
1634	1606	1607	1635	1601	1603	1612	1608	1625	1628
1631	1613	1638	1620	1623	1640	1633	1621	1632	1618
1615	1616	1610	1648	1644					



(a) The trend of FAR (b) The trend of FRR (c) The trend of EER

Fig. 6. The trend of experimental results with the increase of time window size.

TABLE II
THE EXPERIMENTAL RESULTS OF FAR (%), FRR (%), AND EER (%) ON DIFFERENT TIME WINDOW SIZE

Window size (s)	FAR ↓	FRR ↓	EER ↓
0.50	0.13	1.93	0.68
0.75	0.09	1.85	0.59
1.00	0.09	1.83	0.55
1.25	0.08	1.83	0.54
1.50	0.08	1.81	0.52

TABLE III
THE MEAN PERFORMANCE OF FAR (%), FRR (%), EER (%), AND AUC ON FOUR DATASETS

Dataset	FAR ↓	FRR ↓	EER ↓	AUC ↑
HMOG	0.13	1.93	0.68	0.999
BrainRun	0.54	1.63	0.94	0.998
IDNet	1.92	2.75	2.14	0.996
WISDM	1.68	2.31	1.87	0.997

choose a 0.5s time window size of sensory readings to generate training samples in the experiments.

E. Experimental Results on Four Datasets

We evaluate the effectiveness and generality of the proposed continuous authentication system on different datasets. Different behavioral biometric datasets are collected in different environments and conditions, so experimental results on multiple datasets are more convincing. We conduct evaluation experiments using the same experimental parameters and experimental settings on four datasets. As demonstrated in Section VI-D, we apply a 0.5s time window size for sensory readings to generate training samples on four public datasets.

Table III lists the mean experimental results for all smartphone users on each dataset. As shown in Table III, the proposed

TABLE IV
THE DETAIL PERFORMANCE OF TEN RANDOMLY SELECTED USERS FROM FOUR DATASETS

Dataset	User ID	FAR↓	FRR ↓	EER ↓	Dataset	User ID	FAR ↓	FRR↓	EER↓
HMOG	973891	0.07	2.73	0.66	IDNet	u023	1.30	2.85	1.76
	962159	0.02	2.83	0.81		u028	1.50	2.98	1.91
	100669	0.17	2.75	1.07		u009	1.43	2.85	1.78
	220962	0.02	1.38	0.66		u024	1.09	2.59	1.43
	398248	0.11	2.82	0.71		u019	0.58	2.81	2.19
	278135	0.56	2.89	1.28		u040	1.87	2.20	1.96
	733568	0.17	2.81	1.27		u025	1.74	2.93	2.17
	785873	0.30	2.83	1.22		u020	1.37	2.91	2.15
	277905	0.38	2.93	1.09		u017	0.95	2.75	2.20
	594887	0.24	2.19	0.41		u027	0.68	2.86	1.85
BrainRun	frl4fzq	0.98	2.86	2.83	WISDM	1647	0.13	2.83	1.15
	9gykp6g	0.41	1.83	1.76		1609	0	1.74	0.37
	sxvkh3b	0	2.00	0.05		1629	2.58	2.90	2.68
	n0g0zsv	0.03	1.69	0.05		1634	1.07	2.73	2.39
	6jtbpdh	2.35	2.44	2.41		1623	1.37	2.84	1.80
	677l8bq	0.97	2.60	2.36		1635	1.94	2.90	2.46
	y2opfq8	0.05	1.67	0.21		1621	2.70	2.90	2.81
	vo441ru	0.10	2.99	0.97		1618	0	2.70	1.82
	8xih8a	0	1.45	1.35		1616	1.60	2.99	1.90
	fit126a	0	1.72	1.67		1644	0	1.95	0.24

continuous authentication system achieves the authentication performance with 0.13% FAR, 1.93% FRR, 0.68% EER, and 0.999 AUC on the HMOG dataset, and 0.54% FAR, 1.63% FRR, 0.94% EER, and 0.998 AUC on the BrainRun dataset. It also can attain the authentication performance with 1.92% FAR, 2.75% FRR, 2.14% EER and 0.996 AUC on the IDNet dataset, and 1.68% FAR, 2.31% FRR, 1.87% EER and 0.997 AUC on the WISDM dataset. Furthermore, to demonstrate the authentication performance of different smartphone users in detail, we report the detailed authentication performance of ten randomly selected smartphone users with a 0.5s time window size. Without loss of generality, we apply the *random.sample* function[2] to randomly select ten smartphone users from four dataset, respectively. Table IV lists the experimental results of ten randomly selected users in detail from each datasets. As shown in Table IV, the proposed continuous authentication system can achieve excellent performance with FAR (less than 3%), FRR (less than 3%), and EER (less than 3%) for each randomly selected smartphone user on four datasets.

