

Guided Disentangled Representation Learning for Multi-Factor EMG Analysis

Masaki Matsuoka^{a,*}, Xinyue Niu^a, Akira Furui^{a,*}

^a*Graduate School of Advanced Science and Engineering, Hiroshima University, 1-7-1 Kagamiyama, Higashihiroshima, Hiroshima, 739-8521, Japan*

Abstract

Surface electromyographic (EMG) signals, widely used in rehabilitation and prosthetic control, are inherently influenced by multiple generative factors, including force level, forearm orientation, and movement class. Because these factors affect signal characteristics in an entangled manner, where variations in one factor alter features associated with others, EMG-based pattern recognition systems often suffer from degraded robustness and limited generalizability. To address this challenge, we propose a variational autoencoder (VAE)-based framework that learns disentangled latent representations of EMG signals by structuring the latent space around semantically meaningful generative factors. The proposed method combines factor-specific predictors with adversarial gradient reversal. This design aligns each latent subspace with a designated factor and suppresses cross-subspace information leakage, rather than enforcing complete statistical independence among latent variables. Experiments on two public EMG datasets with controlled variations in multiple generative factors, including force level, forearm orientation, and movement class, demonstrate that the proposed framework achieves effective disentanglement, as confirmed by both latent space visualizations and quantitative disentanglement metrics. The learned representations consistently improve downstream multi-task recognition performance and remain effective across different acquisition setups and factor sets, indicating enhanced robustness to variations in underlying factors. Furthermore, latent space traversal analysis reveals that manipulating factor-specific subspaces produces continuous and physically plausible variations in the generated EMG signals, suggesting the potential for controllable EMG data synthesis. These results demonstrate that guided disentangled representation learning provides an effective and principled framework for simultaneously enhancing recognition robustness and enabling interpretable signal generation in EMG applications.

Keywords: Disentangled representation learning, electromyogram (EMG), variational autoencoder (VAE), pattern recognition, data generation.

1. Introduction

Surface electromyographic (EMG) signals are noninvasively recorded bioelectrical signals that reflect muscle activation during contraction. Over the past decades, EMG pattern recognition has contributed substantially to a wide range of applications, including rehabilitation (Cesqui et al., 2013), prosthetic control (Fougner et al., 2012), and human-computer interaction (Xiong et al., 2021).

However, EMG signals are inherently sensitive to various sources of variability, such as changes in force level (Al-Timemy et al., 2013; Scheme and Englehart, 2013; Tkach et al., 2010; Al-Timemy et al., 2015) and forearm posture (Peng et al., 2013; Adewuyi et al., 2017; Pelaez Murciego et al., 2022; Khushaba et al., 2016). Such variations often lead to substantial performance degradation when models are applied under conditions different from those seen during training. To address this issue, these studies have explored strategies such as reducing the influence of confounding factors or incorporating training data collected under diverse conditions, yet exhaustively covering all possible conditions is costly and often impractical.

In addition, practical applications such as prosthetic control require the simultaneous consideration of multiple factors beyond movement class. For example, force levels must be adjusted according to the manipulated object and task context, and the same movement may be executed under different forearm orientations. Motivated by these requirements, several studies have attempted to simultaneously extract multiple factors from EMG signals (Hua et al., 2020; Leone et al., 2023, 2019; Yan et al., 2025; Fang et al., 2022). However, these factors influence EMG signals in an entangled manner—variations in one factor can alter signal characteristics associated with others—making joint extraction challenging and often resulting in sub-optimal representations where the estimation accuracy of individual factors is compromised. For instance, a feature representation optimized for movement classification may inadvertently discard force-level information, while shared representations that attempt to preserve all factors tend to suffer from mutual interference.

Disentangled representation learning (Bengio et al., 2013) aims to learn latent representations in which distinct underlying factors are independently encoded. This approach has been shown to benefit both recognition and generative tasks across diverse domains, including images, videos, natural language, and multimodal data (Wang et al., 2024). Motivated by these advances, we hypothesize that explicitly disentangling the la-

*Corresponding authors.

Email addresses: masakimatsuoka@hiroshima-u.ac.jp (Masaki Matsuoka), xinyueniu@hiroshima-u.ac.jp (Xinyue Niu), akirafurui@hiroshima-u.ac.jp (Akira Furui)

tent factors underlying EMG signals can address the aforementioned challenges. Specifically, disentangled representations can enable flexible recognition frameworks that simultaneously extract information about multiple factors while remaining robust to variations in other factors. Moreover, they naturally support controllable EMG signal generation conditioned on specific factor configurations, thereby reducing data collection costs and facilitating broader practical applications.

To realize this objective, we propose a disentangled representation learning framework for EMG signals based on a variational autoencoder (VAE) (Kingma and Welling, 2014). The proposed method structures the latent space into factor-specific subspaces, each aligned with a corresponding generative factor—force level, forearm orientation, and movement class—through supervisory signals. By jointly optimizing generative reconstruction and factor-specific supervision, the model learns disentangled latent representations aligned with the underlying generative factors of EMG signals. Furthermore, gradient reversal layer is incorporated to suppress the encoding of irrelevant factor information across subspaces, thereby reducing interference among the learned representations. The objective of this work is to align semantically meaningful generative factors with dedicated latent subspaces (factor-aligned disentanglement), rather than to enforce complete statistical independence among latent variables. Because the underlying factors of EMG are intrinsically coupled at the physiological level, exact statistical independence is neither attainable nor required for the targeted recognition and generation tasks.

The contributions of this study are summarized as follows:

- **A disentangled representation learning framework for multi-factor EMG signals.** We formulate the disentanglement of multiple kinematic and physiological factors in EMG signals as an explicit representation learning objective, and propose a VAE-based framework that integrates guided supervision with adversarial mechanisms to achieve factor-specific latent encoding.
- **Improved multi-task EMG recognition enabled by disentangled representations.** The learned factor-specific representations reduce interference among underlying factors, leading to consistent improvements in downstream multi-task EMG recognition across subjects.
- **Controllable EMG signal generation via latent space manipulation.** By independently manipulating factor-specific latent variables, the proposed method enables the generation of EMG signals with desired characteristics, which is experimentally validated through latent traversal analyses.

2. Related Work

2.1. EMG Pattern Recognition and Challenges

EMG pattern recognition aims to decode meaningful information from recorded EMG signals, including motion intentions, force-related muscle activation, and limb configurations. Over the past decades, a wide variety of machine

learning-based approaches, such as linear discriminant analysis (LDA) and support vector machines (SVM) (Scheme and Englehart, 2011), combined with handcrafted feature sets have been proposed. More recently, deep learning-based methods have demonstrated superior performance owing to their capability for learning informative representations directly from raw EMG signals (Phinyomark and Scheme, 2018). However, EMG signals are inherently affected by multiple varying factors, including changes in muscle contraction intensity and limb posture (Al-Timemy et al., 2013; Scheme and Englehart, 2013; Tkach et al., 2010; Al-Timemy et al., 2015; Peng et al., 2013; Adewuyi et al., 2017; Pelaez Murciego et al., 2022; Khushaba et al., 2016). As a result, models trained under specific conditions often suffer from significant performance degradation when applied to different conditions, posing a major challenge for robust EMG pattern recognition.

Furthermore, practical applications often require the simultaneous estimation of multiple factors (Hua et al., 2020; Leone et al., 2023, 2019; Yan et al., 2025; Fang et al., 2022). For instance, object manipulation tasks involve concurrent modulation of force levels according to object properties, while hand posture must be adapted to the spatial configuration of the object. However, feature representations and models designed for a single factor may fail to preserve information relevant to other factors, limiting their applicability in multi-factor estimation scenarios. Moreover, since these factors influence EMG signals in an entangled manner, variations in one factor can degrade the estimation accuracy of others, making joint extraction from shared representations particularly challenging.

These observations suggest that independently modeling individual factors is insufficient for applications requiring multi-factor estimation, and that the entangled nature of these factors in EMG signals calls for representation learning approaches that can explicitly separate and preserve multiple factors within a shared latent space.

2.2. EMG Data Generation

Beyond representation learning, another strategy for improving the robustness of EMG pattern recognition is to expose models to data collected under diverse conditions. However, acquiring large-scale EMG data that adequately covers such variations is time-consuming and often impractical. Consequently, EMG data generation has attracted increasing research interest.

In recent years, deep generative models, including VAEs, generative adversarial networks (GANs), and diffusion models, have been widely adopted for biosignal synthesis across various domains. In the EMG domain, these approaches have also been explored for signal synthesis (Yazawa and Furui, 2026; Gouda et al., 2023; Chen et al., 2022; Xiong et al., 2024). These approaches model the overall distribution of EMG signals and generate statistically plausible samples by sampling from the learned distribution. However, the latent representations learned by these models typically lack clear physiological or kinematic interpretations, limiting the controllability and interpretability of the generated signals.

This limitation highlights the need for generative models that explicitly structure the latent space to reflect meaningful under-

lying factors, enabling independent control and recombination of different kinematic and physiological factors in EMG signals.

2.3. Disentangled Representation Learning

Disentangled representation learning aims to learn latent representations in which different underlying factors of variation are encoded into separate and semantically meaningful components. By factorizing these factors into independent latent components, this approach enhances interpretability, controllability, and generalization. Owing to these advantages, disentangled representation learning has been extensively studied in a variety of settings, including improving inference task performance (Locatello et al., 2020), addressing out-of-distribution issues (Dittadi et al., 2021), few-shot learning (Tokmakov et al., 2019), domain generalization (Lee et al., 2021), controllable data generation (Higgins et al., 2017; Dupont, 2018), and data augmentation (Xu et al., 2021).

