# Rejecting Unknown Gestures Based on Surface-Electromyography Using Variational Autoencoder

Qingfeng Dai, Yongkang Wong, *Member, IEEE*, Mohan Kankanhalli, *Fellow, IEEE*, Xiangdong Li, and Weidong Geng

*Abstract*—The conventional surface electromyography (sEMG)-based gesture recognition systems exhibit impressive performance in controlled laboratory settings. As most systems are trained in a closed-set setting, the systems's performance may see significant deterioration when novel gestures are presented as imposter. In addition, the state-of-the-art generative and discriminative methods have achieved considerable performance on high-density sEMG signals. This can be seen as an unrealistic setting as the real-world muscle computer interface are mainly comprised of sparse multichannel sEMG signals. In this work, we propose a novel variational autoencoder based approach for open-set gesture recognition based on sparse multichannel sEMG signals. Using the predefined corresponding latent conditional distribution of known gestures, the conditional Gaussian distribution of each known gesture is learned. Those samples with low probability density are identified as unknown gestures. The sEMG signals of known gestures are classified using the Kullback-Leibler divergences between the predefined prior distributions and input samples. The proposed approach is evaluated using three benchmark sparse multichannel sEMG databases. The experimental results demonstrate that our approach outperforms the existing open-set sEMG-based gesture recognition approaches and achieves a better trade-off between classifying known gestures and rejecting unknown gestures.

*Index Terms*—Open-set recognition, gesture recognition, surface electromyography, variational autoencoder.

## I. INTRODUCTION

THE Muscle-Computer Interface (MCI) interprets a user's body language by decoding the surface electromyography (sEMG) signal [1]. In the prior arts, sEMG-based gesture recognition has attracted the most investigation [2], [3], [4]. Normally, the closed-set setting, where the type of gestures presented in the training and test set are identical, is adopted. In contrast, a realistic scenario assumes that a system developer could not anticipate all the possible gestures available during the test or deployment phase. This is to say, a user may perform an unknown gesture, whereas the MCI system will identify it as a known one. Thus, it is essential to develop a system based on open-set recognition (OSR) [5] setting.

The objective of open-set sEMG-based gesture recognition is to identify a segment of sEMG signals and classify it as a known gesture or an unknown gesture. Most of the existing OSR approaches are designed for image-relevant tasks such as image recognition and image segmentation [5], [6], [7], [8]. Only a handful of OSR approaches are proposed for sEMG-based gesture recognition [9], [10], [11]. Briefly, Wu et al. [10] trained multiple autoencoders (AEs) to reject unknown gestures with significantly different representations of sEMG feature images. Besides, a discriminative method based on a convolutional prototype network (CPN) constructs multiple prototypes of known motion classes and identifies samples of known/unknown motions using the prototype matching [11]. Despite achieved promising performance on high-density sEMG signals, there are no evaluations on sparse multichannel sEMG datasets, such as the NinaPro databases [2], [12], [13], [14]. Moreover, sparse multichannel sEMG is more manageable than HD-sEMG in terms of data processing and computational requirements. In this paper, we introduce a novel open-set gesture recognition approach based on a variational autoencoder (VAE) with conditional Gaussian distribution learning [15] that enables classifying known gestures and rejecting unknown gestures given the input sparse sEMG signals for more efficiently training.

VAE exhibits the capacity to effectively reconstruct input data while simultaneously enforcing the approximation of a prior distribution by the posterior distribution in the latent feature space. Based on the reconstruction loss of a test sample, a predefined threshold is set to identify whether the given gesture is known or novel. Due to electrode placement, user differences, and muscle fatigue, sEMG signals can exhibit significant variability. Previous methods might not effectively

account for this variability, leading to inadequate rejection of unknown gestures. Gaussian distributions can model a wide range of sEMG representations. By conditional Gaussian distribution learning(CGDL) [15], the model can adapt its understanding of the variability and central tendencies. Therefore, we adopt the techniques utilized in conditional Gaussian learning [15], where a distinct latent conditional distribution in the latent space for every known gesture is predefined. Then Kullback-Leibler divergences between the predefined prior distribution of known gestures and the representations of input sEMG samples are regarded as the confidences that input sEMG signals belong to a specific known gesture. Furthermore, we conduct experiments on three sparse multichannel sEMG datasets including NinaPro DB1, NinaPro DB2 and NinaPro DB5 [2], [12] to validate the effectiveness of our approach.

Our contributions can be summarized as follows:

- We propose a novel variational autoencoder (VAE)-based approach for open-set gesture recognition based on sparse multichannel sEMG. We present how to classify known gestures using a VAE, which is commonly utilized for rejecting unknown gestures.
- We demonstrate the effectiveness of the VAE by utilizing different encoder and decoder architectures to train models for the open-set sEMG-based gesture recognition task. Not only our approach enables rejecting unknown gestures, but also achieves a competitive performance of classifying known gestures.
- We conduct comprehensive evaluations on three sparse multichannel sEMG datasets. The experimental results show that the proposed approach achieves better performance than the state-of-the-art open-set gesture recognition method. Specifically, our approach outperforms the prior arts by 0.04, 0.02 and 0.01 on NinaPro DB1, NinaPro DB2, and NinaPro DB5, using the metric of the area under the receiver operating characteristic (AUROC) which reflects the capacity to accurately reject unknown gestures. Moreover, the ablation studies examine various components of the proposed approach, including architectures of the encoder and decoder, known gestures utilized for training as well as the effects of the VAE.

The remainder of this paper is organized as follows. Section II summarizes the related work. We formulate the problem of open-set gesture recognition based on sEMG and introduce the proposed VAE-based approach in Section III. Section IV describes the evaluation datasets, evaluation metrics, network architectures and implementation details. The experimental results are demonstrated in Section V. Finally, we conclude this paper and discuss future work in Section VI-A.

## II. RELATED WORKS

In this section, we will discuss relevant literature pertaining to our field of work regarding surface electromyography (sEMG) based gesture recognition under the closed-set and open-set scenarios.