F. Comparison With Representative Methods

Most representative continuous authentication methods perform performance evaluation experiments on different datasets, making it difficult for us to evaluate the performance of different representative continuous authentication methods on common datasets. In this paper, to make a quantitative performance comparison, we reproduce some representative continuous authentication methods on four datasets, which are based on conventional machine learning algorithms [4], [15], [17] or deep learning [19], [20], [22], [27], [54]. The HMM, OC-SVM, and IF indicate the Hidden Markov Model, One-Class Support Vector Machine, and Isolation Forest, respectively. The deep learning-based deep feature extraction model in [19], [20], [22], [27], [54] are implemented with the PyTorch framework, and the one-class classifiers are from the scikit-learn library.

Table V lists the EER of different representative authentication methods on four datasets. The experimental results of our reported representative authentication methods may deviate somewhat from the results reported in their papers, that is due to different data preprocessing methods, behavioral biometrics data sources, and experiment parameter settings. As shown in Table V, compared with eight baselines, the proposed continuous authentication system achieves excellent authentication performance on four datasets, which is trained only with normal samples from legitimate users in an unsupervised manner. Baseline [19] obtains the best competitive authentication performance among eight baselines. However, baseline [19] needs a variety of abnormal samples to train the deep feature extraction model to improve its generalization ability and requires huge computational overhead, which needs to be pre-trained with GPU.

G. Comparison With Representative Autoencoder-Based One-Class Classification Methods

We evaluate the effectiveness of representative autoencoder-based one-class classification methods for continuous authentication tasks, such as DSVDD [58][3], GPND [40][4], OCGAN [38][5], RAPP [41][6], and OLED [39][7], which were originally developed for image novelty detection task. Due to the difference in the dimension of training samples, we need to partially modify the network architectures of these

[2]https://docs.python.org/3.7/library/random.html

[3]https://github.com/lukasruff/Deep-SVDD-PyTorch
[4]https://github.com/podgorskiy/GPND
[5]https://github.com/PramuPerera/OCGAN
[6]https://github.com/Aiden-Jeon/RaPP
[7]https://github.com/jewelltaylor/OLED

TABLE V
THE AUTHENTICATION PERFORMANCE OF THE EER (%) OF DIFFERENT REPRESENTATIVE CONTINUOUS AUTHENTICATION METHODS ON FOUR DATASETS

Baseline study	Classifiers	Training data	Dataset			
			HMOG	BrainRun	IDNet	WISDM
Sitová et al. (2015) [4]	Scaled Manhattan	Owner	16.34	18.64	19.07	18.13
Shen et al. (2017) [15]	HMM	Owner	8.46	10.65	10.79	11.93
Li et al. (2020) [20]	OC-SVM	Owner & Impostor	5.85	7.21	7.93	6.76
Cherifi et al. (2021) [17]	HMM	Owner	15.49	17.86	18.31	16.71
Li et al. (2021) [22]	IF	Owner & Impostor	5.07	6.83	7.45	6.15
Wang et al. (2021) [27]	Deep Metric Learning	Owner & Impostor	4.13	5.21	6.97	4.64
Shen et al. (2022) [54]	DeSVDD	Owner	6.97	5.63	11.15	7.81
Hu et al. (2023) [19]	OC-SVM	Owner & Impostor	1.06	1.25	**1.19**	2.98
Our work	Masked Autoencoder	Owner	**0.68**	**0.94**	2.14	**1.87**

TABLE VI
EXPERIMENTAL RESULTS OF REPRESENTATIVE AUTOENCODER-BASED ONE-CLASS CLASSIFICATION METHODS FOR CONTINUOUS AUTHENTICATION TASK ON FOUR DATASETS

Methods	Dataset	FAR ↓	FRR ↓	EER ↓	AUC ↑	Dataset	FAR ↓	FRR ↓	EER ↓	AUC ↑
DSVDD [58] (2018)	HMOG	8.82	2.89	3.53	0.987	BrainRun	7.02	1.89	2.94	0.992
GPND [40] (2018)		14.19	2.89	5.67	0.978		14.24	2.04	4.78	0.974
OCGAN [38] (2019)		17.15	**1.49**	4.08	0.988		15.45	2.06	3.22	0.989
RAPP-AE [41] (2019)		2.10	2.65	2.13	0.996		31.83	2.17	13.33	0.927
OLED [39] (2022)		19.19	2.84	12.15	0.906		6.81	1.92	3.37	0.992
Our method		**0.13**	1.93	**0.68**	**0.999**		**0.54**	**1.63**	**0.94**	**0.998**
DSVDD [58] (2018)	IDNet	23.28	2.91	10.64	0.952	WISDM	16.81	2.73	6.25	0.980
GPND [40] (2018)		8.42	2.83	2.19	0.982		35.99	2.76	9.09	0.951
OCGAN [38] (2019)		28.85	**1.56**	11.36	0.947		57.27	**1.55**	9.35	0.946
RAPP-AE [41] (2019)		45.93	2.91	21.94	0.837		31.97	2.92	9.17	0.962
OLED [39] (2022)		8.23	2.81	5.09	0.985		23.05	2.80	7.87	0.973
Our method		**1.92**	2.75	**2.14**	**0.996**		**1.68**	2.31	**1.87**	**0.997**

representative one-class classification methods, including DSVDD [58], GPND [40], OCGAN [38].