In EMG research, disentangled representation learning has attracted increasing attention, with studies primarily focusing on separating subject-specific information from movement-related information. Fan *et al.* were the first to introduce the concept of disentanglement into EMG feature representations (Fan et al., 2024). Their method adopts a multi-encoder, single-decoder architecture to decompose EMG signals into gesture-specific and subject-specific feature components, demonstrating the superiority of the resulting disentangled features across various downstream tasks. However, their study relies on high-density EMG, which is less practical than sparse multi-channel EMG systems, and depends on hand-crafted features and additional classifiers, limiting end-to-end applicability. To address these limitations, Niu *et al.* proposed a dual-branch architecture that independently encodes gesture-related and subject-related components (Niu and Furui, 2025), incorporating adversarial learning to achieve feature disentanglement from multi-channel EMG signals in an end-to-end manner. Zhang *et al.* investigated EMG feature disentanglement within a VAE framework (Zhang et al., 2025), decomposing EMG representations into subject-specific, gesture-specific, and sample-specific components to achieve effective domain generalization across subjects.

However, existing studies address disentanglement primarily as a means of cross-subject generalization, focusing on the separation of subject-specific and gesture-specific factors. Disentangled representation learning that targets multiple kinematic and physiological factors within an individual subject’s EMG signals, such as force level, forearm orientation, and movement class, has yet to be explored.

3. Proposed Method

3.1. Problem Settings

Let $\mathcal{D} = \{(\mathbf{X}_i, \mathbf{y}_i^{(1)}, \dots, \mathbf{y}_i^{(K)})\}_{i=1}^N$ be a training dataset containing N EMG samples from a single subject, where $\mathbf{X}_i \in \mathbb{R}^{L \times D}$ is an EMG segment of window length L and D channels, and

$\mathbf{y}_i^{(k)} \in \{0, 1\}^{N_k}$ is the one-hot encoded label for the k -th generative factor with N_k classes.

In this study, we consider three generative factors ($K = 3$): force level ($\mathbf{y}_i^{\text{fl}} \in \{0, 1\}^{N_{\text{fl}}}$), forearm orientation ($\mathbf{y}_i^{\text{fo}} \in \{0, 1\}^{N_{\text{fo}}}$), and movement class ($\mathbf{y}_i^{\text{mo}} \in \{0, 1\}^{N_{\text{mo}}}$). Our goal is to learn factor-specific latent representations $\mathbf{z}_i^{\text{fl}}$, $\mathbf{z}_i^{\text{fo}}$, and $\mathbf{z}_i^{\text{mo}}$ from $\mathbf{X}_i$ such that each representation encodes information about its corresponding factor while remaining invariant to the others.

Fig. 1 illustrates the overall architecture of the proposed framework, and Fig. 2 details the design of the guided mechanism. The proposed model consists of two main components: (1) a VAE-based generative backbone that learns a structured latent representation, and (2) a factor-guided mechanism that enforces each latent subspace to encode its designated generative factor.

3.1.1. Generative Model

We model EMG signal generation using a VAE whose latent space is explicitly partitioned to reflect the underlying factors of variation. The latent variable $\mathbf{z}_i$ is decomposed into four factor-specific components: $\mathbf{z}_i^{\text{fl}}$ for force level, $\mathbf{z}_i^{\text{fo}}$ for forearm orientation, $\mathbf{z}_i^{\text{mo}}$ for movement class, and $\mathbf{z}_i^{\text{r}}$ for residual variations not captured by the above factors. This residual subspace is intended to absorb factor-irrelevant nuisance variation, such as trial-to-trial variability and measurement noise, rather than any of the three designated factors.

The prior distribution of the concatenated latent variable $\mathbf{z}_i = (\mathbf{z}_i^{\text{fl}}, \mathbf{z}_i^{\text{fo}}, \mathbf{z}_i^{\text{mo}}, \mathbf{z}_i^{\text{r}})$ is assumed to be a standard normal distribution $p(\mathbf{z}_i) = \mathcal{N}(\mathbf{z}_i | \mathbf{0}, \mathbf{I})$. Given $\mathbf{z}_i$, the EMG signal $\mathbf{X}_i$ is modeled as:

$$p_{\theta}(\mathbf{X}_i | \mathbf{z}_i) = \mathcal{N}(\mathbf{X}_i | \mu_{\theta}(\mathbf{z}_i), \mathbf{I}), \quad (1)$$

where $\mu_{\theta}(\mathbf{z}_i)$ is the output of a decoder parameterized by θ .

Since true posterior $p_{\theta}(\mathbf{z}_i | \mathbf{X}_i)$ is intractable in deep generative models, the VAE framework (Kingma and Welling, 2014) introduces an encoder network to approximate it with a factorized Gaussian distribution:

$$q_{\phi}(\mathbf{z}_i | \mathbf{X}_i) = \mathcal{N}(\mathbf{z}_i | \mu_{\phi}(\mathbf{X}_i), \text{diag}(\sigma_{\phi}(\mathbf{X}_i))), \quad (2)$$

where $\mu_{\phi}(\mathbf{X}_i)$ and $\text{diag}(\sigma_{\phi}(\mathbf{X}_i))$ are produced by an encoder parameterized by ϕ . In the remainder of this section, we use (k) with $k \in \{\text{fl}, \text{fo}, \text{mo}\}$ to collectively index factor-specific latent variables $\mathbf{z}_i^{(k)}$ and their corresponding labels $\mathbf{y}_i^{(k)}$.

3.1.2. Guided Mechanism

To explicitly enforce disentanglement, a guided mechanism is introduced to align each latent subspace with its designated factor. This mechanism consists of two complementary components: factor-specific predictors and adversarial discriminators.

For each factor $k \in \{\text{fl}, \text{fo}, \text{mo}\}$, a predictor $F^{(k)}$ is applied to the corresponding latent variable $\mathbf{z}_i^{(k)}$ to extract factor-specific information. In parallel, a discriminator $G^{(k)}$ receives the remaining latent variables through a gradient reversal layer (GRL) to suppress the encoding of k -th factor information in non-corresponding subspaces. Specifically, the predictions are defined as

$$\hat{\mathbf{y}}_i^{(k)} = F^{(k)}(\mathbf{z}_i^{(k)}), \quad (3)$$

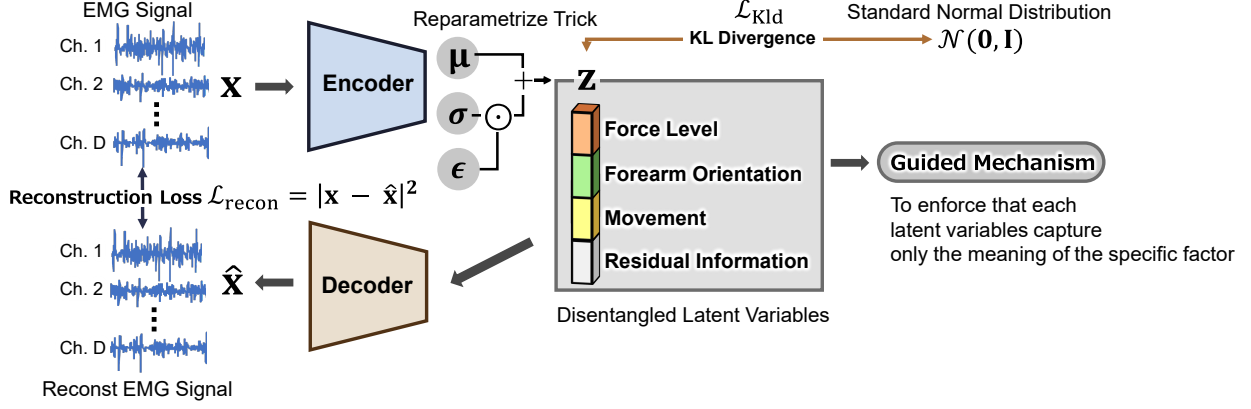


Figure 1: Overview of the proposed method. The raw EMG signals are first encoded to infer the posterior distribution $q_\phi(\mathbf{z} | \mathbf{X})$. Latent variables are sampled via the reparameterization trick, $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The resulting latent representation is semantically partitioned into four subspaces and processed by the guided mechanism to promote disentanglement. Meanwhile, the latent variables are fed into the decoder to reconstruct the EMG signal $\hat{\mathbf{X}}$.

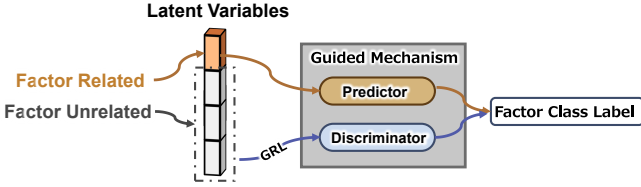


Figure 2: Details of the guided mechanism. Latent variables associated with a specific factor are fed into the corresponding predictor to infer the factor label, while the remaining latent variables are passed to the discriminator. Although the discriminator is trained to predict the factor class, a gradient reversal layer (GRL) is inserted to reverse the gradients during backpropagation, thereby encouraging factor-irrelevant latent variables to suppress information related to that factor.

$$\hat{\mathbf{y}}_i^{\text{adv},(k)} = G^{(k)}(\text{GRL}(\mathbf{z}_i^{\setminus k})), \quad (4)$$

where $\mathbf{z}_i^{\setminus k}$ denotes the concatenation of all latent subspaces other than $\mathbf{z}_i^{(k)}$, i.e., the factor-specific subspaces for $k' \neq k$ together with the residual subspace $\mathbf{z}_i^r$. The GRL inverts the gradient sign during backpropagation, enabling the encoder to learn representations from which the discriminator cannot predict the k -th factor, thereby suppressing information leakage across subspaces.

Through this design, the predictors encourage each subspace to encode its designated factor, while the discriminators discourage the leakage of factor-related information into the non-corresponding subspaces. Since $\mathbf{z}_i^{\setminus k}$ includes the residual subspace $\mathbf{z}_i^r$ for every factor k , the residual subspace is likewise penalized from encoding any factor and is thus constrained to carry only factor-irrelevant variation.

3.2. Objective Function

The model is trained by jointly optimizing the VAE generative objective, factor-guided constraints, and reconstruction refinement terms.

3.2.1. VAE Generative Objective

The generative model is optimized under the β -VAE formulation:

$$\mathcal{L}^{\text{vae}} = \sum_{i=1}^N (-\mathbb{E}_{q_\phi(\mathbf{z}_i | \mathbf{X}_i)} [\log p_\theta(\mathbf{X}_i | \mathbf{z}_i)] + \beta \cdot \text{KL}[q_\phi(\mathbf{z}_i | \mathbf{X}_i) \parallel p(\mathbf{z}_i)]), \quad (5)$$

where $\text{KL}(\cdot)$ denotes the Kullback–Leibler divergence and β controls the regularization strength on the latent representation.