### A. Closed-Set Gesture Recognition Based on sEMG

This approach aims to interpret hand gestures from sEMG signals using computer algorithms under the closed-set setting. Depending on the way of feature extraction, the approaches can be categorized into conventional machine learning approach and deep learning approach [16]. The former extracts handcrafted features such as mean absolute values, waveform lengths, and discrete wavelet transform coefficients [17]. Upon extracting handcrafted features, conventional machine learning classifiers including SVMs, kNNs, Random Forests, and so on [14], [18], and [2] are typically utilized for performing downstream classification. Differently, deep learning-based approaches produce discriminative features using deep neural networks [3], [4], [19], [20], which are more effective and faster. For example, Geng et al. [3] converted the sEMG signals into grayscale images and leveraged a typical convolutional neural network to classify the converted sEMG images. Besides, a hybrid CNN-RNN [19] and XceptionTime [4] are proposed to model the spatial and temporal information of sEMG signals and achieve considerable gesture recognition performance. On the other hand, recent works also have looked at multi-modal gesture recognition for a higher recognition accuracy [17], [21], [22], [23].

### B. Open-Set Gesture Recognition Based on sEMG

In the test phase of closed-set gesture recognition, all the available gestures are already known during training. Nevertheless, it is challenging to ensure these conditions are available in real-world scenarios. Therefore, open-set sEMG-based gesture recognition approaches are investigated to tackle this problem.

Open-set recognition (OSR) problem is firstly proposed for face recognition to detect unknown faces [24]. Specifically, the OSR problem aims to learn a classifier to not only accurately distinguish known classes but also reject unknown ones. Existing approaches can be divided into two categories including discriminative approaches and generative ones. The former utilizes discriminative models such as support vector machine [5] and convolutional neural network [25] to output classification probabilities of rejecting unknown classes through a comparison with a predefined threshold. In addition, metric learning [26], [27]) and prototype learning [28], [29] are commonly used to distinguish between known and unknown samples via their distances on the feature space. With regard to the generative approaches, generative models are trained to learn the probability distribution of each known class. Specifically, if the likelihoods of the test sample estimated from all known class models are low, the sample will be classified as belonging to an unknown class. Typically, generative adversarial networks [30], [31], [32]) and autoencoder [15], [33]) are the most popular generative approaches applied for the OSR problem.

Although there are several typical strategies for traditional OSR problems, open-set gesture recognition based on sEMG is much more challenging. The main challenges lie in the inevitable noise and the low amplitude of sEMG signals that increase the difficulty of finding a proper function
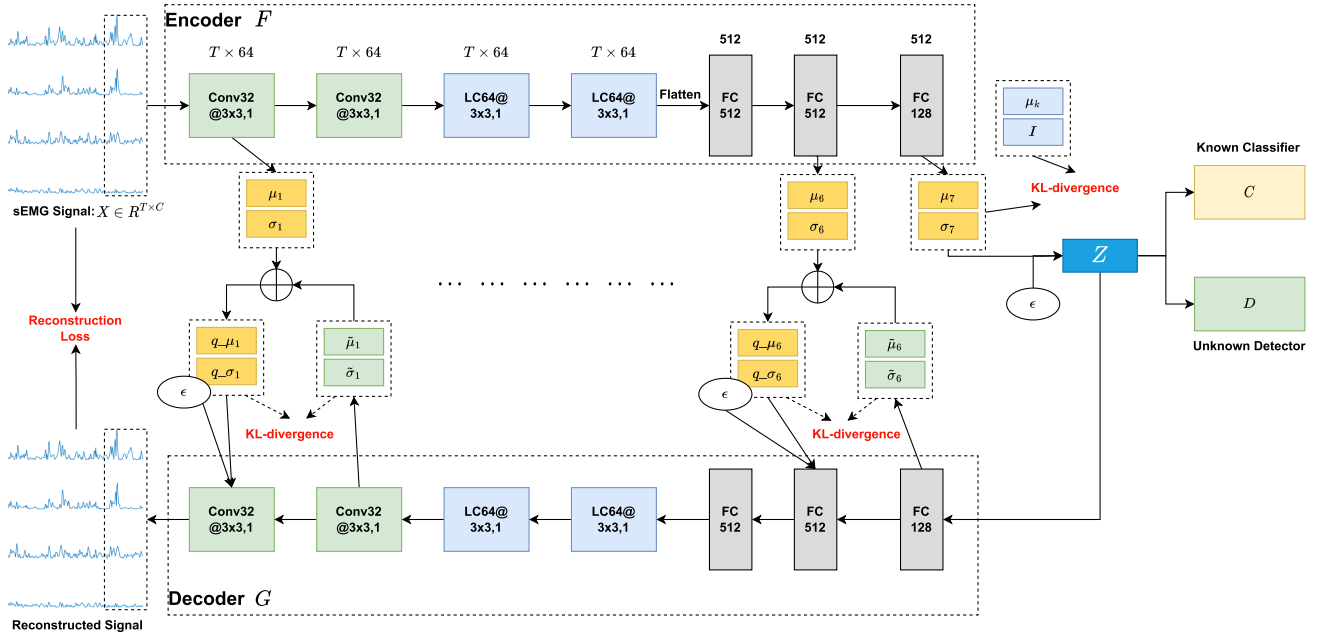
Fig. 1. Diagram of the proposed method based on VAE for open-set sEMG-based gesture recognition. Conv, LC, and FC respectively denote the convolution layer, locally-connected layer and fully-connected layer. The number following the layer name denotes the number of filters, and the numbers after the symbol @ denote the convolution kernel size and stride size.

to identify unknown classes from sEMG signals. Similarly, discriminative approaches utilize well-performed models on sEMG-based gesture recognition to produce probabilities that input sEMG signals belong to unknown gestures. For example, Robertson et al. [34] used a support vector machine classifier to make a tradeoff between error mitigation and unknown rejections in real-time close-loop myoelectric control at a specific threshold. Linear discriminant analysis (LDA) [35] and artificial neural networks (ANN) [9] are also utilized to output entropy function to discriminate between known command gestures and unknown gestures based on sEMG signals. Besides the aforementioned discriminative approaches, Ding et al. [36], [37] proposed hybrid models that combine a multi-class classifier of known classes and an one-class classifier rejecting unknown classes. Wu et al. [10] utilized a reconstruction-based autoencoder (AE) to reject samples with high reconstruction errors and achieved significant performance using high-density surface electromyography. To overcome the respective drawbacks of discriminative and generative approaches, Wu et al. [11] further proposed a method that combines their advantage in an end-to-end manner based on HD-sEMG is proposed using convolutional prototype network (CPN) and prototype matching. Although open-set gesture recognition based on HD-sEMG has been investigated, few sparse multichannel sEMG datasets are utilized to evaluate these approaches.

