

EMG-FRNet: A feature reconstruction network for EMG irrelevant gesture recognition

Wenli Zhang^{1,*}, Yufei Wang¹, Jianyi Zhang², Gongpeng Pang¹

¹ Faculty of Information Technology, Beijing University of Technology, Beijing, China;

² College of Art and Design, Beijing University of Technology, Beijing, China.

SUMMARY With the development of deep learning technology, gesture recognition based on surface electromyography (EMG) signals has shown broad application prospects in various human-computer interaction fields. Most current gesture recognition technologies can achieve high recognition accuracy on a wide range of gesture actions. However, in practical applications, gesture recognition based on surface EMG signals is susceptible to interference from irrelevant gesture movements, which affects the accuracy and security of the system. Therefore, it is crucial to design an irrelevant gesture recognition method. This paper introduces the GANomaly network from the field of image anomaly detection into surface EMG-based irrelevant gesture recognition. The network has a small feature reconstruction error for target samples and a large feature reconstruction error for irrelevant samples. By comparing the relationship between the feature reconstruction error and the predefined threshold, we can determine whether the input samples are from the target category or the irrelevant category. In order to improve the performance of EMG irrelevant gesture recognition, this paper proposes a feature reconstruction network named EMG-FRNet for EMG irrelevant gesture recognition. This network is based on GANomaly and incorporates structures such as channel cropping (CC), cross-layer encoding-decoding feature fusion (CLEDFD), and SE channel attention (SE). In this paper, Ninapro DB1, Ninapro DB5 and self-collected datasets were used to verify the performance of the proposed model. The Area Under the receiver operating characteristic Curve (AUC) values of EMG-FRNet on the above three datasets were 0.940, 0.926 and 0.962, respectively. Experimental results demonstrate that the proposed model achieves the highest accuracy among related research.

Keywords surface EMG signals, human-computer interaction, irrelevant gesture recognition, reconstruction error

1. Introduction

With the rapid development of artificial intelligence and deep learning technology, human-computer interaction through gesture recognition technology has gradually become a hot research topic nowadays. Gesture recognition based on surface electromyography (EMG) signals has advantages over vision-based gesture recognition, such as being less affected by changes in the external environmental background, requiring less computational effort, and offering higher real-time performance (1-2). The surface EMG signals are bioelectrical information obtained from the skin surface, which has the advantages of non-invasive, non-traumatic, and simple operation. The surface EMG signals directly reflect the state of muscle contraction that causes limb movements and contains rich motor information, which can realize the prediction of hand movements intention.

In recent years, pattern recognition technology based on surface myoelectricity has shown promising applications in the field of human-computer interaction, such as intelligent prostheses (3), rehabilitation exoskeletons (4), sign language interpretation (5), etc.

In practical applications of EMG interaction, the recognition of the target gesture category is crucial, along with the need to mitigate various interferences, including irrelevant gestures, electrode displacement, muscle fatigue, and user variations (6-7). Among these interferences, the presence of irrelevant gestures is a common and significant concern. The target gestures refer to the user-defined hand gestures that are used to train the classifier and facilitate human-computer interaction. The irrelevant gestures refer to the unintentional hand gestures made by users during system usage that do not belong to the predefined target categories. In such cases, the classifier is forced

to select one of the trained motions, resulting in erroneous recognition results and compromising the safety of both the device and the user (8-9). In the field of target gesture recognition, numerous studies have achieved high accuracy rates across various hand gesture actions (10-11). In our recent research at our laboratory, Zhang *et al.* (12) proposed the LST-EMG-Net model, which further improves the accuracy of EMG-based gesture recognition. Therefore, the focus of this study lies in the recognition of irrelevant gestures. Existing methods in the field of irrelevant gesture recognition can be mainly categorized into probability-based approaches (13-17) and one-vs-all classification rule-based approaches (18-21).

Probability-based methods: The core idea of this type of method is to effectively differentiate target and irrelevant samples by comparing the classifier's predicted probability values for the test samples with a preset probability threshold. Specifically, the classifier calculates a predicted probability value for the test sample, and if the predicted probability value is higher than the preset threshold, the sample is classified as a target sample; otherwise, it is classified as an irrelevant sample. Scheme *et al.* (13) proposed a method based on linear discriminant analysis (LDA) that generates a confidence score for each decision, providing the ability to reject decisions with scores below a threshold. Robertson *et al.* (14) found limitations in the confidence features of LDA and used support vector machine (SVM) confidence scores to make correct decisions. Tomczynski *et al.* (15) used the entropy function output by an artificial neural network classifier as a loss function and as a criterion for accepting or rejecting gestures. Bao *et al.* (16) generated confidence scores based on the posterior probability of the CNN, estimating the probability of each output of the classifier is correct. Zhou *et al.* (17) proposed a two-layer classifier that combines Gaussian mixture model (GMM) and k-nearest neighbor (KNN) models. The classifier determines that a gesture is irrelevant when the output probabilities of both layers are below a predefined threshold. The probability-based method has the advantages of simple principle and low implementation cost. However, the classification probabilities of many target samples may be low, while those of irrelevant samples may be high. This ultimately affects the accuracy of irrelevant actions discrimination.

Methods based on one-vs-all classification rules: The core idea of this type of method is to train a one-class classifier for each target class, achieving effective discrimination between target and irrelevant samples. Specifically, the test sample is input into all classifiers to obtain binary classification results for each classifier, which are used to determine whether the test sample belongs to the corresponding target class of that classifier. If the test sample does not belong to any known target class, it is classified as irrelevant, otherwise, it is classified as a target sample. Ding *et al.* (18) used a set

of classifiers composed of one-class Gaussian classifiers (GC) to determine whether the input sample belongs to the irrelevant class. The purpose of the Gaussian classifier is to fit a Gaussian distribution to samples belonging to the same target class. Ding *et al.* (19) used a set of classifiers composed of one-class support vector data description (SVDD) to exclude irrelevant motion interference. The purpose of SVDD is to find a minimum volume hyper-sphere to enclose samples belonging to the same target class. Wu *et al.* (20-21) used a set of classifiers composed of one-class autoencoder (AE) to address irrelevant motion interference. The purpose of AE is to reconstruct the input and judge whether the sample belongs to the target class based on the relationship between the reconstruction error and the threshold. Among the methods based on one-vs-all classification rules, some simple machine learning methods such as GC and SVDD are used to distinguish irrelevant gestures. However, these methods assume that the target gestures and irrelevant gestures are significantly different in the feature space, while the gestures in practical applications are indeed unpredictable. In contrast, the AE method can more fully exploit the small differences between the target and irrelevant gestures by calculating the reconstruction error, and improve the discriminative performance and stability of the model. However, this method is currently mostly used in high-density myoelectric systems, and the AE reconstruction process is easily affected by noise, leading to the limited reconstruction performance of the model.