3.2.2. Factor-Guided Constraints

Disentanglement is achieved by jointly minimizing the classification losses on factor-specific subspaces and maximizing adversarial losses on the remaining subspaces. Both objectives are implemented as cross-entropy losses for each factor $k \in \{\text{fl}, \text{fo}, \text{mo}\}$:

$$\mathcal{L}_k^{\text{cls}} = - \sum_{i=1}^N \mathbf{y}_i^{(k)} \log \hat{\mathbf{y}}_i^{(k)}, \quad (6)$$

$$\mathcal{L}_k^{\text{adv}} = - \sum_{i=1}^N \mathbf{y}_i^{(k)} \log \hat{\mathbf{y}}_i^{\text{adv},(k)}, \quad (7)$$

where $\hat{\mathbf{y}}_i^{(k)}$ and $\hat{\mathbf{y}}_i^{\text{adv},(k)}$ are the outputs of the predictor and discriminator defined in (3) and (4), respectively. The overall classification and adversarial losses are given by

$$\mathcal{L}^{\text{cls}} = \sum_k \mathcal{L}_k^{\text{cls}}, \quad \mathcal{L}^{\text{adv}} = \sum_k \mathcal{L}_k^{\text{adv}}. \quad (8)$$

The gradients of $\mathcal{L}^{\text{adv}}$ are reversed by the GRL during backpropagation, effectively suppressing mutual interference among latent subspaces.

3.2.3. Reconstruction Refinement

Preliminary investigations indicate that a standard VAE with only a mean squared error (MSE)-based reconstruction loss

tends to inadequately preserve amplitude-related characteristics in EMG signal reconstruction. To mitigate this limitation, two additional loss terms are introduced to explicitly constrain the signal envelope and spectral content.

The envelope loss penalizes discrepancies in the signal envelopes between the input $\mathbf{X}_i$ and its reconstruction $\hat{\mathbf{X}}_i$:

$$\mathcal{L}^{\text{env}} = \sum_{i=1}^N \|\text{env}(\mathbf{X}_i) - \text{env}(\hat{\mathbf{X}}_i)\|^2, \quad (9)$$

where $\text{env}(\cdot)$ computes the signal envelope by taking the mean absolute value (MAV) and applying a moving-average filter. The spectral loss similarly constrains the amplitude spectrum obtained via the discrete Fourier transform (DFT):

$$\mathcal{L}^{\text{spec}} = \sum_{i=1}^N \|\text{spec}(\mathbf{X}_i) - \text{spec}(\hat{\mathbf{X}}_i)\|^2, \quad (10)$$

where $\text{spec}(\cdot)$ denotes the amplitude spectrum computed via the DFT.

The final training objective combines all loss terms:

$$\mathcal{L} = \mathcal{L}^{\text{vae}} + \lambda_1 \mathcal{L}^{\text{cls}} - \lambda_2 \mathcal{L}^{\text{adv}} + \lambda_3 \mathcal{L}^{\text{spec}} + \lambda_4 \mathcal{L}^{\text{env}}, \quad (11)$$

where $\lambda_1, \lambda_2, \lambda_3$, and λ_4 are weighting coefficients. The negative sign on $\mathcal{L}^{\text{adv}}$ indicates that the discriminators are trained to maximize their prediction accuracy, while the GRL reverses the corresponding gradients with respect to the encoder, encouraging the non-corresponding subspaces to suppress factor-related information.

4. Experiments

4.1. Datasets and Processing

We conducted experiments using two publicly available datasets from the Biosignals Repository. The primary dataset (Dataset 1) (Khushaba et al., 2016) provides label information for three factors: movement class, forearm orientation, and force level. To assess generalizability across acquisition setups and factor sets, we additionally employ a second dataset (Dataset 2) (Khushaba et al., 2014), which provides labels for two factors: forearm orientation and movement class. Since this study relies entirely on previously published datasets and does not involve direct interaction with human subjects, no additional ethical approval was required.

For Dataset 1, EMG signals were collected using a Delsys Bagnoli Desktop EMG System equipped with six electrodes placed around the forearm at a sampling rate of 4000 Hz. A 3D accelerometer (MPU6050, InvenSense) was also attached to the wrist during data collection, but only the EMG signals were used in this study. Each subject performed six movements: grasping (C1), opening (C2), extension (C3), flexion (C4), ulnar deviation (C5), and radial deviation (C6) under three forearm orientations: fully supinated (supination), neutral, and fully pronated (pronation). For each orientation, movements were performed at three force levels: low, medium, and high. Each movement–orientation–force combination was repeated three

times. Force levels were not measured instrumentally; instead, subjects were instructed to use the low level as a reference and to progressively increase their effort for the medium and high levels. In total, 162 trials were collected per subject (6 movements $\times$ 3 orientations $\times$ 3 force levels $\times$ 3 repetitions). Each trial lasted five seconds, followed by a 10-second rest period to minimize fatigue. Data from 10 subjects were used.

For Dataset 2, EMG signals were recorded with seven electrodes (Delsys DE 2.x series sensors processed by the Bagnoli desktop EMG system) placed with approximately equal spacing around the forearm circumference at a sampling rate of 4000 Hz. Each subject performed eight hand motions at five forearm/limb positions (treated as the forearm-orientation factor for this dataset), and each motion–position condition was repeated over six trials. The rest class was excluded from the movement factor, leaving seven movement classes (wrist flexion, wrist extension, pronation, supination, power grip, pinch grip, and open hand). Of the 11 subjects in the dataset, one was excluded because of an inconsistent channel count, leaving 10 subjects for analysis. This dataset thus provides a complementary two-factor setting (orientation $\times$ movement) recorded with a different acquisition setup and a different electrode count.

For both datasets, power-line interference was first removed using a second-order digital notch filter with a rejection band of 49–52 Hz. The EMG signals were then peak-normalized by dividing all channels by the global peak absolute value and downsampled to 500 Hz. Both datasets were recorded at 4000 Hz, and downsampling was performed using the `decimate` function (SciPy) with a decimation factor of 8 (4000 Hz to 500 Hz), which applies an 8th-order Chebyshev Type I anti-aliasing low-pass filter prior to decimation, thereby suppressing spectral content above the new Nyquist frequency of 250 Hz. Together with the notch filter, the effective retained bandwidth is approximately 20–250 Hz, which covers the dominant band of surface EMG. The 500 Hz rate was chosen because the discriminative information for EMG pattern recognition is largely concentrated below 250 Hz; this choice preserves the relevant band while reducing input dimensionality and computational cost, consistent with prior EMG pattern-recognition studies operating at comparable sampling rates (Li et al., 2011; Chen et al., 2017). The signals were finally segmented using a sliding window of 200 ms with a step size of 50 ms, yielding input samples of size 100 along the temporal axis.

We adopt a within-subject protocol in which model training and evaluation are performed independently for each subject; thus the 10 subjects of each dataset correspond to 10 independent within-subject validations. For each condition, the recorded repetitions were partitioned into disjoint training and test sets at the repetition level: for Dataset 1, repetitions 1 and 2 were used for training and repetition 3 for testing, whereas for Dataset 2, trials 1–3 were used for training and trials 4–6 for testing. The sliding-window segmentation was applied within each split separately to prevent leakage of windows across the training and test sets.

4.2. Comparison Methods

To validate the effectiveness of the proposed method, we compare it against a set of baselines and, separately, an ablation study.

Baselines. We compare the proposed method with representative methods spanning linear dimensionality reduction, unsupervised disentanglement, GAN-based disentanglement, and supervised prediction:

- **PCA and ICA** (Pearson, 1901; Comon, 1994): Classical linear dimensionality-reduction methods applied to the EMG features.
- **β -VAE** (Higgins et al., 2017): The proposed generative backbone without the guided mechanism, evaluating the necessity of explicit factor supervision for achieving disentanglement.
- **β -TCVAE** (Chen et al., 2018) and **FactorVAE** (Kim and Mnih, 2018): Unsupervised VAEs that promote disentanglement by penalizing the total correlation among latent variables.
- **InfoGAN** (Chen et al., 2016): A representative GAN-based disentanglement method that maximizes the mutual information between a subset of latent codes and the generated signals.
- **CNN Multi-Predict**: A supervised CNN that shares the same encoder architecture and is trained solely to predict the factors available in each dataset. Its penultimate features are used as the learned representation. This baseline isolates predictability from disentanglement.

The selected baselines share an explicit latent representation whose alignment with the generative factors can be directly quantified. Two related paradigms fall outside this scope: causal representation learning identifies factors from interventional data, which standard EMG acquisition does not provide (Schölkopf et al., 2021), and diffusion-based generative models prioritize sample fidelity over a factor-aligned, controllable latent space. Among GAN-based methods, InfoGAN yields such a latent structure and serves as the representative comparison.

Ablation. To examine how the number of supervised factors affects disentanglement, we also evaluate **partially supervised** variants of the proposed model, in which the guided mechanism supervises only a subset of the factors (i.e., supervision for two of the three factors in Dataset 1, or one of the two factors in Dataset 2).

4.3. Model Configuration and Training Parameters

The architecture of each component is as follows.

- **Encoder**: A 4-layer one-dimensional convolutional neural network (1D-CNN) with output channels of 16, 32, 64, and 128 for each layer, respectively. Each convolutional layer uses a kernel size of 3, followed by batch normalization (BN), a parametric rectified linear unit (PReLU)

activation, and max pooling. The extracted features are compressed along the temporal dimension using an adaptive average pooling layer, and a fully connected network outputs the mean and log-variance of the approximate posterior.