To demonstrate the superiority of our proposed latent representation masked one-class autoencoder compared to representative autoencoder-based one-class classification methods for continuous authentication tasks, we perform comprehensive experiments on the HMOG dataset, BrainRun dataset, IDNet dataset, and WISDM dataset, respectively. Table VI lists the mean experimental results of all smartphone users on four public datasets. As shown in Table VI, representative autoencoder-based one-class classification methods achieve excellent authentication performance. Our proposed lightweight latent representation masked one-class autoencoder outperforms these representative methods with 0.68% EER, 0.94% EER, 2.14% EER, and 1.87% EER on four public datasets, respectively.

We can explain this phenomenon from the following aspects: 1) The characteristics of training data are different. Since these representative works based on autoencoders were designed for image anomaly detection tasks, they are able to perform well in image anomaly detection but may not work as effectively in continuous authentication tasks based on behavioral biometrics. This is because the textural features of images differ greatly from behavioral biometric data, and the performance of autoencoders are highly sensitive to the characteristics of training data; 2) The data preprocessing methods are different. The preprocessing of behavioral biometric data differs from the image preprocessing methods used in image anomaly detection tasks. To reduce the impact of noise and outliers in the behavioral biometric data, we perform data representation processing on the behavioral biometric data. In image anomaly detection tasks, apart from performing data normalization on image data, image augmentation operations are also carried out, such as random scaling, horizontal or central cropping, color transformation, contrast adjustment, and random flipping. Since behavioral biometric data are time-correlated, these image augmentation operations could disrupt the internal correlations within behavioral biometric data; 3) The performance of one-class autoencoder architectures not only depends on the network structure itself but also heavily relies on specific training strategies, such as the optimizer, data augmentation, and regularization strategies. Due to the significant differences between the textural features of images and the structural characteristics of behavioral biometric data, the training strategies tailored for image anomaly detection tasks may not work well in continuous authentication tasks.

We can demonstrate that the proposed latent representation masked one-class autoencoder can better learn the fine-grained and rich semantic information from behavioral biometric data.

H. Ablation Studies

In this section, we study the contributions of each module in our proposed latent representation masked one-class

TABLE VII
ABLATION STUDY OF THE PROPOSED LATENT REPRESENTATION MASKED ONE-CLASS AUTOENCODER ON FOUR DATASETS

Module	Dataset	FAR ↓	FRR ↓	EER ↓
Reconstructor	HMOG	1.97	3.81	2.51
	BrainRun	2.54	3.57	2.92
	IDNet	3.76	4.75	4.33
	WISDM	3.31	4.32	3.81
Reconstructor + MLRG	HMOG	0.64	2.47	1.10
	BrainRun	1.24	2.25	1.47
	IDNet	2.58	3.36	2.88
	WISDM	2.07	3.01	2.39
Full Model	HMOG	0.13	1.93	0.68
	BrainRun	0.54	1.63	0.94
	IDNet	1.92	2.75	2.14
	WISDM	1.68	2.31	1.87

TABLE VIII
COMPUTATIONAL COST AND MODEL SIZE COMPARISON WITH REPRESENTATIVE LIGHTWEIGHT NEURAL NETWORK ARCHITECTURES

Methods	Input size	M-Adds	Params
ShuffleNetV2 [65] (2018)	8×50	7.85M	3.50M
MobileNetV2 [66] (2018)	8×50	7.77M	3.50M
MnasNet [67] (2019)	8×50	14.23M	6.28M
MobileNetV3-Large [68] (2019)	8×50	8.48M	5.48M
MobileNeXt [69] (2020)	8×50	30.72M	11.68M
CA [70] (2021)	8×50	8.98M	3.95M
Our method	8×50	**6.88M**	**0.98M**

autoencoder, namely MLRG and hidden layers reconstruction error, to overall authentication performance. We evaluate the effect of each module on HMOG dataset, BrainRun dataset, IDNet dataset, and WISDM dataset. Specifically, we consider three scenarios. In the first scenario, we only consider the Reconstructor with reconstruction error. In the second scenario, we use the MLRG and the Reconstructor. In the final scenario, we consider the full latent representation masked one-class autoencoder (Reconstructor, MLRG, and hidden layers reconstruction error). Experimental results on four datasets are summarized in Table VII.

As shown in Table VII, we observe that the authentication performance is already high only with the Reconstructor. So even the slightest improvement from this point is remarkable. When the proposed MLRG is introduced to the Reconstructor, the authentication performance of the EER is improved by more than 1.4% on four datasets. When the hidden layers reconstruction error is introduced, the performance of the EER is further improved by 0.4% on four datasets. Overall, the MLRG and hidden layers reconstruction error indeed help to improve the authentication performance of our continuous authentication model.

I. Time Efficiency

Time efficiency is an important indicator to evaluate the continuous authentication system. The shorter time required for each authentication means that attackers can be detected in time, thus avoiding the leakage of personal privacy data on smartphones. The time overhead of the proposed continuous authentication system consists of three parts: 1) A certain time window of sensory data readings to generate training samples ($t1$); 2) The time overhead of data preprocessing ($t2$); 3) The time overhead of hidden layers reconstruction error-based anomaly score calculation during the continuous authentication phase ($t3$).