The detection of irrelevant actions and anomaly detection solve very similar problems. Pang *et al.* (22) pointed out that anomaly detection, also known as outlier detection or novelty detection, refers to the process of detecting instances that deviate significantly from the majority of data. This has been a persistent and active research area for decades. The AE is a well-established method in the field of anomaly detection (23-25). It is trained on normal data and during testing, the reconstruction error of abnormal data is typically much larger than that of normal data. By measuring the magnitude of the reconstruction error, it becomes possible to identify anomalous samples. However, methods based on AE and AE variants are usually susceptible to data noise presented in the training data. With the development of GAN networks, GAN-based anomaly detection has rapidly become a popular deep anomaly detection method. Schlegl *et al.* (26) proposed the AnoGAN for anomaly detection in clinically-assisted retinal diseases. AnoGAN was the first paper to use GAN for anomaly detection. Zenati *et al.* (27) subsequently introduced EGBAD for anomaly detection on handwritten digit image datasets such as MNIST and network intrusion datasets like KDD99. Akcay *et al.* (28) proposed the GANomaly method for detecting dangerous items such as guns and knives in X-ray luggage or package datasets. GANomaly achieved state-

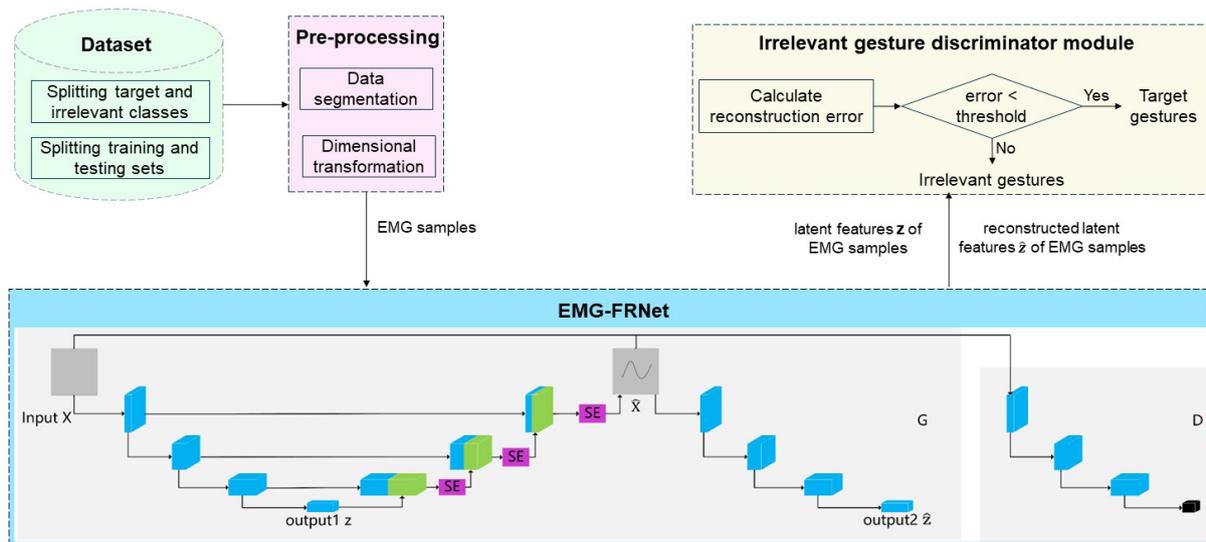


Figure 1. Overall flow chart of irrelevant gesture recognition method. EMG samples are obtained by pre-processing module; then latent features z and reconstructed latent features \hat{z} of EMG samples are obtained by EMG-FRNet module; finally, the reconstructed error between z and \hat{z} is calculated by irrelevant gesture discrimination module and compared with the threshold value to determine the class of EMG samples.

of-the-art performance in statistics and computation. Li *et al.* (29-30) implemented user authentication and improved system and device security based on the GANomaly anomaly detection method and multi-channel surface EMG signals of hand gestures. However, the detection performance of GANomaly in the field of EMG irrelevant gesture recognition still needs further exploration and improvement.

Based on the aforementioned issues, this paper proposes a feature reconstruction network named EMG-FRNet for EMG irrelevant gesture recognition. For the first time, we introduce GANomaly into EMG irrelevant gesture recognition. Building upon this, we incorporate additional structures such as channel cropping (CC), cross-layer encoding-decoding feature fusion (CLEDFFF), and SE channel attention (SE) mechanisms. These enhancements contribute to the improved performance of irrelevant gesture recognition. The main innovative points are as follows:

1) To address the problem that the input of the original GANomaly is a three-channel RGB image, while this paper is a single-channel myoelectric input, the number of feature channels in the network layer is redundant. In this paper, we propose a CC method to optimize the number of channels in the network feature layer, which can significantly reduce the number of network parameters while improving the accuracy of the network.

2) To address the issue of spatial information loss during the downsampling process in the encoder of the original GANomaly generator, which affects the decoding performance, this paper proposes the use of CLEDFFF. This method connects features of different scales from the encoding stage to the decoding stage, compensating for the lost information in the downsampling process of

the encoder. As a result, the feature maps restored by the decoder contain more low-level semantic information, leading to better reconstruction quality.

3) To address the problem that the CLEDFFF method may propagate a large amount of useless information and noise from the encoding process to the decoding layer, although it can achieve parallel transmission of feature information and improve the degree of information reuse, this paper proposes to use SE method. After the CLEDFFF concatenates the feature channels of the encoding and decoding, the decoding stage introduces channel attention mechanism. This mechanism assigns different weights to each feature channel, capturing important feature information while suppressing unimportant channels. This approach helps improve the reconstruction accuracy.