- **Decoder**: A 4-layer transposed 1D-CNN with output channels of 64, 32, 16, and C , respectively, and kernel sizes of 3, 3, 2, and 2, where C is the number of EMG channels of the dataset (six for Dataset 1 and seven for Dataset 2). Each hidden layer is followed by BN and a PReLU activation, while the output layer applies a linear projection without a nonlinear activation.
- **Guided heads**: Each head consists of a 3-layer fully connected network with 256 units per hidden layer, followed by ReLU activation. A softmax function is applied at the output layer to produce a categorical probability distribution.

For all VAE-based models, the latent space dimensionality is set to 8. As a GAN-based model, InfoGAN uses its standard generator, discriminator, and auxiliary Q-network rather than the encoder–decoder described above, with the dimensionality of its latent codes matched to this 8-dimensional latent space for a fair comparison. In the proposed method and partially supervised baselines, the latent representation is evenly partitioned into four two-dimensional subspaces, with the factor-to-subspace assignment kept consistent across datasets. For Dataset 1, dimensions 1–2, 3–4, and 5–6 are aligned with force level, forearm orientation, and movement class, respectively, while dimensions 7–8 form an unsupervised residual subspace. For Dataset 2, which does not include a force factor, forearm orientation and movement retain the same assignment (dimensions 3–4 and 5–6), and the two remaining subspaces (dimensions 1–2 and 7–8) act as residual subspaces. The signal envelope used in $\mathcal{L}^{\text{env}}$ was computed by calculating the mean absolute value and smoothing it with 1D average pooling (window size 5, stride 3).

The training hyperparameters are summarized as follows. For the unsupervised disentanglement baselines, the hyperparameters were selected by grid search, adopting for each dataset the configuration that yielded the highest DCI Disentanglement score. On Dataset 1 and Dataset 2, respectively, this procedure selected $\beta = 0.001$ and 0.1 for the β -VAE, $(\alpha, \beta, \gamma) = (0.0001, 2, 0.01)$ and $(0.001, 1, 0.01)$ for the β -TCVAE, and $(\beta, \gamma) = (0.001, 0.01)$ and $(0.01, 3)$ for FactorVAE. For the proposed method and partially supervised variants, β was fixed to 1, as the factor-specific supervision imposes sufficient structural constraints on the latent space to prevent posterior collapse without relying on a small β . The loss weights were set to $\lambda_1 = 100$, $\lambda_2 = 1$, $\lambda_3 = 1$, and $\lambda_4 = 100$ for the classification, adversarial, spectral, and envelope losses, respectively. For the β -VAE baseline, which does not include classification or adversarial losses, the spectral and envelope loss weights were kept at $\lambda_3 = 1$ and $\lambda_4 = 100$. The VAE-based models—the proposed method, its partially supervised variants, and the β -VAE, β -TCVAE, and FactorVAE baselines—were trained using the

Adam optimizer with a batch size of 256 and a learning rate of 1.0×10^{-3} for 500 epochs. InfoGAN was trained for 2000 epochs, following common practice for stabilizing adversarial training, whereas PCA and ICA are closed-form and require no iterative training. Model training and evaluation were conducted independently for each subject.

4.4. Performance Evaluation

To systematically assess the proposed method, we analyze the learned latent space and evaluate its disentanglement capability, downstream recognition performance, and generative behavior. Unless otherwise stated, all evaluations follow the within-subject train/test split described above. The VAE-based models are trained on the training set, and the latent-space visualizations, DCI metrics, and reconstruction are computed on the held-out test set. For Dataset 1, the three factors (force level, forearm orientation, and movement class) are evaluated, whereas for Dataset 2 the two available factors (forearm orientation and movement class) are evaluated.

4.4.1. Latent Space Analysis

To qualitatively examine the structure of the learned latent space, we visualize latent representations by color-coding samples according to factor labels. This visualization allows us to assess whether each latent subspace predominantly captures a single semantic factor. For comparison, we also visualize the latent space learned by the β -VAE baseline. Since the β -VAE does not explicitly associate latent dimensions with specific factors, factor-wise logistic regression classifiers are trained to identify the two most predictive dimensions for each factor, which are then used for visualization. The details of dimension selection are reported in Section 5.1.

Furthermore, to examine whether the learned latent structure reflects continuous variations in force levels and forearm orientations, we conduct experiments in which certain factor levels are excluded from the training data and subsequently evaluate whether unseen levels are coherently embedded within the learned latent space during inference.

4.4.2. Quantitative Evaluation of Disentanglement

The DCI metric (Eastwood and Williams, 2018) is employed as a quantitative measure of disentanglement. It evaluates learned representations from three complementary aspects: *Disentanglement*, which indicates the extent to which each latent variable captures a single generative factor with minimal influence from other factors; *Completeness*, which reflects whether each factor is primarily encoded by a single latent variable rather than distributed across multiple ones; and *Informativeness*, which quantifies how accurately the ground-truth factors can be predicted from the latent representations.

In this study, the DCI metrics are computed by training a random forest classifier to predict the ground-truth factors from the latent variables. Feature importance scores obtained from the trained classifier are used to quantify disentanglement and completeness, while the cross-entropy loss of factor prediction is used to measure informativeness. The results are reported as averages across 10 subjects.

4.4.3. Downstream Task Evaluation

We evaluate whether the learned disentangled representations benefit downstream recognition and generation tasks.

For recognition, we adopt a multi-task prediction setting that simultaneously classifies the factors of each dataset (force level, forearm orientation, and movement class for Dataset 1; forearm orientation and movement class for Dataset 2). To ensure a fair comparison, two identical multi-task classifiers are independently trained: one using disentangled features extracted by the proposed method (denoted DRL-mt) and the other using features obtained from a CNN with the same encoder architecture, trained end-to-end for multi-task classification (denoted CNN-mt). For DRL-mt, the VAE encoder is first trained on the training set; the multi-task classifier is then trained on latent features obtained by encoding the training set and evaluated on the test set. The CNN-mt baseline is trained on the training set and evaluated on the test set under the identical split. All experiments are repeated three times with different random seeds.

For generation, we first assess reconstruction performance by comparing the reconstructed EMG signals with the original inputs. Reconstruction quality is evaluated using two metrics: the cross-channel envelope correlation (Env. Corr.), which measures whether the relative amplitude balance across channels is correctly preserved, and the spectral root mean squared error (Spec. RMSE), which quantifies the fidelity of frequency characteristics between the original and reconstructed signals. The effects of the envelope and spectral losses on reconstruction quality is analyzed by contrasting models trained with and without these terms. We further evaluate the controllability of data generation by varying the force-related latent dimension while fixing the remaining dimensions, and visualizing the resulting decoder outputs.

4.4.4. Robustness to Noise

To assess robustness under noisy conditions, we evaluate the disentanglement quality of the proposed method when the test signals are corrupted by additive Gaussian noise. The model is trained on clean data, and at test time zero-mean Gaussian noise is injected at signal-to-noise ratios (SNR) ranging from 0 to 40 dB, where the SNR is defined as $\text{SNR} = 10 \log_{10}(P_{\text{signal}}/P_{\text{noise}})$ with P_{signal} and P_{noise} denoting the average power of the signal and the noise, respectively. We report the resulting DCI scores (Disentanglement, Completeness, and Informativeness) as a function of SNR, averaged across subjects of Dataset 1.

5. Results

5.1. Latent Space Analysis

Fig. 3 presents the latent space visualizations for (a) the β -VAE and (b) the proposed method. As similar trends are observed across subjects, Subject 1 is shown as a representative example.

Since the β -VAE does not explicitly associate latent dimensions with specific factors, we train factor-wise logistic regression classifiers on the learned latent variables and select the two

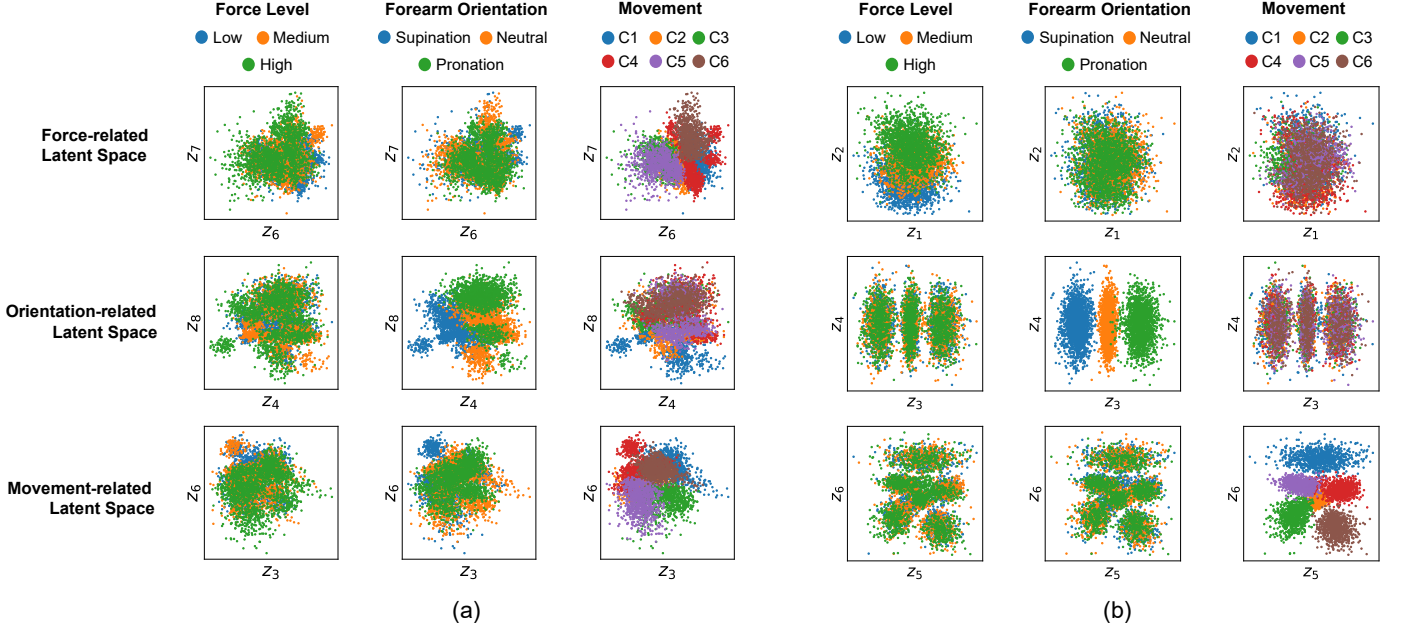


Figure 3: Latent space visualizations of (a) the β -VAE and (b) the proposed method for Subject 1. Each row corresponds to a latent subspace, and each column represents the factor used for color coding. For the β -VAE, factor-wise logistic regression classifiers are trained on the latent variables, and the two dimensions with the highest predictive performance for each factor are selected for visualization. For the proposed method, the latent space is explicitly structured such that z_1, z_2 encode force level, z_3, z_4 encode forearm orientation, and z_5, z_6 encode movement class.