This paper applies a 0.5s time window to segment the sensory readings to generate the training sample ($t1 = 0.5s$). The time overhead of data preprocessing is 20 ms ($t2 = 20$ ms), and the authentication system needs 50 ms to calculate the hidden layers reconstruction error-based anomaly score in each authentication ($t3 = 50$ ms). Then the time overhead of each authentication is 0.57s ($t1 + t2 + t3$). We demonstrate that the proposed continuous authentication system can rapidly detect attackers during the continuous authentication phase.

J. Computational Efficiency Analysis

Neural networks have been widely used in many fields, such as image classification, object detection, anomaly detection, etc. Generally, accuracy is improved by designing deeper neural network architectures, while the computational overhead is ignored. However, considering the limited computing capacity of mobile platforms, it's difficult to deploy state-of-the-art neural network architectures on mobile platforms. We conduct computational efficiency evaluation to verify whether the proposed latent representation masked one-class autoencoder is suitable for deployment on mobile platforms and embedded platforms. We measure the computational cost and neural network model size using the number of multiply-adds (M-Adds) and parameter size (Params). Besides, we compare with state-of-the-art lightweight neural network architectures, such as ShuffleNetV2 [65], MobileNetV3 [68], MnasNet [67], MobileNetV2 [66], CA [70], MobileNeXt [69], which are tuned to mobile and embedded platforms.

As shown in Table VIII, compared with these representative lightweight neural network architectures, the proposed latent representation masked one-class autoencoder has slightly less computation and parameters. We can demonstrate that the proposed latent representation masked one-class autoencoder is lightweight enough for mobile platforms and embedded platforms.

K. Anti-Attack Analysis

As shown in Table III, the proposed continuous authentication system achieves excellent performance on four datasets. However, it does not mean that the proposed continuous authentication system can well resist various attacks from attackers because the distribution of anomalous samples from attackers is diverse and unknown in the real-world environment. This paper considers three kinds of attacks: 1) Attackers attempt to use their behavioral biometrics to attack legitimate users' smartphones; 2) Attackers attempt to imitate the behavioral patterns of legitimate users to attack their smartphones; 3) Attackers attempt to apply

TABLE IX
THE MEAN EXPERIMENTAL RESULTS OF RANDOM ATTACKS WITH THREE ATTACK SCENARIOS IN THE REAL-WORLD ENVIRONMENT

Attack scenario	FAR ↓	FRR ↓	EER ↓
WeChat	1.46	1.71	1.64
TikTok	2.35	1.73	1.91
Website	1.26	1.65	1.42

TABLE X
THE MEAN EXPERIMENTAL RESULTS OF RANDOM ATTACKS USING DIFFERENT ATTACK SAMPLES FROM THREE PUBLIC DATASETS

Attack samples	FAR ↓	FRR ↓	EER ↓
BrainRun	0.04	1.84	1.65
IDNet	0.004	1.91	1.74
WISDM	0.0001	1.85	1.69

behavioral biometrics from public datasets to launch random attacks.

Attackers attempt to use their behavioral biometrics to attack legitimate users' smartphones. To collect behavioral biometric data of attackers when attacking legitimate users' smartphones, we develop a behavioral biometric collection App based on the Android system. As shown in Fig. 1, attackers interact with legitimate users' smartphones using three interaction scenarios (WeChat, TikTok, Website), which are common application scenarios of smartphone users in their daily lives.

- *WeChat:* It is a free application that provides instant messaging, public platform, friends circle and other services for intelligent terminals, which involve frequent key operations for information editing.
- *TikTok:* It is an application for publishing and browsing videos for all ages, and users mainly interact with the TikTok application by sliding up and down on the touchscreen.
- *Website:* Users can search and browse interesting information through website. The interaction process involves simple sliding up and down and complex key operations to complete information retrieval.

We invite 30 volunteers (10 females and 20 males) to participate in the random attack experiment and sign an agreement with them to collect their behavioral biometric data for the attack experiments. Smartphones used by volunteers include Huawei, Xiaomi and Oppo. For each round of experiments (30 rounds of experiments in total), we randomly select one volunteer as the legitimate user, and the rest volunteers acted as attackers. Each attacker uses three different interaction scenarios (WeChat, TikTok, and Website) to interact with legitimate users' smartphones in turn, and we apply the developed data collection App (shown in Fig. 1) to collect behavioral biometrics of attackers in the background.

We report the mean experiment results of the proposed continuous authentication system under random attacks in the real-world environment. As shown in Table IX, our proposed continuous authentication system can still achieve 1.64% EER, 1.91% EER, and 1.42% EER on three interaction scenarios when attacked by multiple real-world attackers. We achieve lower FAR and EER on the WeChat and Website scenarios because users need to use frequent and more complex operations to interact with WeChat and Website applications, such as key operations. And complex key operations are more conducive to our continuous authentication system to learn the behavioral patterns of legitimate users.

Impersonation Attack. Before carrying out the attack, attackers try to imitate the behavioral patterns of legitimate smartphone users interacting with their smartphones by directly observing legitimate smartphone users' actions while they are inputting information.