2. Materials and Methods

The overall process of the proposed method for irrelevant gesture recognition is shown in Figure 1, which consists of three modules: data pre-processing module, feature reconstruction network EMG-FRNet module, and irrelevant gesture discrimination module. They are respectively introduced in sections 2.2, 2.3, and 2.4. First, the data pre-processing module is used to obtain EMG samples by processing the datasets with data segmentation and dimensional transformation, which provides the database for training and testing of the network model. Then, the EMG-FRNet module is used to extract the latent features of the input EMG samples and reconstruct the latent features of the EMG samples, and finally output the latent features z of the EMG samples and the reconstructed latent features \hat{z} . The module trains the network using datasets from the target category and

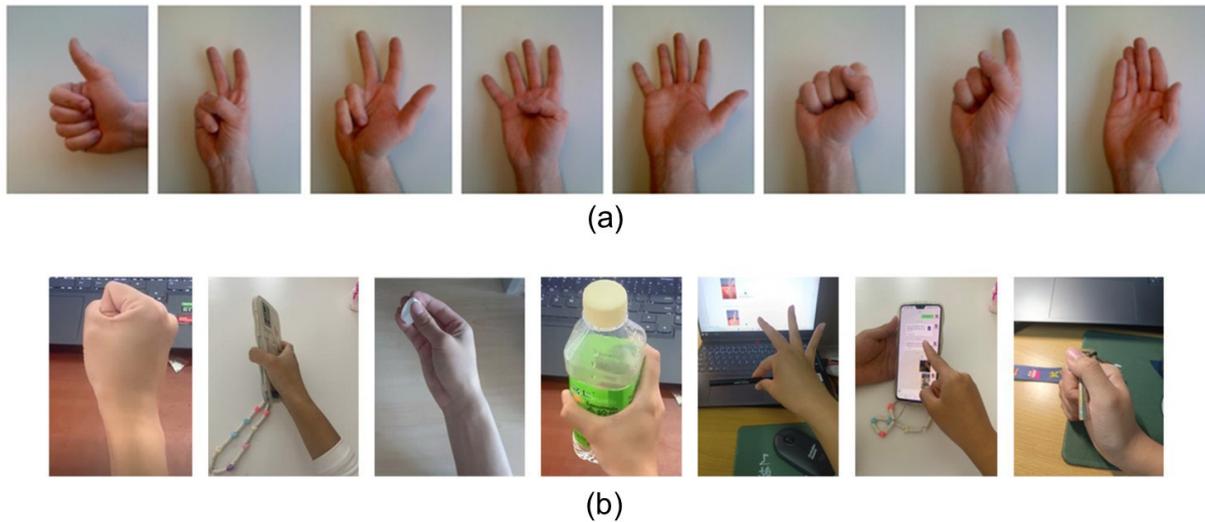


Figure 2. Types of gestures in the dataset used in this paper. (a) 8 gestures in the DB1/DB5 Exercise B dataset; **(b)** 7 gestures in the self-collected dataset.

tests the network using datasets from all categories. The network is tested with a small feature reconstruction error for the target category samples and a large feature reconstruction error for the irrelevant category samples. The feature reconstruction error is the difference between the latent feature z and the reconstructed latent feature \hat{z} . Finally, the irrelevant gesture discrimination module calculates the reconstruction error for z and \hat{z} and compares it with a predefined threshold to determine whether the EMG samples belong to the target gestures or the irrelevant gestures.

2.1. Dataset

The datasets used in this paper for EMG-FRNet are the public datasets NinaproDB1, NinaproDB5, and the self-collected dataset. The self-collected dataset was obtained from stroke patients at Beijing Rehabilitation Hospital, Capital Medical University, and informed consent was obtained from all subjects in accordance with the Declaration of Helsinki, and ethics number 2022bkky-048 was obtained.

1) Public dataset DB1 (31): Eight basic hand postures of DB1 dataset Exercise B (as in Figure 2(a)) were used, and each gesture was acquired 10 times for a total of 10 healthy subjects. Its acquisition device was a 10-channel OttoBock 13E200 with a sampling frequency of 100 Hz. The equipment is manufactured by Ottobock, Germany. 2) Public dataset DB5 (32): Eight basic hand postures of DB5 dataset Exercise B (as in Figure 2(a)) were used, and each gesture was acquired six times for a total of 10 healthy subjects. Their acquisition devices were two Myo EMG bracelets, where each Myo bracelet had 8 channels and a sampling frequency of 200 Hz. The device is manufactured by Canadian company Thalmic Labs. 3) Self-collected dataset: 7 hand movements commonly used in life (as in Figure 2(b)) were used, and each gesture was acquired 6 times for a total of

6 subjects. The acquisition device was a Myo EMG bracelet with 8 channels and a sampling frequency of 200 Hz, manufactured by Thalmic Labs, Canada.

Target category and irrelevant category data division: experiments were set up with 1 gesture as the target category action and the remaining gestures as the irrelevant category actions, traversing all possible situations. Specifically, a total of 8 experiments were conducted per subject in DB1 and DB5, and 7 experiments were conducted per subject in the self-collected dataset.

Training set and test set data division: the training set has only target category data, and the test set has both target category and irrelevant category data. Specifically, in the DB1 dataset experiments, the 1st, 3rd, 4th, 6th, 8th, 9th, and 10th gesture repetitions of the target category are used to build the training set, and the 2nd, 5th, and 7th gesture repetitions of the target category and all gesture repetitions of the irrelevant category are used to build the test set. In DB5 and self-collected dataset experiments, the 1st, 3rd, 4th, and 6th gesture repetitions of the target category are used to construct the training set, and the 2nd and 5th repetitions of the target category and all gesture repetitions of the irrelevant category are used to construct the test set.

2.2. Data pre-processing

First, for all datasets in this paper, only eight channels of EMG data were used. The specific reasons are as follows: the most important muscle group for EMG gesture recognition is concentrated around the forearm brachioradialis muscle below the elbow. Commercially available eight-channel EMG bracelets are able to cover this part of the muscle. Additionally, the configuration of such bracelets is portable and has wide practical application prospects. Therefore, this data acquisition

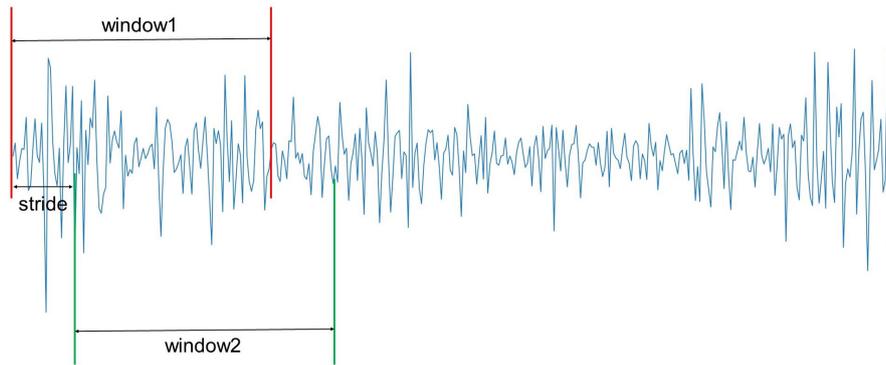


Figure 3. Schematic diagram of data segmentation. Windows1 and Windows2 represent myoelectric windows with a length of 128 sampling points, and stride represents a step size of 16 sampling points.