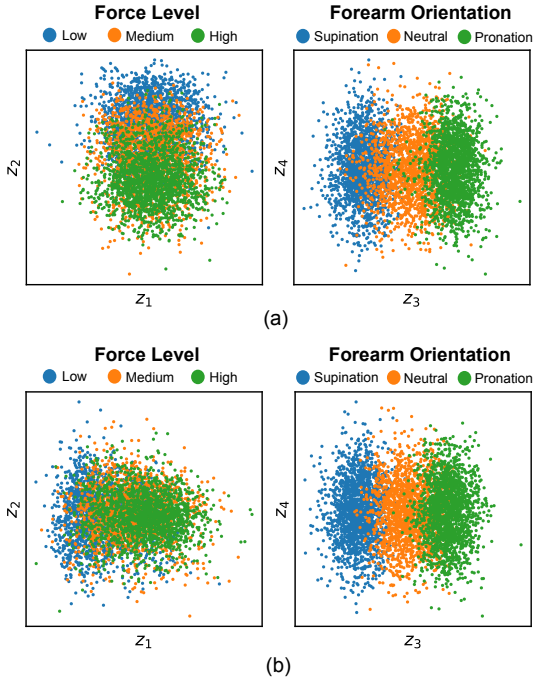


Figure 4: Latent space visualizations of the proposed method for Subject 1 when certain factor levels are excluded during training. (a) The medium force level and the neutral forearm orientation are excluded. (b) The high force level and the supinated forearm orientation are excluded.

most predictive dimensions for each factor. The resulting assignments are $\mathbf{z}^{\beta, \text{fl}} = (z_6, z_7)$ for force level, $\mathbf{z}^{\beta, \text{fo}} = (z_4, z_8)$ for forearm orientation, and $\mathbf{z}^{\beta, \text{mo}} = (z_3, z_6)$ for movement class.

In the β -VAE latent space, factor-related information is only partially captured, and the representations remain entangled. Samples corresponding to different force levels do not form clearly separable clusters in the force-related subspace, indicating that force information is not faithfully encoded. The forearm-orientation and movement subspaces exhibit partial separation along their designated factors, but simultaneously reflect other factors (e.g., the forearm-orientation subspace also reveals movement-related clustering) suggesting that multiple factors are concurrently encoded within each subspace.

In contrast, the proposed method yields a markedly different structure. Each factor is clearly separated into its corresponding latent subspace: samples form well-defined clusters when color-coded by the designated factor, yet remain largely intermixed under other factor colorings, indicating limited cross-subspace interference. These results demonstrate that the proposed method effectively learns disentangled representations in which each latent subspace predominantly encodes its designated factor.

Fig. 4 presents the results of the latent space continuity analysis, which examines whether continuous factors (force level and forearm orientation) are embedded with meaningful ordering and spacing, even for factor levels unseen during training. In Fig. 4 (a), the medium force level and neutral forearm orientation are excluded from training. During inference, samples from these unseen conditions are smoothly mapped to intermediate positions along the corresponding latent dimensions, situated between the adjacent observed levels. In Fig. 4 (b),

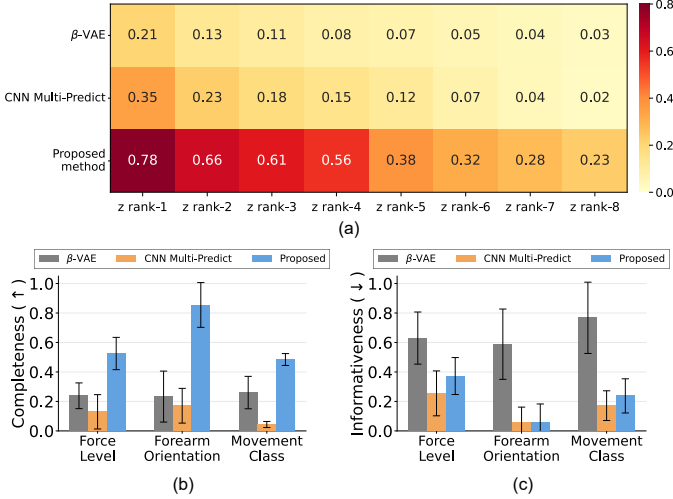


Figure 5: Detailed DCI analysis comparing the β -VAE, CNN Multi-Predict, and the proposed method on Dataset 1, averaged across 10 subjects. (a) Per-dimension Disentanglement scores. For each method, the eight latent dimensions are sorted in descending order of score—per subject, then averaged across subjects—and labelled z rank-1 through z rank-8; because the β -VAE and CNN Multi-Predict latents are not factor-aligned and the Disentanglement metric is factor-agnostic, this ranking enables a like-for-like comparison across methods. (b) Factor-wise Completeness scores. Higher values indicate that each factor is predominantly encoded by fewer latent variables. (c) Factor-wise Informativeness scores. Lower values indicate more accurate prediction of the ground-truth factors from the latent representations. Error bars in (b) and (c) indicate standard deviations across subjects.

the high force level and supination forearm orientation are excluded. Along the force-related latent dimension, the high-force condition is mapped to the opposite end relative to the low-force condition, with the medium force level positioned in between. Similarly, supination and pronation occupy opposite ends of the forearm-orientation dimension, while the neutral orientation lying between them. These observations indicate that the proposed method preserves meaningful ordering and continuity in the latent space, even for factor levels not observed during training.

5.2. Quantitative Evaluation of Disentanglement

Table 1 compares the DCI scores of the proposed method with those of the baselines on both datasets. On both Dataset 1 and Dataset 2, the proposed method achieves the highest Disentanglement and Completeness scores among all compared methods, indicating that factor-related information is more consistently concentrated within dedicated latent subspaces rather than distributed across multiple dimensions. This advantage holds by a large margin over the unsupervised baselines: the proposed method attains Disentanglement and Completeness of 0.644/0.621 on Dataset 1 and 0.808/0.664 on Dataset 2, several times higher than any unsupervised method (Table 1). The ablation in Table 2 further shows that, on both datasets, performance among the partially supervised variants improves as more factors receive explicit supervision, suggesting that comprehensive factor guidance is essential for effective disentanglement.

Two aspects of the baseline results warrant attention. First, the linear methods PCA and ICA attain moderate Disentan-

glement and Completeness scores, but this does not reflect genuine multi-factor disentanglement. Beyond the tabulated DCI metrics, a per-dimension linear-probe analysis (a logistic-regression classifier fitted per latent dimension on the held-out test representations; full results in the Supplementary Material) makes this explicit: for PCA on Dataset 1, only forearm orientation is decoded well above chance from a single latent dimension (≈ 0.74), whereas force level and movement stay near their chance levels (≈ 0.36 and ≈ 0.19 ; chance 0.333 and 0.167), averaged across the 10 subjects, and ICA shows the same single-factor pattern (≈ 0.77 for orientation). Their moderate scores therefore arise from cleanly encoding one factor while discarding the others; the VAE-based baselines, by contrast, encode all factors above chance but in an entangled form. Second, the supervised CNN Multi-Predict baseline attains the best Informativeness yet Disentanglement and Completeness far below those of the proposed method on both datasets (e.g., 0.169 versus 0.644 in Disentanglement on Dataset 1; Table 1). This dissociation, reproduced on both datasets, shows that accurate factor prediction does not by itself yield disentangled representations: a representation can predict the factors well while remaining strongly entangled. In contrast, the proposed method attains both high structural disentanglement and a low Informativeness loss (second only to CNN Multi-Predict on both datasets), indicating that the learned representations are simultaneously disentangled and informative.

That the same method ranking and the same predictability–disentanglement dissociation recur on Dataset 2—recorded under a different acquisition setup and with a different factor set (forearm orientation $\times$ movement)—indicates that the factor-aligned disentanglement achieved by the proposed method transfers across datasets rather than being specific to a single benchmark.

Fig. 5 provides a more detailed breakdown. The per-dimension Disentanglement scores in Fig. 5(a) show that the proposed method attains substantially higher scores than both the β -VAE and the CNN Multi-Predict baseline at every rank, confirming the structural effect of the guided mechanism. The factor-wise analysis in Fig. 5(b) and (c) further reveals that the improvements of the proposed method over the unsupervised β -VAE are consistent across all three factors; the CNN Multi-Predict baseline attains the best Informativeness among the three methods yet the lowest Completeness (Fig. 5(b)), and force level exhibits the highest Informativeness loss among the three factors. In terms of Completeness, forearm orientation achieves a substantially higher score than force level and movement class, which show comparably lower values.

5.3. Performance on Downstream Tasks

5.3.1. Performance on Multi-Task Recognition

Fig. 6 compares the classification accuracy of DRL-mt and CNN-mt for the multi-task recognition on Dataset 1, reported as the mean over three runs with different random seeds. Each point represents an individual subject, with points above the diagonal indicating improved performance by DRL-mt. Overall, the proposed DRL-mt achieves higher average accuracy than the baseline CNN-mt across all three factors: force

Table 1: DCI scores on Dataset 1 (three factors) and Dataset 2 (two factors), under the within-subject protocol (mean $\pm$ standard deviation over 10 subjects per dataset).