## III. METHOD

This section delineates the proposed method based on variational autoencoder (VAE) [38] for open-set sEMG-based gesture recognition (Fig. 1). As a probabilistic graphical model, a VAE is capable of accurately reconstructing input data while also enforcing the posterior distribution $q_\phi(z|x)$ in the latent space to closely approximate a prior distribution $p_\theta(z)$. Therefore, an appropriately trained VAE is expected to

accurately discern known data and identify biased samples as unknown via the reconstruction loss. Nevertheless, the VAE cannot classify known categories because all samples follow a single distribution. To enable classifying known categories, we predefine a latent conditional distribution $q_\phi(z|x, k)$ for each known class (denoted as $k$th) in the latent space via conditional Gaussian distribution learning(CGDL) [15]. Subsequently, samples of the $k$th known gesture are required to approximate a multivariate Gaussian distribution $p_\theta^{(k)}(z) = \mathcal{N}(z; \mu_k, I)$, where $\mu_k$ is the mean value of the $k$th distribution. Because samples of each gesture conform to its corresponding prior distribution, those samples with low probability density are identified as unknown gestures. With regard to samples of known gestures, Kullback-Leibler divergences between the predefined prior distributions and input samples are calculated as the confidences. The gesture with the highest level of confidence is regarded as the predicted one of the model. The complex and diverse patterns of sEMG signals demand a more discriminative architecture of VAE. Besides, certain unknown gestures can closely resemble known gestures, such as the thumb up and fist, which implies that their corresponding Gaussian distributions may be too similar. To tackle this issue, we utilize a parameter-tuning strategy based on Bayesian optimization to find the optimal probability threshold for a better trade-off between rejecting unknown gestures and classifying known ones.

### A. Problem Formulation

This paper focuses on open-set sEMG-based gesture recognition. Given a training set of sEMG signals with data-label pairs $\mathcal{D}_{\text{Train}} = \{(x_i \in X_{\text{Train}}, y_i \in Y_{\text{Train}} = \{1, \cdots, C_{\text{Train}}\}) | i = 1, 2, \cdots, m\}$ and a test set $\mathcal{D}_{\text{Test}} = \{(x_j \in X_{\text{Test}}, y_j \in Y_{\text{Test}} = \{1, \cdots, C_{\text{Train}}, \cdots, C_{\text{Test}}\}) | j = 1, 2, \cdots, n\}$, where $x_i, x_j$

denotes the sEMG signals and $y_i$, $y_j$ for gesture labels. As an open-set task, the $C_{\mathrm{Train}}$ is the known classes and $\{C_{\mathrm{Train}} + 1, \cdots, C_{\mathrm{Test}}\}$ is the unknown classes. Our target is utilizing $\mathcal{D}_{\mathrm{Train}}$ to train a model $f$ to accurately classify the subset $\mathcal{D}_{\mathrm{Test}}^{\mathrm{known}} = \{x_i \in X_{\mathrm{Test}}, y_i \in \{1, \cdots, C_{\mathrm{Train}}\}\}$ and detect unknown gestures in the subset $\mathcal{D}_{\mathrm{Test}}^{\mathrm{unknown}} = \{x_i \in X_{\mathrm{Test}}, y_i \in \{C_{\mathrm{Train}} + 1, \cdots, C_{\mathrm{Test}}\}\}$. Specifically, the minimization of open space risk $R_{\mathcal{O}}$ and empirical risk $R_{\epsilon}$ is solved to find the optimal function $f$ as follows:

$$\arg\min_{f}\{R_{\mathcal{O}}(f) + \lambda_r R_{\epsilon}(f(\mathcal{D}_{\mathrm{Train}}))\}$$

where $f(x) > 0$ means that the sEMG sample $x$ belongs to a known gesture. The empirical risk $R_{\epsilon}$ denotes the average classification loss of all samples in the $\mathcal{D}_{\mathrm{Train}}$ and $\lambda_r$ is the weight of classification loss.

## B. Constructing Variational Auto-Encoder (VAE)

A VAE usually consists of an encoder and a decoder, whose loss function is represented as $\mathcal{L}(\theta; \phi; x)$ where $\phi$ and $\theta$ are the parameters of the encoder and the decoder, respectively. A latent representation $z$ is obtained after the input sEMG signal $x$ is fed into the encoder. Conversely, a reconstructed sEMG sample $\hat{x}$ is obtained by feeding the latent representation $z$ into the decoder. The loss function of VAE is commonly defined as follows:

$$\mathcal{L}(\theta; \phi; x) = -D_{KL}\big(q_\phi(z|x) \,\|\, p_\theta(z)\big) + \mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x|z)\big] \tag{1}$$

where $q_\phi(z|x)$ denotes the approximate posterior distribution, $p_\theta(z)$ is the prior distribution of the latent representation $z$, and $p_\theta(x|z)$ represents the conditional possibility of input sample $x$ given a latent representation $z$. The first item of Eq. (1) is the KL divergence between the approximate posterior distribution and the prior distribution. This loss item enforces the similarity between $q_\phi(z|x)$ and $p_\theta(z)$. The second item of $\mathcal{L}(\theta; \phi; x)$ is often represented using the reconstruction loss.

The prior distribution $p_\theta(z)$ is assumed to be an independent multivariate Gaussian distribution, denoted by $p_\theta = \mathcal{N}(z; 0, I)$. Then the variational posterior distribution obtained from the trained VAE is formulated as follows:

$$q_\phi(z|x) = \mathcal{N}(z; \mu, \sigma^2 I)$$

where the mean value $\mu$ and variance $\sigma$ of Gaussian distribution are outputted by the trained VAE. The latent representation is $z = \mu + \sigma \odot \epsilon$ and $\epsilon$ follows the standard normal distribution. The conditional information of the $k$th gesture is introduced into VAE and the prior distribution $p_\theta(z)$ is denoted as $\mathcal{N}(z; \mu_k, I)$, where the vector $\mu_k$ is obtained by passing the one-hot encoding of each gesture through a fully-connected layer. Thus, the KL divergence is calculated as follows:

$$-D_{KL}\big(q_\phi(z|x) \,\|\, p_\theta(z)\big)$$
$$= \int \mathcal{N}(z; \mu, \sigma^2)\big(\log \mathcal{N}(z; \mu_k, I) - \log \mathcal{N}(z; \mu, \sigma^2)\big)dz$$
$$= \frac{1}{2}\sum_{j=1}^{J}\Big(1 + \log(\sigma_j^2) - (\mu_j - \mu_j^{(k)})^2 - \sigma_j^2\Big)$$

The architecture of the proposed model for open-set sEMG-based gesture recognition is derived from GengNet [3]. It consists of a number of modules, including an encoder $\mathcal{F}$, a decoder $\mathcal{G}$, a classifier $\mathcal{C}$ for known gestures and a detector $\mathcal{D}$ for unknown gestures. We delineate the details of these modules in the following:

*1) Encoder $\mathcal{F}$:* The architecture of $\mathcal{F}$ follows that of GengNet and a module for calculating the mean and variance is integrated at each layer of $\mathcal{F}$. Specifically, the output feature $x_l$ of the $l$th layer is transformed into a one-dimensional representation and then it is fed into a fully-connected layer to obtain the mean and variance values. To prevent information loss, we replace the original ReLU activation function with PReLU.