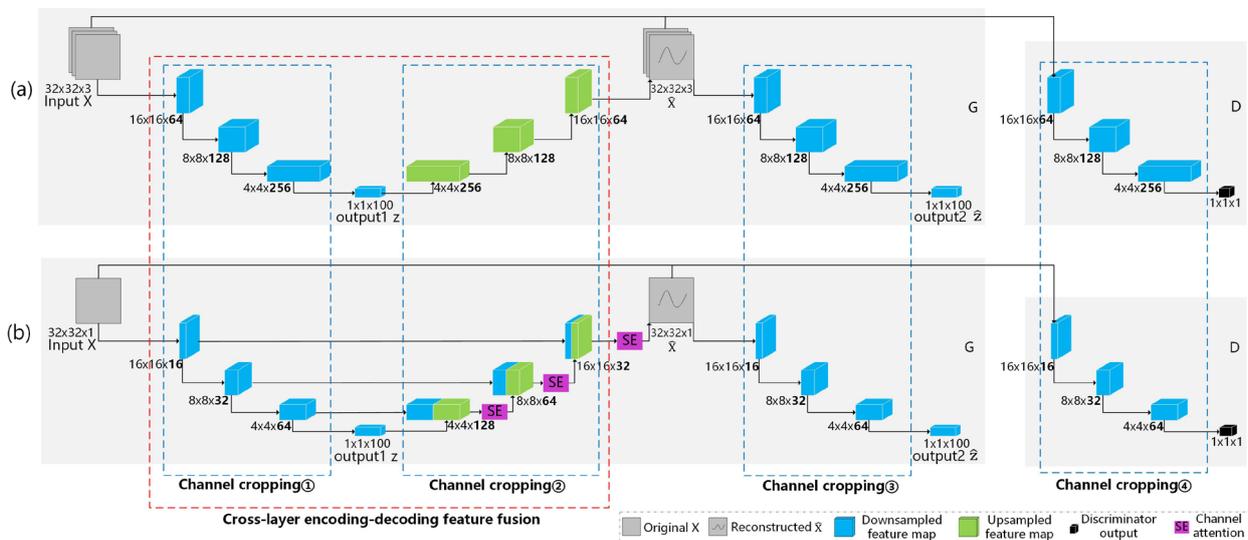


Figure 4. Model structure diagram. (a) Structure of the original GANomaly model; (b) Structure of the EMG-FRNet model.

scheme has become the first choice for many EMG gesture recognition studies.

Data segmentation: This study utilizes the sliding window method to segment the multi-channel EMG signals and obtain muscle activity samples of hand gestures. Specifically, as shown in Figure 3, we select windows of length 128 from the time series data and slide them with a fixed step size of 16. This means that multiple windows are extracted, with each window containing 128 consecutive time values. In this case, the data size is 1×128 . Considering the presence of 8 channels, the data size of the EMG window samples is 8×128 . This method can effectively segment the EMG signals to extract useful time-domain features and provide a database for subsequent model training and testing.

Data dimensional transformation: To enhance the feature learning in GAN networks, this study conducts a data dimensional transformation. Specifically, the segmented multichannel EMG samples are transformed from a 2D matrix of size 8×128 to a 2D matrix of size 32×32 , which better matches the input

format of the network.

2.3. Feature reconfiguration network EMG-FRNet model

The original GANomaly network architecture, as shown in Figure 4(a), consists of two main components: the generator G and the discriminator D. The generator network G is composed of an encoder G_{E1} , a decoder G_D , and another encoder G_{E2} , forming an "encoder-decoder-encoder" structure. Firstly, G_{E1} learns the latent features z of the input data X through downsampling. Then, G_D upsamples z to generate \hat{X} . \hat{X} is further downsampled by G_{E2} to learn the feature representation \hat{z} of \hat{X} . \hat{z} is the reconstructed latent feature of the input data. G_{E2} employs the same network structure as G_{E1} . The main purpose of the discriminator network D is to distinguish between the real input data X and the reconstructed input data \hat{X} . During training, the parameters of the generator network G and the discriminator network D are alternately updated. During testing, the discriminator D is discarded. Ultimately, in the testing phase, the network obtains the latent features z and reconstructed latent

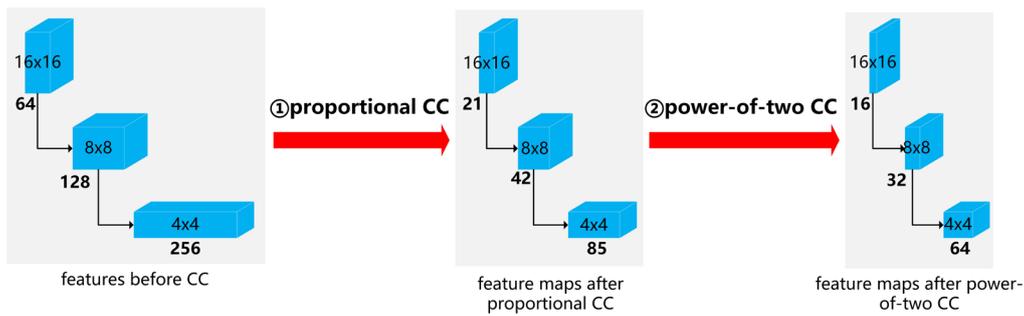


Figure 5. CC methods. Including proportional CC and power-of-two CC. Bold numbers in the figure represent the number of channels in the feature maps.

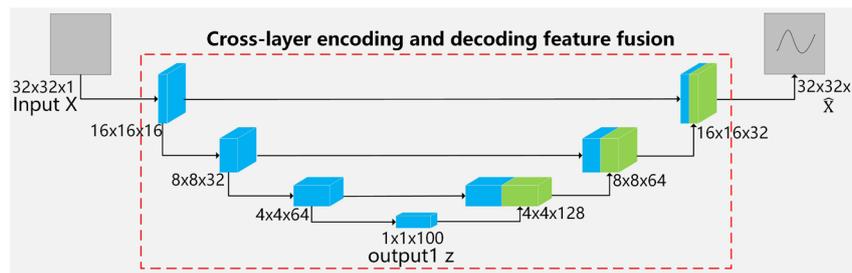


Figure 6. Schematic diagram of CLEDFF. Downsampled feature maps and upsampled feature maps for feature fusion.

features \hat{z} of the input data.