Model	Dataset 1			Dataset 2		
	Dis. ($\uparrow$)	Com. ($\uparrow$)	Inf. ($\downarrow$)	Dis. ($\uparrow$)	Com. ($\uparrow$)	Inf. ($\downarrow$)
<i>Unsupervised</i>						
PCA	0.148 $\pm$ 0.127	0.229 $\pm$ 0.218	0.737 $\pm$ 0.252	0.047 $\pm$ 0.045	0.034 $\pm$ 0.038	1.488 $\pm$ 0.059
ICA	0.164 $\pm$ 0.120	0.256 $\pm$ 0.193	0.766 $\pm$ 0.226	0.031 $\pm$ 0.028	0.030 $\pm$ 0.025	1.487 $\pm$ 0.055
β -VAE	0.102 $\pm$ 0.047	0.244 $\pm$ 0.084	0.662 $\pm$ 0.135	0.152 $\pm$ 0.113	0.172 $\pm$ 0.127	1.578 $\pm$ 0.097
β -TCVAE	0.070 $\pm$ 0.046	<u>0.293 $\pm$ 0.110</u>	0.988 $\pm$ 0.099	<u>0.294 $\pm$ 0.103</u>	<u>0.447 $\pm$ 0.046</u>	1.152 $\pm$ 0.101
FactorVAE	0.070 $\pm$ 0.046	0.285 $\pm$ 0.066	0.761 $\pm$ 0.138	0.221 $\pm$ 0.074	0.293 $\pm$ 0.084	1.247 $\pm$ 0.118
InfoGAN	0.089 $\pm$ 0.030	0.089 $\pm$ 0.032	0.663 $\pm$ 0.099	0.147 $\pm$ 0.105	0.128 $\pm$ 0.060	0.956 $\pm$ 0.234
<i>Supervised</i>						
CNN Multi-Predict	<u>0.169 $\pm$ 0.072</u>	0.115 $\pm$ 0.053	0.162 $\pm$ 0.082	0.120 $\pm$ 0.055	0.056 $\pm$ 0.026	0.137 $\pm$ 0.040
Proposed method	0.644 $\pm$ 0.088	0.621 $\pm$ 0.078	<u>0.223 $\pm$ 0.080</u>	0.808 $\pm$ 0.065	0.664 $\pm$ 0.083	<u>0.152 $\pm$ 0.073</u>

Bold and underlined values indicate the best and second-best per column, respectively.

Table 2: Ablation study: DCI scores of the partially supervised variants, in which the guided mechanism supervises only a subset of the factors (mean $\pm$ standard deviation over 10 subjects per dataset). The proposed (fully supervised) method is shown for reference.

Supervised factors	Dataset 1			Dataset 2		
	Dis. ($\uparrow$)	Com. ($\uparrow$)	Inf. ($\downarrow$)	Dis. ($\uparrow$)	Com. ($\uparrow$)	Inf. ($\downarrow$)
w/o force level	0.488 $\pm$ 0.089	0.456 $\pm$ 0.070	0.422 $\pm$ 0.093	—	—	—
w/o forearm orientation	0.314 $\pm$ 0.050	0.364 $\pm$ 0.056	0.570 $\pm$ 0.065	0.406 $\pm$ 0.050	0.320 $\pm$ 0.010	0.842 $\pm$ 0.032
w/o movement	0.433 $\pm$ 0.067	0.500 $\pm$ 0.070	0.709 $\pm$ 0.080	0.313 $\pm$ 0.063	0.406 $\pm$ 0.096	0.975 $\pm$ 0.067
Proposed (all factors)	0.644 $\pm$ 0.088	0.621 $\pm$ 0.078	0.223 $\pm$ 0.080	0.808 $\pm$ 0.065	0.664 $\pm$ 0.083	0.152 $\pm$ 0.073

“w/o” denotes a factor that is *not* supervised. “—” indicates a factor not present in the dataset.

level (79.88% vs. 77.25%), forearm orientation (97.35% vs. 96.29%), and movement (94.60% vs. 90.37%). These improvements are consistent at the individual-subject level, with accuracy gains observed in 7 out of 10 subjects for force level, 9 out of 10 for forearm orientation, and 7 out of 10 for movement classification.

The same advantage carries over to Dataset 2, whose two factors (forearm orientation and movement class) are evaluated in Fig. 7. Here as well, DRL-nt improves the average accuracy over CNN-nt for both factors—forearm orientation (93.79% vs. 87.10%; 10 of 10 subjects) and movement (95.30% vs. 90.82%; 9 of 10 subjects)—with the per-subject gains again concentrated above the diagonal. That the multi-task recognition benefit of the disentangled representations recurs on a dataset with a different acquisition setup and factor set confirms that the improvement is not specific to Dataset 1.

5.3.2. Generation Capacity Evaluation

We first examine the reconstruction ability of the proposed method, which serves as the foundation for EMG data generation. Fig. 8 visualizes the input signals (a) and the corresponding reconstructed signals obtained under different training configurations: (b) without the envelope and spectral losses, and (c) with both losses included. Without the envelope and spectral losses (Fig. 8 (b)), the reconstructed signals are largely concentrated around zero, failing to preserve amplitude characteristics. When both loss terms are incorporated (Fig. 8 (c)), the overall

Table 3: Reconstruction performance under different loss configurations, averaged across 10 subjects.

Loss Configuration	Env. Corr. ($\uparrow$)	Spectral RMSE ($\downarrow$) [$\times 10^{-3}$]
$\mathcal{L}^{\text{vae}}$ only	0.396 $\pm$ 0.209	5.35 $\pm$ 1.83
$\mathcal{L}^{\text{vae}} + \mathcal{L}^{\text{spec}}$	0.871 $\pm$ 0.060	3.02$\pm$1.01
$\mathcal{L}^{\text{vae}} + \mathcal{L}^{\text{env}}$	<u>0.910$\pm$0.037</u>	5.40 $\pm$ 1.79
$\mathcal{L}^{\text{vae}} + \mathcal{L}^{\text{spec}} + \mathcal{L}^{\text{env}}$	0.916$\pm$0.036	<u>3.04$\pm$1.00</u>

signal amplitude and the relative amplitude ratios across channels are reproduced more faithfully.

Table 3 quantifies these observations. With $\mathcal{L}^{\text{vae}}$ loss alone, the envelope correlation remains low, confirming the amplitude suppression observed in Fig. 8 (b). Incorporating $\mathcal{L}^{\text{env}}$ substantially improves envelope preservation, while $\mathcal{L}^{\text{spec}}$ primarily reduces spectral distortion. The combination of both losses achieves the best overall reconstruction fidelity.

Next, we evaluate whether factor-specific data generation can be achieved through latent variable manipulation. Fig. 9 illustrates EMG signal generated by varying the force-related latent variables while fixing the remaining dimensions. As similar trends are observed across subjects, a representative example from Subject 9, in whom force levels are relatively well separated, is shown for clarity. Traversing the force-related latent dimension produces diverse EMG signal patterns with a monotonic increase in signal amplitude, consistent with the known

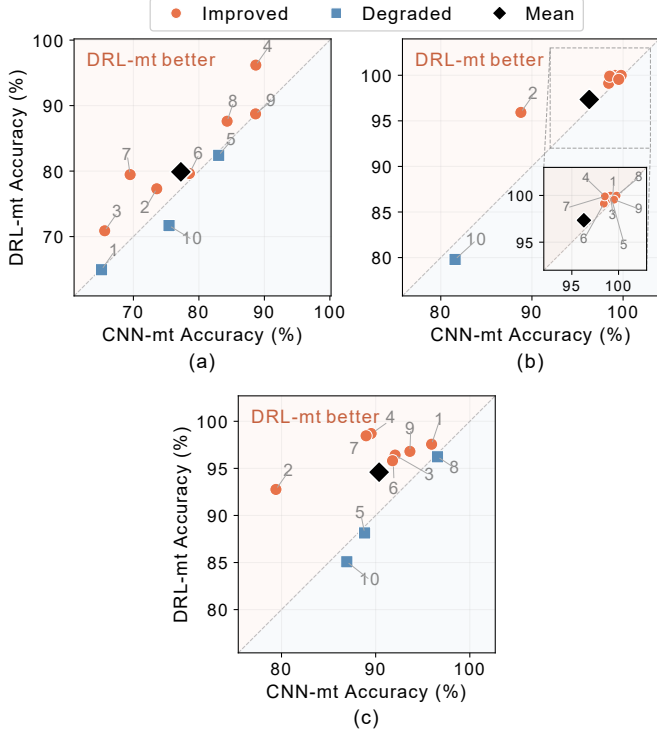


Figure 6: Multi-task recognition accuracy for (a) force level, (b) forearm orientation, and (c) movement class. The horizontal and vertical axes represent the accuracy (%) of CNN-mt and DRL-mt, respectively. The dashed diagonal line indicates equivalent performance; points above this line indicate that DRL-mt achieved higher accuracy than CNN-mt. Orange circles denote subjects whose performance improved, blue squares denote those whose performance declined, and black diamonds indicate the mean accuracy across the 10 subjects

physiological relationship between force exertion and EMG amplitude. These results suggest that the proposed method learns independently structured, factor-specific latent subspaces that enable controlled EMG signal generation through manipulation of individual latent variables.

5.4. Robustness to Noise

Fig. 10 shows the DCI scores of the proposed method on Dataset 1 as a function of the test-time SNR, where the dashed lines indicate the noise-free reference of the proposed method. As the SNR decreases, the Disentanglement and Completeness scores of the proposed method decline gracefully from their noise-free reference (dashed lines), remaining close to the clean-data values down to moderate SNR and dropping more markedly only at low SNR. Notably, even under noise these scores stay above the *noise-free* scores of all baseline methods (Table 1) down to moderate SNR levels, so the proposed method under perturbation still surpasses the undisturbed baselines in disentanglement structure. The Informativeness similarly worsens only gradually until low SNR. These results indicate that the factor-aligned structure learned by the proposed method is comparatively robust to additive noise.