*2) Decoder $\mathcal{G}$:* The architecture of $\mathcal{G}$ is equivalent to that of the inverse encoder, except the convolution layers are replaced by the transpose convolution layers and *Tanh* is employed as the activation function of the last layer. This module aims to reconstruct sEMG signals using parameter sampling from the learned distribution. In this decoder, a probability ladder architecture [39] is utilized to exchange bottom-up and top-down information for restoring the information that is dismissed by the encoder $\mathcal{F}$. The information is also represented in terms of a Gaussian distribution and its mean, and a Gaussian distribution and its variance are as follows:

$$q\_\mu_l = \frac{\widetilde{\mu}_l \widetilde{\sigma}_l^{-2} + \mu_l \sigma_l^{-2}}{\widetilde{\sigma}_l^{-2} + \sigma_l^{-2}}$$
$$q\_\sigma_l^2 = \frac{1}{\widetilde{\sigma}_l^{-2} + \sigma_l^{-2}}$$

*3) Classifier $\mathcal{C}$ for Known Gestures:* Firstly, the classifier $\mathcal{C}$ calculates the probability densities of each gesture using the latent representation $z$ and Gaussian prior distributions. Then, the probability densities of each gesture are normalized through the Softmax function. After that, the gesture with the highest confidence level is regarded as the final classification result. And the probability densities of the $k$th gesture are calculated as follows:

$$p(z|\mu_k, I) = \frac{1}{\sqrt{(2\pi)^N}} exp\left(-\frac{1}{2}(z - \mu_k)^T(z - \mu_k)\right)$$

*4) Detector $\mathcal{D}$ for Unknown Gestures:* Given the well-trained encoder $\mathcal{F}$ and decoder $\mathcal{G}$, a specific multivariate Gausssian distribution $f_k(z) = \mathcal{N}(z; m_k, \sigma_k^2)$ of each gesture can be constructed. Then, we calculate the mean $m_k$ and variance $\sigma_k^2$ of the latent representations of all the training samples of the $k$th gesture in the feature space. Thus, the probability that a sample being drawn from the distribution $f_k(z)$ is defined as follows:

$$P_k(z) = 1 - \int_{m_0 - |z_0 - m_0|}^{m_0 + |z_0 - m_0|} \cdots \int_{m_n - |z_n - m_n|}^{m_n + |z_n - m_n|} f_k(\mathbf{t})d\mathbf{t}$$

where $n$ denotes the dimension of the feature space and $z$ is represented as $(z_1, z_2, \cdots, z_n)$.

## C. Evaluation Procedure

We also investigate the hidden information that the reconstruction loss carries, while the reconstruction loss with inputs from known gestures is typically smaller than that of unknown gestures [40]. In practical implementation, the threshold $\tau_r$ for reconstruction error is determined by identifying a fixed ratio of training sEMG data that is classified as known. Overall, the evaluation procedure of our approach is shown in Algorithm 1.

---

**Algorithm 1** The Procedure of Evaluation for Open-Set sEMG-Based Gesture Recognition

---

**Input:** A test sample $X$
**Input:** The trained encoder $\mathcal{F}$, decoder $\mathcal{D}$ and classifier for known gestures $\mathcal{C}$
**Input:** The threshold probability $\tau_l$ for Gaussian distribution
**Input:** The threshold $\tau_r$ for reconstruction error
**Input:** With regard to each gesture, denoted as $k$, the latent representation of the correctly classified training sample $x_{i,k}$ is $z_{i,k}$
  1: **for** $k = 1, \ldots, K$ **do**
  2:     $m_k = mean_i(z_{i,k})$
  3:     $\sigma_k^2 = var_i(z_{i,k})$
  4:     $f_k(z) = \mathcal{N}(z; m_k, \sigma_k^2)$
  5: **end for**
  6: The latent representation $Z = \mathcal{F}(X)$
  7: The predicted gesture $y_{pred} = argmax(\mathcal{C}(Z))$
  8: The reconstructed sEMG signal $\tilde{X} = \mathcal{G}(X)$
  9: The reconstructed error $R = \|X - \tilde{X}\|_1$
10: **if** $\forall k \in \{1, \ldots, K\}$, $P_k(Z) < \tau_l$ or$R > \tau_r$ **then**
11:     $X$ is predicted as unknown
12: **else**
13:     $X$ is predicted as a known gesture $y_{pred}$
14: **end if**

---

## D. Loss Function

The total loss function of our proposed approach is as follows:

$$\mathcal{L} = -(\mathcal{L}_r + \beta \mathcal{L}_{KL} + \lambda \mathcal{L}_c)$$

This loss function consists of three loss items including reconstruction loss ($\mathcal{L}_r$), KL divergence ($\mathcal{L}_{KL}$) and classification loss ($\mathcal{L}_c$). The reconstruction loss is obtained by means of the $L_1$ distance and the classification loss is computed using the cross-entropy loss function. The KL divergence is leveraged to approximate the posterior distribution to the prior distribution of each gesture by maximizing the similarity between the intermediate representations as follows:

$$\mathcal{L}_{KL} = -\frac{1}{L}[D_{KL}\left(q_\phi(z \mid x, k) \| p_\theta^{(k)}(z)\right) \\ + \sum_{l=1}^{L-1} D_{KL}(q_\theta(\tilde{x}_l \mid \tilde{x}_{l+1}, x) \| q_\theta(\tilde{x}_l \mid \tilde{x}_{l+1}))]$$

where

$$q_\theta(\tilde{x}_l \mid \tilde{x}_{l+1}, x) = \mathcal{N}\left(\tilde{x}_l; q_-\mu_l, q_-\sigma_l^2\right)$$

$$q_\theta(\tilde{x}_l \mid \tilde{x}_{l+1}) = \mathcal{N}\left(\tilde{x}_l; \tilde{\mu}_l, \tilde{\sigma}_l^2\right)$$