The original GANomaly network has shown good performance in image anomaly detection, but its performance still needs to be improved when applied to EMG gesture recognition. Therefore, this paper proposes a feature reconstruction network, EMG-FRNet, for EMG gesture recognition. The model structure of EMG-FRNet is shown in Figure 4(b). It is an improvement on the original GANomaly network by adding CC, CLEDFF, and SE.

(1) CC: As shown in Figure 4(a), the original GANomaly network takes in three-channel RGB images, while this paper's input is single-channel EMG samples. The network contains a large number of redundant feature channels, which increases the number of parameters in the network, makes training more difficult, and affects the performance of the network.

In order to reduce the redundant feature channels of the original GANomaly network, we propose a CC method. This method consists of two parts: proportional CC and power-of-two CC. Proportional CC reduces the number of feature channels in the network according to the channel proportion of input samples. Power-of-two CC ensures that the number of channels after cropping can still be divisible by 2, to maximize the computer's processing capability. The CC method proposed in this paper is illustrated in Figure 5. In the proportional CC part, the feature maps of the original GANomaly network are cropped according to a 3:1 ratio, which is due to the channel ratio of three-channel RGB images and single-channel EMG samples. In the power-of-two CC part, the number of channels after cropping is required to be an

integer power of two. Figure 5 shows the CC process of the downsampling feature maps. The number of channels in the downsampling feature layer is modified from (64, 128, 256) to (16, 32, 64). Similar modifications are made to the upsampling feature layer, with the number of channels changed from (256, 128, 64) to (64, 32, 16). Furthermore, the number of channels for the latent features z and \hat{z} output by the network remains at 100, as shown in Figure 4. This configuration allows for more information to be contained in the latent features z and \hat{z} , which leads to more accurate reconstruction error calculation in the irrelevant gesture recognition module in Figure 1.

(2) CLEDFF: During the downsampling encoding process G_{E1} of the original GANomaly network from the input data X to the latent feature z , the original data is compressed continuously through the use of convolution and pooling operations. This compression can result in information loss, which limits the available feature information during the upsampling decoding process G_D from the latent feature z to the reconstructed data \hat{X} . As a result, the reconstruction performance of the generator is restricted and the recovery effect of the original data is affected.

To compensate for the information loss during the downsampling process, this paper proposes a CLEDFF method. As shown in Figure 6, the feature blocks of the encoder G_{E1} are visualized as blue, and the feature blocks of the decoder G_D are visualized as green. This method adds a series of skip connections between the encoder G_{E1} and decoder G_D of the generator G . These skip connections can directly transmit the original

high-resolution feature information to the decoder, allowing effective fusion of feature information during the encoding and decoding process, thus avoiding information loss caused by downsampling. The implementation steps of this method are as follows: in the decoding and recovery process from latent feature z to \hat{X} , first, z is upsampled by $2\times$ to obtain the $4 \times 4 \times 64$ feature block of decoder G_D . Secondly, the $4 \times 4 \times 64$ feature blocks of encoder G_{E1} and decoder G_D are concatenated into a $4 \times 4 \times 128$ feature block, which is then upsampled by $2\times$ to obtain the $8 \times 8 \times 32$ feature block of decoder G_D . Then, the $8 \times 8 \times 32$ feature blocks of encoder G_{E1} and decoder G_D are concatenated into an $8 \times 8 \times 64$ feature block, which is further upsampled by $2\times$ to obtain the $16 \times 16 \times 16$ feature block of decoder G_D . Finally, the $16 \times 16 \times 16$ feature blocks of encoder G_{E1} and decoder G_D are concatenated into a $16 \times 16 \times 32$ feature block, which is upsampled by $2\times$ to obtain \hat{X} .

(3) SE: Since CLEDFD directly connects the same-scale encoding and decoding features, this connection mechanism enables cross-layer information transmission and can compensate for information loss. However, same-scale features often contain similar but not exactly the same information, so the feature information transmitted through CLEDFD may contain redundant information. The existence of this redundant feature information can reduce the network's generalization ability and may lead to overfitting problems.

To avoid the problem of redundant information in CLEDFD, this paper proposes the use of SE mechanism. As shown in Figure 7, after obtaining the concatenated feature maps M through CLEDFD, this method uses the SE mechanism to obtain the feature maps M -weight containing weight information. The SE mechanism mainly consists of three steps: Squeeze, Excitation, and Scale (33).

Squeeze: Firstly, the feature concatenation map M of the encoder and decoder is reduced in dimensionality through global average pooling, resulting in a numerical representation for each feature channel, and yielding a feature representation z . This is shown in Equation (1), where z represents the feature representation, M represents the feature concatenation map, and H , W , and C represent the height, width, and number of channels of the feature concatenation map.

$$z = \text{Squeeze}(M) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W M(i, j) \quad (1)$$

Excitation: Next, the feature representation z is non-linearly transformed and mapped into a weight vector s . This process is accomplished through two fully connected layers, where different numerical values in s represent the weight information of different channels. As shown in Equation (2), where s represents the weight vector, W_1 represents the parameters of the first fully connected layer, Relu is the activation function of the first fully connected layer, W_2 represents the parameters of the second fully connected layer, and Sigmoid is the activation function of the second fully connected layer.

$$s = \text{Excitation}(z) = \text{sigmoid}(W_2 \text{Relu}(W_1 z)) \quad (2)$$

Scale: Finally, the weight vector s is applied to the original feature maps M to obtain the weighted feature maps M -weight. Specifically, the feature concatenation map is weighted by multiplying it with the weight vector s generated in the third step, resulting in a feature concatenation map containing weight information. As shown in Equation (3), where M -weight represents the weighted feature maps.

$$M - \text{weight} = \text{Scale}(M, s) = M \times s \quad (3)$$

2.4. Irrelevant gesture discriminator module

The process of this module is shown in Figure 1. Firstly, the reconstruction error between the feature vectors z and \hat{z} is calculated using L2 distance, as shown in equation (4). Then, the reconstruction error is compared with a pre-defined threshold value. If the reconstruction error is greater than the threshold, it is classified as an irrelevant gesture; otherwise, it is classified as the target gesture. The equation for the classification is shown in equation (5), where 0 represents the target gesture and 1 represents the irrelevant gesture.