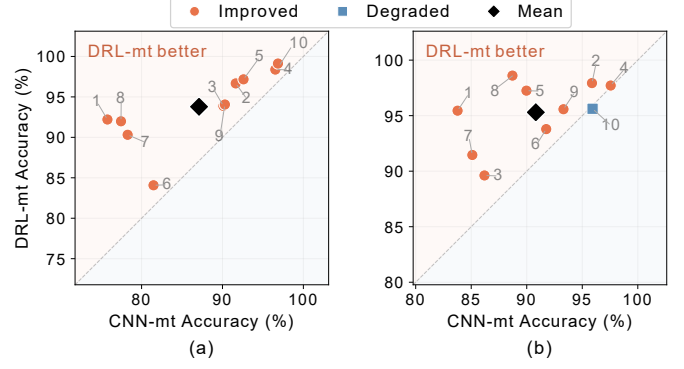


Figure 7: Multi-task recognition accuracy on Dataset 2 for (a) forearm orientation and (b) movement class. The horizontal and vertical axes represent the accuracy (%) of CNN-mt and DRL-mt, respectively. Points above the dashed diagonal indicate that DRL-mt achieved higher accuracy than CNN-mt. Orange circles denote subjects whose performance improved, blue squares denote those whose performance declined, and black diamonds indicate the mean accuracy across the 10 subjects.

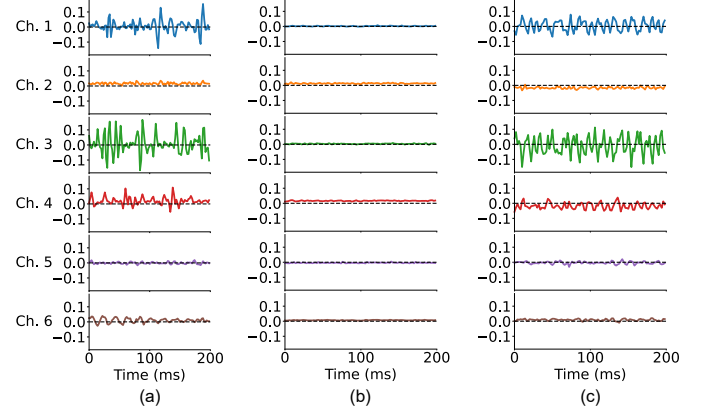


Figure 8: Reconstruction performance comparison for Subject 1. (a) Original EMG signal. (b) Reconstructed EMG signal without the envelope and spectral losses. (c) Reconstructed EMG signal with the proposed method including both losses. The vertical and horizontal axes indicate normalized amplitude and time (ms), respectively.

6. Discussion

6.1. Mechanism and Representation Learning

Both the latent space visualizations and the quantitative DCI evaluation indicate that the proposed method effectively learns disentangled latent representations. The per-dimension analysis in Fig. 5(a) further confirms that the proposed method attains substantially higher per-dimension Disentanglement than the β -VAE and the CNN Multi-Predict baseline across all ranks, evidencing the structural effect of the guided mechanism. This observation is consistent with the findings of Locatello *et al.* (Locatello *et al.*, 2019), who demonstrated that disentangled representations cannot be reliably identified without appropriate inductive biases, such as supervision or explicit constraints imposed on the latent space. The proposed guided mechanism provides such biases through two complementary pathways: supervised prediction that encourages each latent subspace to

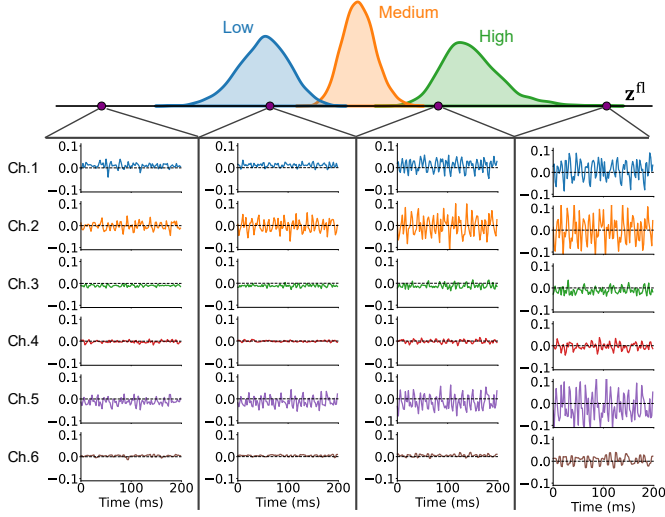


Figure 9: EMG signals generated by traversing the force-related latent dimension for Subject 9. The top panel shows the learned distribution along this dimension, and each signal below is generated by fixing the latent variable to the value indicated by the corresponding marker. A monotonic increase in signal amplitude is observed as the latent variable shifts from lower to higher values. The vertical and horizontal axes indicate normalized amplitude and time (ms), respectively.

encode its designated generative factor, and adversarial constraints that suppress the encoding of factor-related information in non-corresponding subspaces. Together, these mechanisms facilitate the learning of latent representations in which the underlying generative factors are effectively disentangled.

The proposed method does not enforce, nor does it claim, statistical independence of the latent subspaces; rather, it imposes inductive biases that align each subspace with a designated generative factor. This distinction matters because predictability and disentanglement are separate notions, as the CNN Multi-Predict baseline makes concrete: a representation optimized purely for factor prediction attains the best Informativeness yet remains strongly entangled (Table 1). This dissociation reflects that the DCI Disentanglement and Completeness scores, derived from random-forest feature-importance distributions, quantify the structural alignment between latent variables and factors rather than mere predictability. Given that force level, forearm orientation, and movement class are physiologically coupled, this factor-aligned structuring, rather than exact statistical independence, is what underlies the observed improvements in recognition robustness and controllable generation.

The effectiveness of this label-guided approach also suggests a potential direction for reducing the dependence on explicit labels. Since certain EMG signal characteristics are naturally correlated with specific factors (e.g., overall amplitude with force levels, relative channel distributions with forearm orientations), leveraging such prior knowledge may enable weakly supervised disentanglement and warrants further investigation.

It is also worth noting that, among the three factors, force level was the least well represented in the latent space, likely reflecting the subjectively assigned—and hence noisier—force

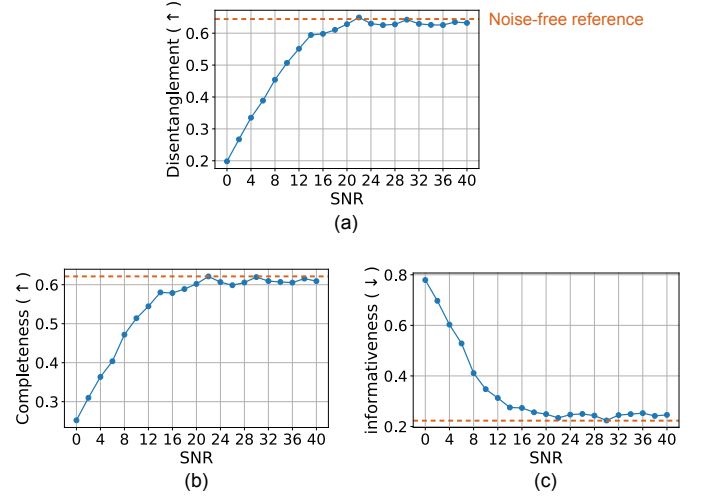


Figure 10: DCI scores of the proposed method on Dataset 1 as a function of the test-time signal-to-noise ratio (SNR): (a) Disentanglement, (b) Completeness, and (c) Informativeness, averaged across subjects. The model is trained on clean data and evaluated under additive Gaussian noise injected at test time. Dashed lines indicate the corresponding noise-free reference of the proposed method.

labels in this dataset (further discussed in Section 6.5).

6.2. Improved Performance on Downstream Recognition Tasks

Figs. 6 and 7 show that predicting generative factors from the disentangled representations achieves performance exceeding that obtained by direct multi-task prediction from the original signals, suggesting that the disentangled features encode factor-specific information that facilitates more robust downstream learning.

A key observation is that the improvement is consistent across *all* factors rather than confined to a single one. On Dataset 1, DRL-mt improves over CNN-mt for every factor, with the largest gain for movement class (+4.23 percentage points); the same trend holds on Dataset 2. We attribute this broad improvement to the intrinsically entangled nature of the three factors in EMG signals: because a change in one factor alters the signal features associated with the others, a representation that does not separate them suffers from mutual interference. By aligning each factor with a dedicated subspace and suppressing cross-subspace leakage, the proposed method reduces this inter-factor interference, which benefits the recognition of all factors simultaneously.

Force level shows a gain in 7 of 10 subjects on Dataset 1. Rather than a property specific to the force–movement relationship, this reflects supervision quality: in this dataset, force levels were assigned subjectively—subjects were merely instructed to progressively increase their effort—so the force labels are the noisiest among the three factors. This interpretation is supported by the factor-wise DCI analysis: as shown in Fig. 5(c), force level exhibits the highest Informativeness loss, indicating that it is the most difficult factor to predict from the learned representations. Consistent with this, the ablation (Table 2) shows that withholding force-level supervision reduces the overall Disentanglement score only slightly (0.644 →

0.488), the smallest reduction among the three factors, whereas withholding forearm-orientation supervision causes the largest reduction ($0.644 \rightarrow 0.314$); the model thus gains comparatively little from force supervision, in line with the noisy, subjectively assigned force labels. The Completeness pattern in Fig. 5(b) indicates that forearm orientation is the factor most amenable to compact encoding within its assigned subspace, whereas force level and movement class are more prone to being distributed across multiple latent dimensions, consistent with their greater susceptibility to entanglement.

It should be noted that the performance difference between DRL-mt and CNN-mt may partly reflect the regularization effect of the VAE’s generative objective in addition to disentanglement. However, the consistent correspondence between improved DCI scores and recognition gains across different supervision conditions (Table 2) suggests that the factor-specific structuring of the latent space plays a meaningful role in the observed improvements.

Although disentangled representations improve performance for most subjects, a degradation was observed for a few subjects. This suggests that strict factor separation is not universally optimal; for recognition tasks targeting a specific factor, certain task-relevant information associated with other factors may still be beneficial, and enforcing complete separation may limit the expressiveness of the learned representations. These results highlight a potential trade-off between representation disentanglement and task-specific performance, indicating that the appropriate degree of disentanglement may depend on both the target task and individual subject characteristics.