### TABLE I
### SPECIFICATIONS OF THE EVALUATED sEMG DATASETS

| | NinaPro DB1 [2] | NinaPro DB2 [2] | NinaPro DB5 [12] |
|---|---|---|---|
| Gestures | 52 | 50 | 53 |
| Subjects | 27 | 40 | 10 |
| Sessions | 1 | 1 | 1 |
| Trials | 10 | 6 | 6 |
| Channels | 10 | 12 | 16 |
| Sampling Rate | 100Hz | 2000Hz | 200Hz |
| Training Trials | 1,3,4,6,8,9 | 1,3,4,6 | 1,3,4,6 |
| Validation Trials | 2,7 | 2 | 2 |
| Testing Trials | 5,10 | 5 | 5 |
| Unknown Gestures | 2,5 | 2,5 | 2,5 |

## IV. EXPERIMENT SETTINGS

In this section, we will briefly introduce the benchmark datasets used in our experiments, the training settings, evaluation metrics, and implementation details.

### A. Datasets

Three sparse multichannel sEMG datasets, i.e., NinaPro DB1, NinaPro DB2 [2], and NinaPro DB5 [12] are utilized to train and evaluate the proposed VAE-based approach on open-set gesture recognition based on sEMG. The details of these three datasets are described in Table I. The training trials, validation trials and testing trials columns denote the number of trials used for training, validation and testing respectively. The unknown gestures column represents the gestures selected as unknown ones in the experiments.

### B. Data Preprocessing

Due to the high noise of sEMG signals, it is necessary to preprocess the raw signals in the experiments. Firstly, a Butterworth low-pass filter, whose cutoff frequency coefficient is set as 0.2, is applied to denoise the raw sEMG signals. Then, NinaPro DB2 and NinaPro DB5 are downsampled to 200Hz. To obtain the sEMG inputs with the same window length, sliding windows of lengths 20, 40, and 40 are used to segment the processed sEMG signals in a stride of 1 for NinaPro DB1, NinaPro DB2, and NinaPro DB5 respectively.

### C. Evaluation Metrics

To evaluate the performance of the proposed approach on open-set sEMG-based gesture recognition, the gestures of each dataset are divided into two parts including known gestures and unknown gestures. Additionally, the datasets are split into a training set, a validation set and a testing set in terms of trials. The specific dividing strategy refers to Table I. To perform a comprehensive comparison, four kinds of evaluation metrics are employed:

*Precision:* Intra-session gesture recognition precision is adopted for evaluation. Testing is conducted on sEMG data that contains $K + N$ gestures, with the output being $K + 1$ classification results. Therefore, *Precision* is computed as follows:

$$Precision = \sum_{k=1}^{K} \frac{TP_k}{(1+K)(TP_k + FP_k)}$$

TABLE II

AUROC AVERAGED OVER ALL SUBJECTS OF FOUR APPROACHES ON THREE SPARSE MULTICHANNEL sEMG DATASETS INCLUDING NINAPRO DB1, NINAPRO DB2 AND NINAPRO DB5. THE RESULTS IN BOLD ENTRIES INDICATE THE BEST PERFORMANCES

| | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 |
|---|---|---|---|
| OpenMax [25] | 0.64±0.07 | 0.69±0.08 | 0.67±0.12 |
| Multiple-AEs [10] | 0.66±0.06 | 0.68±0.07 | 0.67±0.10 |
| CPN [11] | 0.70±0.10 | 0.71±0.11 | 0.68±0.11 |
| Ours | **0.74±0.07** | **0.73±0.10** | **0.69±0.11** |

*Recall:* The weighted average recall of all gestures

$$Recall = \sum_{k=1}^{K} \frac{TP_k}{(1+K)(TP_k + FN_k)}$$

*F1-Score:* The harmonic mean between the *Precision* and *Recall*, which is a comprehensive evaluation metric for evaluating a model.

$$F1\text{-}Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

*AUROC:* The ROC curve is a graphical representation of the classifier's true positive rate versus its false positive rate, where each point on the curve represents a different threshold value for classifying the known and unknown gestures. The AUROC is the area under the ROC curve, which ranges from 0 to 1. The higher the AUROC, the better the performance of the evaluated model.

### D. Network Architecture

The architectures of $\mathcal{F}$ and $\mathcal{G}$ are derived from the GengNet [3], which achieve strong performance on NinaPro databases. Specifically, GengNet is composed of two convolutional layers, two locally connected layers and three fully connected layers. Each convolutional layer consists of 64 $3 \times 3$ filters with a stride of 1 and a zero padding of 1. Each locally connected layer consists of 64 $1 \times 1$ filters. The three fully connected layers consist of 512, 512 and 128 hidden units, respectively. Batch normalization and the ReLU nonlinearity function are applied to each layer, and dropout is adopted to prevent overfitting.

### E. Implementation Details

Our VAE-based approach is implemented with PyTorch. For all the three datasets, we utilize an optimizer of RAdam whose initial learning rate, weight decay and batch size are set as 0.001, 0.01 and 512 respectively for improved convergence and generalization. Besides, the learning rate is divided by 10 after each epoch and the total number of training epochs is set to 28 following the settings in GengNet [3]. The weights of loss items including classification loss and KL divergence are set to 100.0 and 1.0, respectively. With regard to the window lengths of input signals, 20, 40 and 40 are respectively adopted for NinaPro DB1, NinaPro DB2 and NinaPro DB5, to account for time delay constraints. We can see that the same hyperparameters are used for all three NinaPro variants which indicates that our model is not fine-tuned for each dataset.

## V. EXPERIMENTAL RESULTS

### A. Evaluation on Benchmark Dataset

To demonstrate the effectiveness of our proposed VAE-based approach, we conduct a comprehensive comparison with existing approaches on open-set gesture recognition based on sEMG in the evaluation metrics of AUROC, *Precision*, *Recall* and *F1-Score*. For a fair comparison, several existing open-set sEMG-based gesture recognition approaches are implemented. The details of these approaches are as follows:

- **OpenMax** [25] It is the most commonly used approach for open-set recognition where an OpenMax layer is proposed to replace the Softmax layer that is often used in the deep classification task. In this paper, we regard this approach as a baseline.
- **Multiple-autoencoders** [10] It trains multiple autoencoders (AEs) to reject representations from any unseen gesture that appeared significantly different from known gestures.
- **Convolutional Prototype Network (CPN)** [11] This approach constructs a CNN feature extractor and multiple prototypes of known gestures. Then samples of known or unknown gestures are identified using prototype matching.