$$\text{error} = \left\| z - \hat{z} \right\|_2 \quad (4)$$

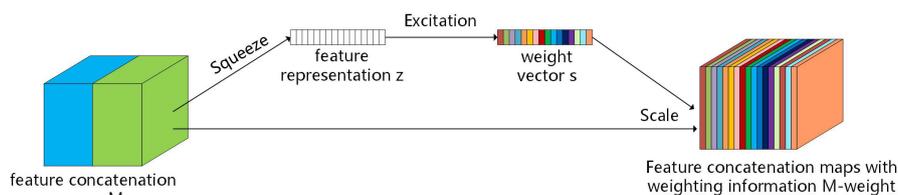


Figure 7. SE Diagram. The Squeeze operation obtains the feature representation z for each channel of the feature concatenation map M ; the Excitation operation obtains the weight s for each channel; the Scale operation obtains the feature concatenation map M -weight containing weight information.

$$\text{label} = \begin{cases} 0, & \text{error} \leq \text{threshold} \\ 1, & \text{error} > \text{threshold} \end{cases} \quad (5)$$

The principles for threshold selection are as follows: Firstly, the reconstruction errors of the feature vectors from all samples in the test set are sorted. Secondly, these values are iterated in ascending order to serve as the current threshold. Subsequently, the samples are predicted as either belonging to the target category or the irrelevant category based on the threshold. Next, by combining the predicted labels with the true labels, the false positive rate (FPR) and true positive rate (TPR) are calculated for each threshold. Finally, the optimal threshold is determined by selecting the value at which the difference between the TPR and FPR is minimized, ensuring the best trade-off between TPR and FPR.

3. Results

3.1. Experimental environment and parameter settings

The computer configuration used in the experiments of this study is as follows: an Intel Core i5-8250U CPU processor (with 8GB of memory), an NVIDIA GeForce 940MX graphics card (with 2GB of memory), and the Windows 10 operating system. The network model was trained and tested using the Python 3.7 programming language in the PyTorch 1.2.0 deep learning framework.

The training parameter settings of the EMG-FRNet: Adam optimizer was used with a smoothing constant (β_1, β_2) of (0.5, 0.999), the initial learning rate lr was set to $2e-3$, the batch size was set to 64, and the number of training epochs was set to 200.

3.2. Evaluation Indicators

In this paper, the area under the Receiver Operating Characteristic (ROC) curve (AUC) is used to evaluate the performance of the model. AUC values range from 0 to 1, and the larger the AUC value, the better the performance of the model.

The confusion matrix is the basis for drawing the ROC curve. In the confusion matrix, TP represents true positive, indicating that the sample's true class is positive and the model recognition result is also positive. Similarly, FN represents false negative, FP represents false positive, and TN represents true negative (34).

Based on the formulas (6) and (7), the FPR and TPR can be obtained. FPR represents the ratio of negative samples that are incorrectly classified as positive. TPR represents the ratio of positive samples that are correctly classified as positive.

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

For binary classification tasks, a fixed threshold can be set to obtain a (FPR, TPR) pair. By plotting the (FPR, TPR) pairs corresponding to different thresholds on a coordinate system, the ROC curve can be obtained. The ROC curve represents the recognition performance of the model under different thresholds.

Furthermore, this paper uses the AUC to quantitatively evaluate the performance of the model in recognizing irrelevant gestures.

3.3. Comparison experiment

In this paper, the proposed EMG-FRNet for EMG-based irrelevant gesture recognition is compared with the existing methods based on SVDD (19) and AE (20). These comparative algorithms are all state-of-the-art methods in the field of irrelevant gesture recognition and have demonstrated good performance in this area, thus we chose them as the comparison algorithms in this study.

The experimental settings are as described in Section 2.1, where each subject undergoes multiple experiments, with a different target gesture category set for each experiment. DB1 and DB5 each have 10 subjects, and each subject performs 8 different target gesture experiments. The self-collected dataset consisted of 6 participants, with each participant performing 7 different target gesture experiments. Firstly, the AUC values are recorded for each subject when setting different target gestures. Then, the AUC values corresponding to different target gestures are averaged to obtain the subject's AUC value. Finally, the AUC values of all subjects in the dataset are averaged to obtain the dataset's AUC value. Figure 8 shows the AUC line graph for different subjects in each dataset. Table 1 shows the AUC values for each dataset.

From Figure 8, it can be seen that for different subjects, compared to the SVDD and AE comparison algorithms, the AUC value of the EMG-FRNet model can always maintain a high and relatively stable level. Specifically, the AUC values on different datasets are described in Table 1.

In the aforementioned comparative experiments, SVDD performed the worst. The reason for this may be that the traditional one-class support vector machine method is used, based on spherical hyperplanes. This method leads to poor classification performance in the case of complex data distributions such as EMG samples. The AE achieves improvement based on the SVDD. This may be because AE uses an autoencoder method for feature learning and extraction, which can better explore the intrinsic features of data. The EMG-FRNet achieves improvement based on the AE. The reason for this may be that the method utilizes the advantages of Generative Adversarial Networks (GANs) in anomaly detection. It also combines strategies such as CC, CLEDF, and SE to further improve the model's performance in EMG

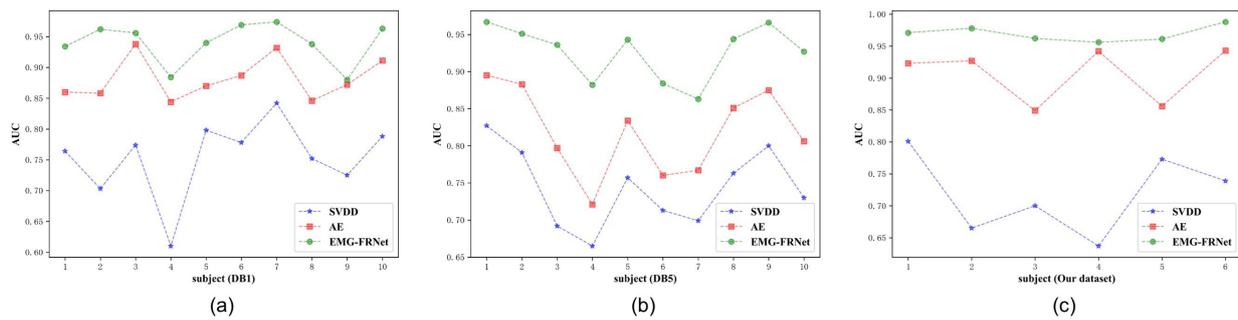


Figure 8. AUC line chart for different subjects in each dataset. (a)-(c) represent the AUC line charts of SVDD, AE, and EMG-FRNet on different subjects in DB1, DB5, and the self-collected dataset, respectively.