In this paper, we invite 20 volunteers to participate in impersonation attack experiments and sign an agreement with them to collect their behavioral biometric data for the impersonation attack experiments. We conduct 20 rounds of experiments, with each volunteer performing the experiment as a legitimate user. In each round of the impersonation experiment, we randomly selected one of the 20 volunteers to act as the legitimate user, while the remaining 19 participants assumed the role of attackers to carry out the impersonation attack. In the impersonation attack experiments, attackers initially observe the legitimate users' actions during their interaction with smartphones to learn their behavioral patterns. Subsequently, the attackers replicate these patterns to interact with the smartphones of the legitimate users. During the impersonation attack phase, we utilize the developed data collection App to collect the behavioral biometrics of attackers in the background. We evaluate the effectiveness of our proposed continuous authentication system in detecting impersonation attacks by calculating the FAR. The experimental results show that our proposed continuous authentication system attains an average FAR of 2.17% during impersonation attacks. We can demonstrate that the attacker is unable to substantially improve the success rate of attacks by mimicking the behavioral patterns of the legitimate user.

Using behavioral biometrics from public datasets to launch attacks. We choose smartphone users from the HMOG dataset as legitimate users to train our authentication system, and we apply the behavioral biometrics from three datasets (BrainRun dataset, IDNet dataset, and WISDM dataset) as anomalous samples to attack the continuous authentication system. Behavioral biometrics from different datasets are collected under different conditions (lab environment or unconstrained real-world environment), which can well reflect the distribution diversity of anomalous samples from attackers.

During the training phase, we randomly select one smartphone user's behavioral biometrics from the HMOG dataset (90 smartphone users) as normal samples to train the continuous authentication system. We perform 90 rounds (90 legitimate users) of experiments and report the mean experimental results. Table X lists the mean experimental results of the proposed continuous authentication system under random attacks from three datasets. As shown in Table X, we can achieve the mean performance of 0.04% FAR, 0.004% FAR, and 0.0001% FAR when we apply anomalous samples from three datasets to attack the continuous authentication system, respectively.

Experimental results demonstrate that the proposed continuous authentication system can resist random attacks and

impersonation attack from various unknown distributions of attack samples from attackers in three attack scenarios.

VII. Discussion

Only learn to model the distribution of normal samples. A long time of deep learning-based continuous authentication schemes [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33] are trained as a binary or multi-class classification task, which needs abnormal samples from attackers to train authentication models. It means that they can only detect attacks after learning the distribution of abnormal samples from attackers. However, we can't learn the distribution of abnormal samples a priori in the real-world environment because the distribution of abnormal samples from attackers is diverse and unknown. In this paper, the proposed continuous authentication system is trained only with normal samples in an unsupervised manner, which detects the attacker through the reconstruction error of the sample. We can detect unknown attackers more accurately in the real-world environment.

Evaluation of the superiority of the proposed one-class classification algorithm. We reproduce some representative continuous authentication methods on four public datasets (Section VI-F), and make detailed performance evaluation and comparison. More importantly, as we all know, the advanced autoencoder-based one-class classification algorithms based on deep learning will be first proposed and applied in computer vision. In this paper, we reproduce some representative autoencoder-based one-class classification algorithms [38], [39], [40], [41], [58] for the continuous authentication task (Section VI-G), which were originally developed for novelty detection task in the computer vision field. As shown in Table VI, experimental results demonstrate that the proposed latent representative masked one-class autoencoder achieves excellent authentication performance among these advanced autoencoder-based one-class classification methods.

Lightweight enough for mobile platform. Deep learning-based continuous authentication schemes [20], [21], [22], [23], [28], [31], [32] improve the accuracy of authentication through more complex and deep neural network architectures. However, the more complex and deeper neural network architecture means a lot of computational overhead and prediction latency, which are difficult to deploy on mobile and embedded platforms. This paper proposes a lightweight latent representation masked one-class autoencoder, which is suitable for deploying on platforms with limited computing power (Section VI-J), such as smartphones and embedded platforms.

Comprehensive attacks analysis in real-world environment. Current continuous authentication schemes [4], [6], [9], [10], [15], [20], [21], [22], [23], [28], [30], [31], [32], [33], [54] based on classical machine learning or deep learning pay more attention to improving authentication performance on their datasets, while ignoring usability of these proposed authentication schemes in the real-world environment. They do not evaluate the ability to resist various attacks of attackers in the real-world environment. To fill this gap, we develop a behavioral biometric data collection App based on the Android system (shown in Fig. 1) and evaluate the ability of the proposed continuous authentication scheme to resist three kinds of attacks in the real-world environment (Section VI-K).

Limited analysis of behavioral biometrics based on multiple sessions and multiple modalities. Although our proposed continuous authentication scheme has achieved satisfactory performance on four public datasets, there remain the following limitations in the preprocessing of behavioral biometric data: 1) The temporal orders between multiple sessions have not been considered; 2) Authentication performance evaluation based on multimodal behavioral biometrics has not been conducted, such as the fusion of motion movement patterns and touch dynamics. Like [6], [15], [20], [21], [22], we have only focused on the evaluation of continuous authentication performance based on motion movements. In our future work, we will concentrate on evaluating authentication performance within the context of multiple sessions and multimodal behavioral biometrics. Because this approach is inherently more complex, as it necessitates the consideration of a variety of conditions. Moreover, it more accurately reflects the challenges that are encountered in real-world scenarios.