Firstly, we compare our approach with the aforementioned three approaches using the evaluation metric of AUROC to validate its performance on unknown gesture rejection. Table II shows the average AUROC using testing sEMG data from all subjects of the four approaches. Our approach outperforms CPN, Multiple-AEs, OpenMax with an improvement on NinaPro DB1 and NinaPro DB2. Specifically, our approach improves the AUROC metric by +0.04, +0.08, +0.10 on NinaPro DB1, +0.02, +0.05, +0.04 on NinaPro DB2, +0.01, +0.02, +0.02 on NinaPro DB5 compared with CPN, Multiple-AEs and OpenMax, respectively. Note that AUROC metric reflects the ability to detect unknown gestures, our approach performs best to detect unknown gestures.

Secondly, our approach is compared with the other three methods under the scenario to classify $K + 1$ gestures including $K$ known gestures and one unknown category. Specifically, the metrics *Precision*, *Recall* and *F1-Score* in Section IV-C are adopted to evaluate the effectiveness of our approach. The performances of these four methods are shown in Table III. We can see that the three used evaluation metrics are consistent in validating the performance of classifying known gestures and rejecting unknown ones. Specifically, using the *Precision* metric, our approach outperforms CPN, Multiple-AEs, and OpenMax by +0.01, +0.03, and +0.06, respectively, on NinaPro DB1. Thus, our approach improves the performances of rejecting unknown gestures alongside classifying known gestures on NinaPro DB1. However, CPN performs best on NinaPro DB2 and NinaPro DB5 compared with the other three methods. Because discriminative methods can directly learn classification boundaries, the discriminative method CPN outperforms the other three methods in most scenarios. As a generative method, our approach achieves competitive performance against CPN and even performs better on NinaPro DB1. Note that our approach surpasses compared open-set sEMG-based gesture recognition approach

TABLE III

*Precision*, *Recall* AND *F1-Score* AVERAGED OVER ALL SUBJECTS OF FOUR APPROACHES ON THREE SPARSE MULTICHANNEL sEMG DATASETS INCLUDING NINAPRO DB1, NINAPRO DB2 AND NINAPRO DB5. THE RESULTS IN BOLD ENTRIES INDICATE THE BEST PERFORMANCES

| Method | *Precision* | | | *Recall* | | | *F1-Score* | | |
|---|---|---|---|---|---|---|---|---|---|
| | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 |
| OpenMax [25] | 0.70±0.07 | 0.71±0.10 | 0.74±0.05 | 0.71±0.05 | 0.69±0.07 | 0.76±0.07 | 0.70±0.06 | 0.69±0.08 | 0.72±0.07 |
| Multiple-AEs [10] | 0.73±0.07 | 0.73±0.08 | 0.75±0.05 | 0.73±0.07 | 0.71±0.08 | 0.76±0.06 | 0.71±0.05 | 0.69±0.10 | 0.74±0.06 |
| CPN [11] | 0.75±0.06 | **0.75±0.09** | **0.78±0.07** | 0.75±0.06 | **0.77±0.07** | **0.81±0.05** | 0.73±0.06 | **0.75±0.08** | **0.79±0.06** |
| Ours | **0.76±0.05** | 0.74±0.09 | 0.76±0.06 | **0.76±0.05** | 0.74±0.09 | 0.78±0.05 | **0.75±0.05** | 0.72±0.09 | 0.76±0.05 |

TABLE IV

COMPARISON OF OUR APPROACH USING THREE DIFFERENT NETWORK ARCHITECTURES FOR $\mathcal{F}$ AND $\mathcal{G}$ ON NINAPRO DB1, NINAPRO DB2 AND NINAPRO DB5. THE RESULTS IN **BOLD** ENTRIES INDICATE THE BEST PERFORMANCES

| Network Architectures | AUROC | | | *Precision* | | |
|---|---|---|---|---|---|---|
| | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 |
| GengNet [3] | **0.74±0.07** | **0.73±0.10** | 0.69±0.11 | **0.76±0.05** | **0.74±0.09** | **0.76±0.06** |
| XceptionTime [4] | 0.72±0.08 | 0.70±0.09 | 0.70±0.06 | 0.74±0.07 | 0.74±0.11 | 0.75±0.07 |
| TCN [41] | 0.69±0.09 | 0.71±0.10 | **0.71±0.08** | 0.71±0.09 | 0.73±0.07 | 0.76±0.09 |

TABLE V

COMPARISON OF VAE WITH CGDL AND THE TYPICAL VAE ON NINAPRO DB1, NINAPRO DB2 AND NINAPRO DB5. THE RESULTS IN **BOLD** ENTRIES INDICATE THE BEST PERFORMANCES

| | AUROC | | | *Precision* | | |
|---|---|---|---|---|---|---|
| | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 | NinaPro DB1 | NinaPro DB2 | NinaPro DB5 |
| VAE+CGDL | **0.74±0.07** | 0.73±0.10 | **0.69±0.11** | **0.76±0.05** | **0.74±0.09** | **0.76±0.06** |
| VAE | 0.73±0.04 | **0.74±0.11** | 0.68±0.05 | 0.74±0.07 | 0.72±0.06 | 0.72±0.09 |

for rejecting unknown gestures, our approach achieves a better trade-off between classifying known gestures and rejecting unknown gestures.

### B. Ablation Study

*1) Effects of CGDL:* To demonstrate the concrete factors that contribute to the effectiveness of our approach, we conduct a fair comparison between the proposed approach and the typical VAE without CGDL on NinaPro DB1, NinaPro DB2 and NinaPro DB5. In the experiments, GengNet [3] is selected as the architecture of $\mathcal{F}$ and $\mathcal{G}$. As shown in Table V, VAE with CGDL outperforms the typical VAE in the metric of *Precision*. However, their performances in rejecting unknown gestures are comparable when using the AUROC metric. These experimental results indicate that the introduction of CGDL could mitigate the degradation of classifying known gestures while sustaining the capability to reject unknown gestures.