Table 1. AUC values for each dataset (Comparison experiment)

Model Name	DB1	DB5	Our dataset
SVDD (19)	0.744	0.753	0.719
AE (20)	0.882	0.819	0.907
EMG-FRNet	0.940	0.926	0.969

irrelevant gesture recognition tasks.

In conclusion, the proposed EMG-FRNet achieves state-of-the-art (SOTA) performance in the task of EMG irrelevant gesture recognition.

3.4. Ablation experiments

To validate the effectiveness of the proposed method, the following ablation experiments were conducted: based on the original GANomaly network, CC, CLEDF, and SE were successively added. In the legend below, GANomaly is represented as Baseline, GANomaly+CC is represented as Model 1, GANomaly+CC+CLEDF is represented as Model 2, and GANomaly+CC+CLEDF+SE (EMG-FRNet) is represented as Model 3.

The experimental setup is described in Section 2.1. Figure 9 demonstrates the AUC fold plots for different subjects for each dataset. Table 2 demonstrates the AUC values for each dataset.

From Figure 9, it can be seen that the performance of Baseline, Model 1, Model 2, and Model 3 increases in turn. Specifically, the AUC values on different datasets are described in Table 2.

In the ablation experiments, Model 1 achieves improvement based on the Baseline. The reason for this improvement is that the CC reduces the impact of redundant features in the original GANomaly network on model performance. Model 2 further improves upon Model 1. The reason for this improvement is that the CLEDF can compensate for information loss during downsampling. Model 3 has been improved from model 2. The reason for this improvement is that the SE can avoid the problem of redundant information brought by feature fusion, making the network focus on more important features and improving the reconstruction

performance of the model.

In summary, the proposed model EMG-FRNet achieves the best performance in the task of recognizing irrelevant gestures. CC, CLEDF, and SE all improve the model's performance to varying degrees.

4. Discussion

Currently, the recognition performance of most studies on EMG irrelevant gesture recognition is unstable. This paper establishes a connection between the EMG irrelevant gesture recognition and anomaly detection fields, and for the first time applies GANomaly to EMG irrelevant gesture recognition. Based on this, a feature reconstruction network, EMG-FRNet, is proposed for EMG irrelevant gesture recognition. The network exhibits a small feature reconstruction error for target class samples and a large feature reconstruction error for irrelevant class samples, which improves the ability to distinguish between target and irrelevant samples. We verify the feasibility of the proposed method through experiments, and the results show that our method can maintain a high level of performance in all EMG datasets.

In this paper, we have achieved high reliability in distinguishing target gestures from multiple irrelevant gestures. However, in practical myoelectric interaction applications, there are often multiple types of target gestures, which not only require us to distinguish between target and irrelevant gestures but also to distinguish between different categories of target gestures. In addition, there are other types of interference in practical myoelectric interaction applications, such as electrode displacement, muscle fatigue, user differences, etc., which can lead to poor interaction effects. Therefore, for future research, we will further explore the following aspects: (i) Based on the EMG-FRNet method proposed in this paper, we aim to achieve the distinction between multiple target gestures and multiple irrelevant gestures, as well as the distinction between different categories of target gestures. (ii) Additionally, we will seek corresponding solutions for other types of interference, with the goal of improving the interaction effects in

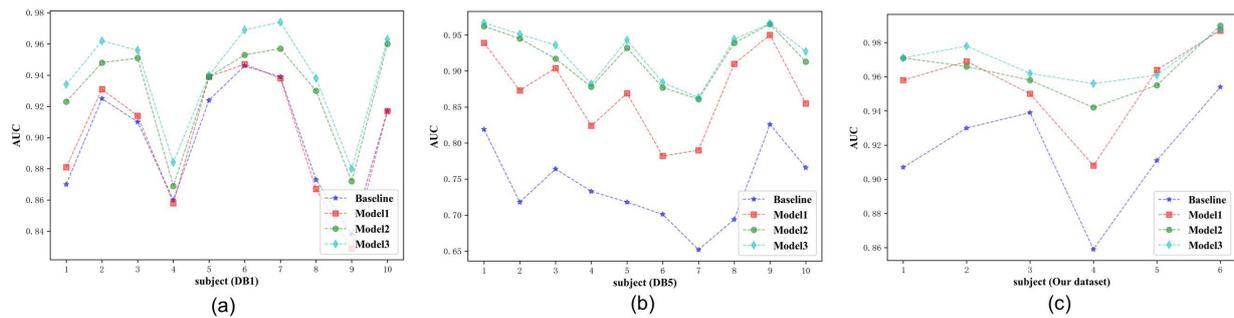


Figure 9. AUC line chart for different subjects in each dataset. (a)-(c) represent the AUC line charts of GANomaly, GANomaly+CC, GANomaly+CC+ CLEDF, and EMG-FRNet on different subjects in DB1, DB5, and the self-collected dataset, respectively.

Table 2. AUC values for each dataset (Ablation experiments)

Items	CC	CLEDF	SE	DB1	DB5	Our dataset
Baseline				0.900	0.739	0.915
Model 1	√			0.902	0.870	0.956
Model 2	√	√		0.930	0.919	0.963
Model 3	√	√	√	0.940	0.926	0.969

practical myoelectric applications.

Funding: This study was partially supported by Beijing Municipal Science & Technology Commission No.Z221100007422068

Conflict of Interest: The authors have no conflicts of interest to disclose.

Author statement: Zhang W and Wang Y conceived the ideas and designed the methodology. Zhang W and Wang Y implemented the technical pipeline, conducted the experiments, and analyzed the results. Zhang J and Pang G provided the dataset for the experiments; All authors discussed, wrote the manuscript, and gave final approval for publication.