VIII. Conclusion

In this paper, we propose a lightweight latent representation masked one-class autoencoder to learn the distribution of behavioral biometrics from legitimate users, which is trained only with normal samples in an unsupervised manner. More importantly, we apply a masked latent representation generator (MLRG) to generate the low dimensional latent representation with discriminative semantic information, and then the generated latent representation is used as a mask to cover the optimal part of the latent representation generated from the encoder of the Reconstructor. The Reconstructor can learn discriminative and fine-grained deep features from behavioral biometrics data of legitimate users by reconstructing the input based on the masked latent representation. Besides, we construct a continuous authentication system based on the proposed lightweight latent representation masked one-class autoencoder. Experimental results demonstrate that the proposed continuous authentication system achieves excellent authentication performance of 0.68% EER, 0.94% EER, 2.14% EER, and 1.87% EER on four datasets, respectively. Additionally, it is highly resilient to random attacks from unknown distribution of anomalous samples in the real-world environment.

References

[1] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," in *Proc. 4th USENIX Workshop Offensive Technol.*, 2010, pp. 1–7.

[2] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "Accessory: Password inference using accelerometers on smartphones," in *Proc. 12th Workshop Mobile Comput. Syst. Appl.*, 2012, pp. 1–6.

[3] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan, "Shoulder surfing defence for recall-based graphical passwords," in *Proc. 7th Symp. Usable Privacy Secur.*, 2011, pp. 1–12.

[4] Z. Sitová et al., "HMOG: New behavioral biometric features for continuous authentication of smartphone users," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 5, pp. 877–892, May 2015.