*2) Backbone Architecture of VAE:* To validate whether the architectures of the encoder $\mathcal{F}$ and the decoder $\mathcal{G}$ affect the performance, we conduct a fair comparison using our approach with various architectures of $\mathcal{F}$ and $\mathcal{G}$. Specifically, XceptionTime [4]) and TCN [41] are applied as the backbone architecture of $\mathcal{F}$ and $\mathcal{G}$ for this validation. The former contains multiple XceptionTime modules which are composed of several separable convolutional layers and max pooling layers. The latter is composed of 1D fully-convolutional network and dilated convolutions enabling an exponentially large receptive field. Table IV reports the performance of our approach with three kinds of architectures of $\mathcal{F}$ and $\mathcal{G}$. As shown, GengNet outperforms XceptionTime and TCN. In addition, the proposed approach consistently outperforms the baseline with each architecture, which indicates our

approach is model-agnostic and can be applied to common ones. In this experiment, the evaluation metrics of AUROC and *Precision* are adopted for comparison.

*3) Variation on Known Gestures:* Here, we examine the effects of known gestures used for training on the performance of open-set sEMG-based gesture recognition. To cope with this, we divide the gestures into three groups and then conduct a 3-fold cross-validation. Specifically, two groups are selected as known gestures for training and validation, but the evaluation is performed on test data of all gestures. In this experiment, we select NinaPro DB1 and mix the sEMG data of all subjects to dismiss the influence brought by the individual differences of subjects. As shown in Table VI, the performances of rejecting unknown gestures using different kinds of known gestures are comparable. However, the known gestures used for training in our approach could affect the performance in the metric *F1-score*. Specifically, when the second group and third group are selected as known gestures, our approach achieved an *F1-score* of 0.83 on NinaPro DB1, which is much higher than the results obtained in the other two settings.

*4) Effects of VAE:* In this part, we explore the effects brought by the VAE on the features extracted using $\mathcal{F}$. To do this, we train the GengNet on NinaPro DB1 under the closed-set and open-set scenarios respectively. Then we utilize TSNE for visualizing the deep sEMG features, with a perplexity set to 30 and 5000 iterations. From each gesture, we randomly select 1000 samples for visualization. The specific visualization results of TSNE are displayed in Fig. 2. From Fig 2a, we can see that the features extracted by GengNet under the closed-set scenario are discretely distributed throughout the feature space and exhibit mutual overlap. In our approach (Fig 2b),

TABLE VI
OPEN-SET sEMG-BASED GESTURE RECOGNITION
PERFORMANCE WITH DIFFERENT KNOWN GESTURES
FOR TRAINING ON NINAPRO DB1

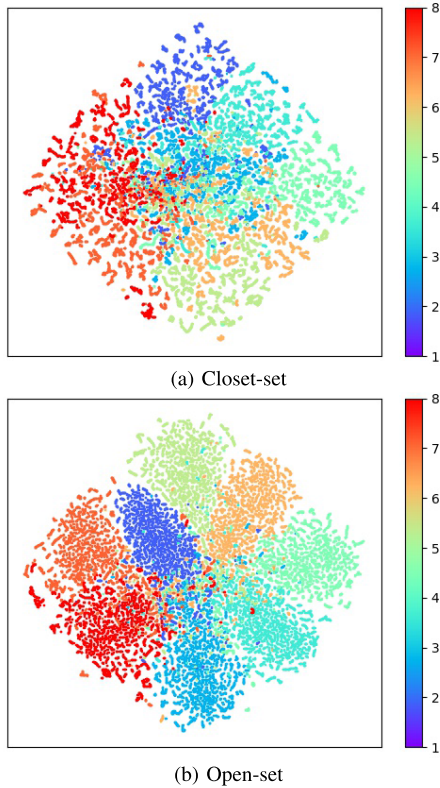| Groups for Training | Dataset | AUROC | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| 0,1 | NinaPro DB1 | 0.73 | 0.74 | 0.59 | 0.65 |
| 0,2 | NinaPro DB1 | 0.74 | 0.75 | 0.60 | 0.66 |
| 1,2 | NinaPro DB1 | 0.72 | 0.74 | 0.56 | 0.83 |



(a) Closet-set



(b) Open-set

Fig. 2. The visual comparison of features extracted by GengNet under the closed-set and open-set scenarios using our approach. Features of eight gestures are displayed in this figure and the 7th and 8th gestures belong to the unknown category.

the features of each gesture adhere to their respective prior Gaussian distributions by minimizing KL divergence. In this way, the extracted features of the same gesture can aggregate together, mitigating the open space risk.

## VI. CONCLUSION AND DISCUSSION

### A. Discussion

In this part, we discuss the reason why the introduction of VAE can improve the open-set gesture recognition performance and our future work. We first recall the framework of our approach, where the conditional Gaussian distribution learning [15]) is employed to define different prior distributions for all gestures, enabling classification for known gestures through comparison between distributions in terms of KL divergence. Although there are other competitive generative networks, such as GAN [42] and Glow [43], the reconstruction ability and latent space representation of VAE mitigates the negative impact caused by high noise in the sEMG signals. In addition, VAE has a probabilistic formulation where the encoder learns to approximate the posterior distribution of the latent variables given the sEMG data from a specific gesture.

### B. Conclusion

Most existing sEMG-based gesture recognition approaches are investigated under closed-set settings, which means that gestures during testing are already known when training. However, in a realistic scenario, the user may perform an unknown gesture which will be misclassified as a known gesture by a closed-set muscle-computer interface. Therefore, we propose a novel open-set sEMG-based gesture recognition approach based on the variational autoencoder (VAE) and we evaluated it on 3 public sparse multichannel benchmark databases. Different from the typical VAE, we utilize conditional Gaussian distribution learning [15] to enable classifying known gestures. Experimental results indicate that the proposed approach achieves better performance in rejecting unknown gestures compared with three existing approaches. Specifically, our approach achieves respective AUROC improvements of +0.04, +0.02, and +0.01 on NinaPro DB1, NinaPro DB2, and NinaPro DB5, compared with the state-of-the-art approach. On the other hand, our approach achieves a competitive performance in classifying known gestures against the discriminative open-set recognition methods. So the proposed approach could reach a better trade-off between classifying known gestures and rejecting unknown ones. We also conducted an ablation study to demonstrate the effects of known gestures for training and VAE on rejecting unknown gestures. The experimental results show that our approach is model-agnostic to the architecture of VAE and the introduction of VAE could mitigate the open space risk. Besides, the effects of known gestures used for training were investigated and we found that the performance of rejecting unknown gestures was minimally impacted, which means our approach is robust to training data.

Our future work will focus on enhancing the performance of rejecting unknown gestures meanwhile making the recognition accuracy of known gestures comparable to that under the closed-set settings. One way is to apply a more discriminative network as the encoder of VAE, whose performance on sEMG-based gesture recognition is superior. Another way is to replace the VAE with the latest generative networks, such as the diffusion probabilistic model [44]. We will also investigate improving the inter-subject or inter-session open-set gesture recognition performance based on our work considering its more practical prospects.