References

- Lv Z, Xiao F, Wu Z, Liu Z, Wang Y. Hand gestures recognition from surface electromyogram signal based on self-organizing mapping and radial basis function network. *Biomed Signal Process Control*. 2021; 68:102629.
- Zhang Z, He C, Yang K. A Novel Surface Electromyographic Signal-Based Hand Gesture Prediction Using a Recurrent Neural Network. *Sensors (Basel)*. 2020; 20:3994.
- Xu Z, Tian Y, Li Y. sEMG Pattern Recognition of Muscle Force of Upper Arm for Intelligent Bionic Limb Control. *J Bionic Eng*. 2015; 12, 316–323.
- Cases C, Baldovino R, Manguerra M, Dupo V. An EMG-based Gesture Recognition for Active-assistive Rehabilitation. 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). DOI: 10.1109/HNICEM51456.2020.9400132
- Yang X, Chen X, Cao X, Wei S, Zhang X. Chinese Sign Language Recognition Based on an Optimized Tree-Structure Framework. *IEEE J Biomed Health Inform*. 2017; 21:994-1004.
- Xu H, Xiong A. Advances and Disturbances in sEMG-Based Intentions and Movements Recognition: A Review. *IEEE Sens J*. 2021; 21:3019-13028. doi: 10.1109/JSEN.2021.3068521.
- Li X, Chen S, Zhang H, Samuel OW, Wang H, Fang P, Zhang X, Li G. Towards reducing the impacts of unwanted movements on identification of motion intentions. *J Electromyogr Kinesiol*. 2016; 28:90-98.
- Chang J, Phinyomark A, Bateman S, Scheme E. Wearable EMG-Based Gesture Recognition Systems During Activities of Daily Living: An Exploratory Study. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020; 2020:3448-3451.
- Robertson JW. Toward an enhanced understanding of rejection in pattern recognition-based myoelectric control. University of New Brunswick, 2019.
- Chen H, Zhang Y, Zhou D, Liu H. Improving Gesture Recognition by Bidirectional Temporal Convolutional Networks. *International Conference on Robotics and Rehabilitation Intelligence*. Springer, Singapore, 2020; 413-424.
- Rahimian E, Zabihi S, Asif A, Farina D, Atashzar SF, Mohammadi A. TEMGNet: Deep Transformer-based Decoding of Upperlimb sEMG for Hand Gestures Recognition. *arXiv:2109.12379*, 2021. <https://doi.org/10.48550/arXiv.2109.12379>
- Zhang W, Zhao T, Zhang J, Wang Y. LST-EMG-Net: Long short-term transformer feature fusion network for sEMG gesture recognition. *Front Neurobot*. 2023; 17:1127338.
- Scheme E J, Hudgins B S, Englehart K B. Confidence-based rejection for improved pattern recognition myoelectric control. *IEEE Trans Biomed Eng*. 2013; 60:1563-1570
- Robertson J W, Englehart K B, Scheme E J. Effects of confidence-based rejection on usability and error in pattern recognition-based myoelectric control. *IEEE J Biomed Health Inform*. 2019; 23:2002-2008.
- Tomczyński J, Kaczmarek P, Mańkowski T. Hand gesture-based interface with multichannel sEMG band enabling unknown gesture discrimination. 2015 10th International Workshop on Robot Motion and Control (RoMoCo). DOI: 10.1109/RoMoCo.2015.7219713
- Bao T, Zaidi S AR, Xie SQ, Yang PF, Zhang ZQ. CNN Confidence Estimation for Rejection-Based Hand Gesture Classification in Myoelectric Control. *IEEE Trans Hum Mach Syst*. 2021; 52:99-109.

17. Zhou S, Fei F, Yin K. Toward improving the reliability of discrete movement recognition of sEMG signals. *Applied Sciences*. 2022; 12: 3374.
18. Ding Q, Li Z, Zhao X, Xiao Y, Han J. Real-time myoelectric prosthetic-hand control to reject outlier motion interference using one-class classifier. 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC). DOI: 10.1109/YAC.2017.7967385
19. Ding Q, Zhao X, Han J, Bu C, Wu C. Adaptive Hybrid Classifier for Myoelectric Pattern Recognition Against the Interferences of Outlier Motion, Muscle Fatigue, and Electrode Doffing. *IEEE Trans Neural Syst Rehabil Eng*. 2019; 27:1071-1080.
20. Wu L, Chen X, Chen X, Zhang X. Rejecting novel motions in high-density myoelectric pattern recognition using hybrid neural networks. *Front Neurobot*. 2022; 16:862193.
21. Wu L, Zhang X, Zhang X, Chen X, Chen X. Metric learning for novel motion rejection in high-density myoelectric pattern recognition. *Knowl Based Syst*. 2021; 227:107165.
22. Pang G, Shen C, Cao L, Hengel A. Deep learning for anomaly detection: A review. *ACM Comput Surv (CSUR)*, 2021; 54:1-38.
23. Bati E, Çalışkan A, Koz A, Alatan AA. Hyperspectral anomaly detection method based on auto-encoder. *Proc. SPIE 9643, Image and Signal Processing for Remote Sensing XXI*, 96430N. 2015; <https://doi.org/10.1117/12.2195180>.
24. Cowton J, Kyriazakis I, Plötz T, Bacardit J. A Combined Deep Learning GRU-Autoencoder for the Early Detection of Respiratory Disease in Pigs Using Multiple Environmental Sensors. *Sensors (Basel)*. 2018; 18:2521.
25. An J, Cho S. Variational autoencoder based anomaly detection using reconstruction probability. *Special lecture on IE*. 2015; 2:1-18. <http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf>
26. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings*. Cham: Springer International Publishing, 2017: 146-157.
27. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR. Efficient gan-based anomaly detection. *arXiv:1802.06222*, 2018. <https://doi.org/10.48550/arXiv.1802.06222>
28. Akcay S, Atapour-Abarghouei A, Breckon TP. GANomaly: Semi-supervised anomaly detection via adversarial training. *Computer Vision – ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part III* 14. Springer International Publishing, 2019; 622-637.
29. Li Q, Luo Z, Zheng J. Deep Learning-based User Authentication with Surface EMG Images of Hand Gestures. *Annu Int Conf IEEE Eng Med Biol Soc*. 2021; 2021:2038-2041.
30. Li Q, Luo Z, Zheng J. A new deep anomaly detection-based method for user authentication using multichannel surface EMG signals of hand gestures. *IEEE Trans Instrum Meas*. 2022; DOI: 10.1109/TIM.2022.3164162
31. Atzori M, Gijsberts A, Heynen S, Hager AGM, Deriaz O, Van Der Smagt P, Castellini C, Caputo B, Müller H. Building the Ninapro database: A resource for the biorobotics community. 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechanics (BioRob). *IEEE*, 2012; DOI: 10.1109/BioRob.2012.6290287
32. Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H, Atzori M. Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLoS One*. 2017; 12:e0186132.
33. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018; DOI: 10.1109/CVPR.2018.00745
34. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information processing & management*. 2009; 45:427-437.

Received May 21, 2023; Revised June 21, 2023; Accepted June 29, 2023.

*Address correspondence to:

Wenli Zhang, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China.
E-mail: zhangwenli@bjut.edu.cn

Released online in J-STAGE as advance publication June 30, 2023.