[5] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, "Silentsense: Silent user identification via touch and movement behavioral biometrics," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 187–190.
[6] Y. Li, H. Hu, and G. Zhou, "Using data augmentation in continuous authentication on smartphones," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 628–640, Feb. 2019.
[7] S. Vhaduri and C. Poellabauer, "Multi-modal biometric-based implicit authentication of wearable device users," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 12, pp. 3116–3125, Dec. 2019.
[8] W.-H. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *Proc. Int. Conf. Inf. Syst. Secur. Privacy*, 2015, pp. 1–11.
[9] Y. Li, B. Zou, S. Deng, and G. Zhou, "Using feature fusion strategies in continuous authentication on smartphones," *IEEE Internet Comput.*, vol. 24, no. 2, pp. 49–56, Mar./Apr. 2020.
[10] C. Shen, Y. Zhang, X. Guan, and R. A. Maxion, "Performance analysis of touch-interaction behavior for active smartphone authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 3, pp. 498–513, Mar. 2016.
[11] W.-H. Lee and R. B. Lee, "Sensor-based implicit authentication of smartphone users," in *Proc. 47th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw.*, 2017, pp. 309–320.
[12] M. W. A. El-Soud, T. Gaber, F. AlFayez, and M. M. Eltoukhy, "Implicit authentication method for smartphone users based on rank aggregation and random forest," *Alexandria Eng. J.*, vol. 60, no. 1, pp. 273–283, 2021.
[13] R. Dave et al., "Hold on and swipe: A touch-movement based continuous authentication schema based on machine learning," 2022, *arXiv:2201.08564*.
[14] T. P. Thao, "Location-based behavioral authentication using GPS distance coherence," 2020, *arXiv: 2009.08025*.
[15] C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion, "Performance analysis of multi-motion sensor behavior for active smartphone authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 1, pp. 48–62, Jan. 2018.
[16] A. Roy, T. Halevi, and N. Memon, "An HMM-based multi-sensor approach for continuous mobile authentication," in *Proc. IEEE Mil. Commun. Conf.*, 2015, pp. 1311–1316.
[17] F. Cherifi, M. Omar, and K. Amroun, "An efficient biometric-based continuous authentication scheme with hmm prehensile movements modeling," *J. Inf. Secur. Appl.*, vol. 57, 2021, Art. no. 102739.
[18] S. K. Al Kork, I. Gowthami, X. Savatier, T. Beyrouthy, J. A. Korbane, and S. Roshdi, "Biometric database for human gait recognition using wearable sensors and a smartphone," in *Proc. 2nd Int. Conf. Bio- Eng. Smart Technol.*, 2017, pp. 1–4.
[19] M. Hu, K. Zhang, R. You, and B. Tu, "AuthConFormer: Sensor-based continuous authentication of smartphone users using a convolutional transformer," *Comput. Secur.*, vol. 127, 2023, Art. no. 103122.
[20] Y. Li, H. Hu, Z. Zhu, and G. Zhou, "Scanet: Sensor-based continuous authentication with two-stream convolutional neural networks," *ACM Trans. Sensor Netw.*, vol. 16, no. 3, pp. 1–27, 2020.
[21] Y. Li, P. Tao, S. Deng, and G. Zhou, "Deffusion: CNN-based continuous authentication using deep feature fusion," *ACM Trans. Sensor Netw.*, vol. 18, no. 2, pp. 1–20, 2021.
[22] Y. Li, J. Luo, S. Deng, and G. Zhou, "CNN-based continuous authentication on smartphones with conditional wasserstein generative adversarial network," *IEEE Internet Things J.*, vol. 9, no. 7, pp. 5447–5460, Apr. 2022.
[23] T. Zhu, Z. Weng, G. Chen, and L. Fu, "A hybrid deep learning system for real-world mobile user authentication using motion sensors," *Sensors*, vol. 20, no. 14, 2020, Art. no. 3876.
[24] L. He, C. Ma, C. Tu, and Y. Zhang, "Gait2Vec: Continuous authentication of smartphone users based on gait behavior," in *Proc. IEEE 25th Int. Conf. Comput. Supported Cooperative Work Des.*, 2022, pp. 280–285.
[25] X. Zeng, X. Zhang, S. Yang, Z. Shi, and C. Chi, "Gait-based implicit authentication using edge computing and deep learning for mobile devices," *Sensors*, vol. 21, no. 13, 2021, Art. no. 4592.
[26] M. P. Centeno, Y. Guan, and A. van Moorsel, "Mobile based continuous authentication using deep features," in *Proc. 2nd Int. Workshop Embedded Mobile Deep Learn.*, 2018, pp. 19–24.
[27] C. Wang, Y. Xiao, X. Gao, L. Li, and J. Wang, "A framework for behavioral biometric authentication using deep metric learning on mobile devices," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 19–36, Jan. 2023.
[28] M. Abuhamad, T. Abuhmed, D. Mohaisen, and D. Nyang, "AUToSen: Deep-learning-based implicit continuous authentication using smartphone sensors," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5008–5020, Jun. 2020.
[29] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li, "Deep learning-based gait recognition using smartphones in the wild," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3197–3212, 2020.
[30] G. Giorgi, A. Saracino, and F. Martinelli, "Using recurrent neural networks for continuous authentication through gait analysis," *Pattern Recognit. Lett.*, vol. 147, pp. 157–163, 2021.
[31] S. Amini, V. Noroozi, A. Pande, S. Gupte, P. S. Yu, and C. Kanich, "DeepAuth: A framework for continuous user re-authentication in mobile apps," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 2027–2035.
[32] S. Mekruksavanich and A. Jitpattanakul, "Deep learning approaches for continuous authentication based on activity patterns using mobile sensing," *Sensors*, vol. 21, no. 22, 2021, Art. no. 7519.
[33] H. C. Volaka, G. Alptekin, O. E. Basar, M. Isbilen, and O. D. Incel, "Towards continuous authentication on mobile phones using deep learning models," *Procedia Comput. Sci.*, vol. 155, pp. 177–184, 2019.
[34] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[35] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3379–3388.
[36] Y. Zhang et al., "Adversarially learned one-class novelty detection with confidence estimation," *Inf. Sci.*, vol. 552, pp. 48–64, 2021.
[37] Y. Xia, X. Cao, F. Wen, G. Hua, and J. Sun, "Learning discriminative reconstructions for unsupervised outlier removal," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1511–1519.
[38] P. Perera, R. Nallapati, and B. Xiang, "OCGAN: One-class novelty detection using GANs with constrained latent representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2898–2906.
[39] J. T. Jewell, V. R. Khazaie, and Y. Mohsenzadeh, "One-class learned encoder-decoder network with adversarial context masking for novelty detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, pp. 3591–3601.
[40] S. Pidhorskyi, R. Almohsen, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6823–6834.
[41] K. H. Kim et al., "RaPP: Novelty detection with reconstruction along projection pathway," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–14.
[42] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2536–2544.
[43] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16000–16009.
[44] C. Wei, H. Fan, S. Xie, C.-Y. Wu, A. Yuille, and C. Feichtenhofer, "Masked feature prediction for self-supervised visual pre-training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14668–14678.
[45] C. Feichtenhofer et al., "Masked autoencoders as spatiotemporal learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 35946–35958.
[46] S. Woo et al., "ConvNext V2: Co-designing and scaling convnets with masked autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 16133–16142.
[47] Z. Tong, Y. Song, J. Wang, and L. Wang, "VideoMAE: Masked autoencoders are data-efficient learners for self-supervised video pre-training," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 10078–10093.
[48] Y. Pang, W. Wang, F. E. Tay, W. Liu, Y. Tian, and L. Yuan, "Masked autoencoders for point cloud self-supervised learning," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2022, pp. 604–621.
[49] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, "ViTPose: Simple vision transformer baselines for human pose estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 38571–38584, 2022.
[50] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 136–148, Jan. 2013.
[51] C. Shen, Y. Zhang, Z. Cai, T. Yu, and X. Guan, "Touch-interaction behavior for continuous user authentication on smartphones," in *Proc. Int. Conf. Biometrics*, 2015, pp. 157–162.
[52] W. Meng, Y. Wang, D. S. Wong, S. Wen, and Y. Xiang, "TouchWB: Touch behavioral user authentication based on web browsing on smartphones," *J. Netw. Comput. Appl.*, vol. 117, pp. 1–9, 2018.