## REFERENCES

[1] A. Chowdhury, R. Ramadas, and S. Karmakar, "Muscle computer interface: A review," in *ICoRD*. India: Springer, 2013, pp. 411–421.

[2] M. Atzori et al., "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 1, pp. 1–13, Dec. 2014.

[3] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, p. 36571, Nov. 2016.

[4] E. Rahimian, S. Zabihi, S. F. Atashzar, A. Asif, and A. Mohammadi, "Xceptiontime: Independent time-window xceptiontime architecture for hand gesture classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1304–1308.

[5] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, Jul. 2013.

[6] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.

[7] S. Kong and D. Ramanan, "OpenGAN: Open-set recognition via open data generation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 813–822.

[8] H.-M. Yang, X.-Y. Zhang, F. Yin, Q. Yang, and C.-L. Liu, "Convolutional prototype network for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2358–2370, May 2022.

[9] J. Tomczynski, P. Kaczmarek, and T. Mankowski, "Hand gesture-based interface with multichannel sEMG band enabling unknown gesture discrimination," in *Proc. 10th Int. Workshop Robot Motion Control (RoMoCo)*, 2015, pp. 52–57.

[10] L. Wu, X. Zhang, X. Zhang, X. Chen, and X. Chen, "Metric learning for novel motion rejection in high-density myoelectric pattern recognition," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107165.

[11] L. Wu, A. Liu, X. Zhang, X. Chen, and X. Chen, "Unknown motion rejection in myoelectric pattern recognition using convolutional prototype network," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4305–4314, Mar. 2022.

[12] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PLoS ONE*, vol. 12, no. 10, Oct. 2017, Art. no. e0186132.

[13] F. Palermo, M. Cognolato, A. Gijsberts, H. Müller, B. Caputo, and M. Atzori, "Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1154–1159.

[14] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *J. NeuroEng. Rehabil.*, vol. 14, no. 1, pp. 1–14, Dec. 2017.

[15] X. Sun, Z. Yang, C. Zhang, K.-V. Ling, and G. Peng, "Conditional Gaussian distribution learning for open set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13480–13489.

[16] A. Phinyomark and E. Scheme, "EMG pattern recognition in the era of big data and deep learning," *Big Data Cognit. Comput.*, vol. 2, no. 3, p. 21, Aug. 2018.

[17] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-electromyography-based gesture recognition by multi-view deep learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019.

[18] A. Doswald, F. Carrino, and F. Ringeval, "Advanced processing of sEMG signals for user independent gesture recognition," in *Proc. 13th Medit. Conf. Med. Biol. Eng. Comput.*, Seville, Spain. Cham, Switzerland: Springer, 2014, pp. 758–761.

[19] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS ONE*, vol. 13, no. 10, Oct. 2018, Art. no. e0206049.

[20] X. Chen, Y. Li, R. Hu, X. Zhang, and X. Chen, "Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1292–1304, Apr. 2021.

[21] Y. Jiang, L. Song, J. Zhang, Y. Song, and M. Yan, "Multi-category gesture recognition modeling based on sEMG and IMU signals," *Sensors*, vol. 22, no. 15, p. 5855, Aug. 2022.

[22] L. Xu, K. Zhang, G. Yang, and J. Chu, "Gesture recognition using dual-stream CNN based on fusion of sEMG energy kernel phase portrait and IMU amplitude image," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103364.

[23] A. Fatayer, W. Gao, and Y. Fu, "sEMG-based gesture recognition using deep learning from noisy labels," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 9, pp. 4462–4473, Sep. 2022.

[24] F. Li and H. Wechsler, "Open set face recognition using transduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1686–1697, Nov. 2005.

[25] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, Jun. 2016, pp. 1563–1572.

[26] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial reciprocal points learning for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8065–8081, Nov. 2022.

[27] G. Chen et al., "Learning open set network with discriminative reciprocal points," in *Computer Vision—ECCV*. Glasgow, U.K.: Springer, 2020, pp. 507–522.

[28] H.-M. Yang, X.-Y. Zhang, F. Yin, and C.-L. Liu, "Robust classification with convolutional prototype learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3474–3482.

[29] J. Cen, P. Yun, J. Cai, M. Y. Wang, and M. Liu, "Deep metric learning for open world semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15333–15342.

[30] Z. Ge, S. Demyanov, Z. Chen, and R. Garnavi, "Generative OpenMax for multi-class open set classification," 2017, *arXiv:1707.07418*.

[31] L. Ditria, B. J. Meyer, and T. Drummond, "OpenGAN: Open set generative adversarial networks," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Nov. 2020, pp. 1–19.

[32] P. Perera et al., "Generative-discriminative feature representations for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11811–11820.

[33] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4016–4025.

[34] J. W. Robertson, K. B. Englehart, and E. J. Scheme, "Effects of confidence-based rejection on usability and error in pattern recognition-based myoelectric control," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2002–2008, Sep. 2019.

[35] E. J. Scheme, B. S. Hudgins, and K. B. Englehart, "Confidence-based rejection for improved pattern recognition myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1563–1570, Jun. 2013.

[36] Q. Ding, X. Zhao, J. Han, C. Bu, and C. Wu, "Adaptive hybrid classifier for myoelectric pattern recognition against the interferences of outlier motion, muscle fatigue, and electrode doffing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1071–1080, May 2019.

[37] Q. Ding, Z. Li, X. Zhao, Y. Xiao, and J. Han, "Real-time myoelectric prosthetic-hand control to reject outlier motion interference using one-class classifier," in *Proc. 32nd Youth Academic Annu. Conf. Chinese Assoc. Automat. (YAC)*, 2017, pp. 96–101.

[38] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[39] C. Kaae Sønderby, T. Raiko, L. Maaløe, S. Kaae Sønderby, and O. Winther, "Ladder variational autoencoders," 2016, *arXiv:1602.02282*.

[40] P. Oza and V. M. Patel, "C2AE: Class conditioned auto-encoder for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2302–2311.

[41] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Improved gesture recognition based on sEMG signals and TCN," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1169–1173.

[42] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[43] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible $1 \times 1$ convolutions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Montréal, QC, Canada: Curran Associates, 2018, pp. 1–10.

[44] Y. Song and S. Ermon, "Denoising diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 9248–9259.