[53] P. S. Teh, N. Zhang, S.-Y. Tan, Q. Shi, W. H. Khoh, and R. Nawaz, "Strengthen user authentication on mobile devices by using user's touch dynamics pattern," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 10, pp. 4019–4039, 2020.
[54] Z. Shen, S. Li, X. Zhao, and J. Zou, "MMAuth: A continuous authentication framework on smartphones using multiple modalities," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 1450–1465, 2022.
[55] A. Buriro, B. Crispo, F. Delfrari, and K. Wrona, "Hold and sign: A novel behavioral biometrics for smartphone user authentication," in *Proc. IEEE Secur. Privacy Workshops*, 2016, pp. 276–285.
[56] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
[57] D. M. Tax and R. P. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, 2004.
[58] L. Ruff et al., "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2018, pp. 4393–4402.
[59] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "Cutpaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9664–9674.
[60] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.," *J. Mach. Learn. Res.*, vol. 11, no. 12, 2010.
[61] X. Yan, H. Zhang, X. Xu, X. Hu, and P.-A. Heng, "Learning semantic context from normal samples for unsupervised anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 3110–3118.
[62] M. D. Papamichail, K. C. Chatzidimitriou, T. Karanikiotis, N.-C. I. Oikonomou, A. L. Symeonidis, and S. K. Saripalle, "BrainRun: A behavioral biometrics dataset towards continuous implicit authentication," *Data*, vol. 4, no. 2, 2019, Art. no. 60.
[63] I. Dataset, 2016. [Online]. Available: https://signet.dei.unipd.it/research/human-sensing/
[64] G. M. Weiss, "WISDM smartphone and smartwatch activity and biometrics dataset," *UCI Mach. Learn. Repository: WISDM Smartphone Smartwatch Activity Biometrics Dataset Data Set*, vol. 7, pp. 133190–133202, 2019.
[65] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.
[66] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
[67] M. Tan et al., "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2820–2828.
[68] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.
[69] D. Zhou, Q. Hou, Y. Chen, J. Feng, and S. Yan, "Rethinking bottleneck structure for efficient mobile network design," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 680–697.
[70] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722.

Mingming Hu received the MS degree in computer science and technology from Beijing Jiaotong University, in 2017, and the PhD degree from the Institute of Information Engineering, Chinese Academy of Sciences, in 2024. Currently, he is with the Yellow River Henan Bureau. He has published papers at venues, such as Computers & Security, IEEE Internet of Things Journal. His research interests include continuous authentication and deep learning.

Ding Wang (Member, IEEE) received the PhD degree in information security from Peking University, in 2017. He was supported by the "Boya Postdoctoral Fellowship" with Peking University from 2017 to 2019. He is currently a full professor with Nankai University. As the first author (or the corresponding author), he has published more than 90 papers at venues, such as IEEE S&P, ACM CCS, NDSS, *USENIX Security*, *IEEE Transactions on Dependable and Secure Computing*, and *IEEE Transactions on Information Forensics and Security*. His research has been reported by more than 200 medias, such as Daily Mail, Science Daily, Forbes, IEEE Spectrum and Communications of the ACM, appeared in the Elsevier 2017 "Article Selection Celebrating Computer Science Research in China," and resulted in the revision of the authentication guideline NIST SP800-63-2. His research interests include passwords, authentication, and provable security. He has been involved in the community as the PC chair/TPC member for more than 60 international conferences, such as NDSS 2023-2025, ACM CCS 2022, IEEE EuroS&P 2025, PETS 2022-2024, ACSAC 2020-2024, RAID 2024/2023, ACM AsiaCCS 2025/2022/2021, FC 2025, IFIP SEC 2018-2021, ICICS 2018-2024, and SPNCE 2020-2022. He has received the ACM China Outstanding Doctoral Dissertation Award, the Best Paper Award from INSCRYPT 2018, the Outstanding Youth Award of China Association for Cryptologic Research, the Young Scientist Nomination Award for Powerful Nation, and the First Prize of Natural Science Award of Ministry of Education.

Chen Li received the PhD degree in computer system architecture from the Institute of Information Engineering (IIE), Chinese Academy of Sciences (CAS), in 2023. She is currently an engineer with the Institute of Information Engineering (IIE), Chinese Academy of Sciences (CAS). Her current research interest is cloud data center security.

Yang Xu received the master's degree in oil and gas field development from the China University of Petroleum (Beijing), in 2018. He is currently working toward the PhD degree with the Institute of Information Engineering, Chinese Academy of Sciences, with a research focus on cloud data center security. His research interests include intrusion detection systems and explainable artificial intelligence.

Bibo Tu received the MS degree in computer software and theory from the Huazhong University of Science and Technology, in 2002, and the PhD degree in computer architecture from the Institute of Computing Technology, Chinese Academy of Sciences, in 2009. He is a professor with the Institute of Information Engineering, Chinese Academy of Sciences. He has published papers at venues, such as *IEEE Transactions on Dependable and Secure Computing*, *IEEE Transactions on Computers, Computers & Security*, *IEEE Internet of Things Journal*. His research interests include software defined data center and security.