6.3. Generation Capacity of the Proposed Model

Beyond recognition, the disentangled latent space also provides a foundation for controllable EMG signal generation, which first requires faithful signal reconstruction. The analysis of reconstruction performance reveals that an MSE-based reconstruction loss alone is insufficient to preserve amplitude-related information in EMG signals. Since EMG signals are approximately zero-mean, minimizing pointwise errors tends to emphasize mean-level signal components while underconstraining amplitude and spectral characteristics. As quantified in Table 3, incorporating $\mathcal{L}^{\text{env}}$ and $\mathcal{L}^{\text{spec}}$ yields complementary improvements in envelope preservation and spectral fidelity, respectively, confirming that explicitly constraining these characteristics during training is effective.

Regarding signal generation, Fig. 9 shows that manipulating the force-related latent variable produces meaningful and continuous variations in the generated EMG signals, indicating that the model has learned a semantically structured latent space in which changes along specific dimensions lead to interpretable modifications of the generated signals.

Although the present analysis focuses on force-related generation as a representative case, the same principle applies to other factor-specific subspaces, enabling generation conditioned on arbitrary combinations of force level, forearm orientation, and movement class. Such controllability could facilitate data augmentation by generating EMG signals under diverse

factor conditions, potentially reducing the collection costs and alleviating the burden on subjects.

6.4. Practical Implications and Robustness

The proposed framework has several practical implications. The simultaneous decoding of force level, forearm orientation, and movement class supports more dexterous and context-aware prosthetic control, where grip force and posture must be adapted jointly to the manipulated object. In rehabilitation, factor-aligned representations enable the monitoring of effort and posture over time. Furthermore, controllable EMG synthesis via latent manipulation can reduce the calibration burden on users and provide data augmentation across factor conditions, alleviating the cost of exhaustive data collection.

Robustness to signal degradation is also relevant for these applications. As shown by the noise-robustness analysis (Fig. 10), the disentanglement quality of the proposed method degrades gracefully under additive Gaussian noise, suggesting that the factor-aligned latent structure is not overly sensitive to amplitude-level perturbations. This tolerance, however, is characterized only for additive Gaussian noise, which is a first-order approximation of real disturbances; fully realistic in-situ noise such as motion artifacts and electrode shift is beyond the present scope and is discussed below as a direction for future work.

6.5. Limitation

Despite the promising results, several limitations warrant further investigation. On the modeling side, the proposed method relies on supervised annotations of generative factors, requiring subjects to perform movements under all combinations of factor conditions, which can be time-consuming and costly in practice. Future work will explore incorporating inductive biases related to EMG signal characteristics through model design or data processing strategies to reduce this reliance. Relatedly, the reconstruction quality of EMG signals remains limited, which may restrict downstream applications such as data augmentation; improving reconstruction fidelity, particularly for fine-grained signal characteristics, is an important direction for future research.

On the evaluation side, the study is based on 10 subjects per dataset under a within-subject protocol. While this yields ten independent within-subject validations per dataset, a larger and more diverse subject pool would further strengthen the conclusions. Our framework is moreover designed for within-subject, within-session use. Generalization to unseen subjects constitutes a cross-subject domain-shift problem. This problem is orthogonal to the factor disentanglement addressed here, and it is the target of prior cross-subject disentanglement methods that separate the subject factor. Cross-session generalization—across electrode re-donning or separate recording days—is a further domain-shift axis, likewise orthogonal to the within-session factor disentanglement studied here. Unifying such subject- and session-level shifts with factor-level disentanglement is left for future work. Transferability across acquisition setups, in contrast, is already supported by the additional

Dataset 2 experiment, in which the framework was retrained per setup and remained effective. The robustness evaluation, in turn, considers only additive Gaussian noise; fully realistic in-situ disturbances such as motion artifacts and electrode shift remain to be investigated.

Finally, existing public EMG datasets present inherent limitations. In particular, force-level annotations are often ambiguously defined, and multi-factor EMG datasets with simultaneous annotations remain scarce, which constrains large-scale cross-dataset validation. The development of datasets with well-defined and consistent factor annotations will be essential for advancing research on disentangled representations in EMG signals.

7. Conclusion

In this study, we presented a VAE-based framework for learning disentangled representations from EMG signals by explicitly imposing inductive biases on the latent space. Instead of relying solely on implicit regularization, the proposed approach introduces guided mechanisms that align latent dimensions with semantically meaningful generative factors. Both qualitative visualizations and quantitative DCI evaluations demonstrate that such explicit guidance enables effective disentanglement of force level, forearm orientation, and movement class.

The learned disentangled representations consistently improve downstream multi-task recognition performance by reducing interference among generative factors, while also enabling controllable EMG signal generation through manipulation of factor-specific latent variables. These benefits recur on a second dataset recorded under a different acquisition setup and factor set, and the learned factor-aligned structure degrades only gracefully under additive noise, indicating that the improvements are not specific to a single benchmark. These results highlight guided disentanglement as a principled representation learning approach that unifies recognition and generation within a single framework for EMG signals.

Overall, this work demonstrates that incorporating explicit inductive biases into latent representation learning provides an effective pathway toward semantically meaningful and task-relevant EMG representations. Future work will focus on improving generation fidelity through more expressive decoder architectures and reconstruction objectives, and on reducing the reliance on explicit supervisory labels by leveraging prior knowledge of factor-related signal properties.

CRedit Author Statement

Masaki Matsuoka: Methodology, Software, Formal analysis, Writing - Original Draft, Writing - Review and Editing, Visualization. **Xinyue Niu:** Methodology, Writing - Review and Editing. **Akira Furui:** Conceptualization, Resources, Writing - Reviewing and Editing, Visualization, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number JP23K28128.

Data availability

Datasets are available at <https://rami-khushaba.com/biosignals-repository>. The code can be obtained upon request.

References

- Adewuyi, A.A., Hargrove, L.J., Kuiken, T.A., 2017. Resolving the effect of wrist position on myoelectric pattern recognition control. *Journal of NeuroEngineering and Rehabilitation* 14, 39. doi:10.1186/s12984-017-0246-x.
- Al-Timemy, A.H., Bugmann, G., Escudero, J., Outram, N., 2013. A preliminary investigation of the effect of force variation for myoelectric control of hand prosthesis, in: *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE. pp. 5758–5761. doi:10.1109/EMBC.2013.6610862.
- Al-Timemy, A.H., Khushaba, R.N., Bugmann, G., Escudero, J., 2015. Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24, 650–661. doi:10.1109/TNSRE.2015.2463174.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1798–1828. doi:10.1109/TPAMI.2013.50.
- Cesqui, B., Tropea, P., Micera, S., Krebs, H.I., 2013. EMG-based pattern recognition approach in post stroke robot-aided rehabilitation: a feasibility study. *Journal of NeuroEngineering and Rehabilitation* 10, 75. doi:10.1186/1743-0003-10-75.
- Chen, H., Zhang, Y., Zhang, Z., Fang, Y., Liu, H., Yao, C., 2017. Exploring the relation between EMG sampling frequency and hand motion recognition accuracy, in: *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE. pp. 1024–1028. doi:10.1109/SMC.2017.8122765.
- Chen, R.T., Li, X., Grosse, R.B., Duvenaud, D.K., 2018. Isolating sources of disentanglement in variational autoencoders. *Advances in Neural Information Processing Systems* 31.
- Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., Abbeel, P., 2016. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, in: *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2172–2180.
- Chen, Z., Qian, Y., Wang, Y., Fang, Y., 2022. Deep convolutional generative adversarial network-based EMG data enhancement for hand motion classification. *Frontiers in Bioengineering and Biotechnology* 10, 909653. doi:10.3389/fbioe.2022.909653.
- Comon, P., 1994. Independent component analysis, a new concept? *Signal Processing* 36, 287–314. doi:10.1016/0165-1684(94)90029-9.
- Dittadi, A., Träuble, F., Locatello, F., Wüthrich, M., Agrawal, V., Winther, O., Bauer, S., Schölkopf, B., 2021. On the transfer of disentangled representations in realistic settings, in: *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Dupont, E., 2018. Learning disentangled joint continuous and discrete representations. *Advances in Neural Information Processing Systems* 31.
- Eastwood, C., Williams, C.K., 2018. A framework for the quantitative evaluation of disentangled representations, in: *Proceedings of the International Conference on Learning Representations (ICLR)*.

- Fan, J., Jiang, X., Liu, X., Meng, L., Jia, F., Dai, C., 2024. Surface EMG feature disentanglement for robust pattern recognition. *Expert Systems with Applications* 237, 121224. doi:10.1016/j.eswa.2023.121224.
- Fang, B., Wang, C., Sun, F., Chen, Z., Shan, J., Liu, H., Ding, W., Liang, W., 2022. Simultaneous sEMG recognition of gestures and force levels for interaction with prosthetic hand. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30, 2426–2436. doi:10.1109/TNSRE.2022.3199819.
- Fougner, A., Stavadahl, Ø., Kyberd, P.J., Losier, Y.G., Parker, P.A., 2012. Control of upper limb prostheses: Terminology and proportional myoelectric control—A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20, 663–677. doi:10.1109/TNSRE.2012.2196711.
- Gouda, M.A., Hong, W., Jiang, D., Feng, N., Zhou, B., Li, Z., 2023. Synthesis of sEMG signals for hand gestures using a IDDCGAN. *Bioengineering* 10, 1353. doi:10.3390/bioengineering10121353.
- Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A., 2017. beta-VAE: Learning basic visual concepts with a constrained variational framework, in: *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Hua, S., Wang, C., Xie, Z., Wu, X., 2020. A force levels and gestures integrated multi-task strategy for neural decoding. *Complex & Intelligent Systems* 6, 469–478. doi:10.1007/s40747-019-0130-2.
- Khushaba, R.N., Al-Timemy, A., Kodagoda, S., Nazarpour, K., 2016. Combined influence of forearm orientation and muscular contraction on EMG pattern recognition. *Expert Systems with Applications* 61, 154–161. doi:10.1016/j.eswa.2016.05.031.
- Khushaba, R.N., Takruri, M., Miro, J.V., Kodagoda, S., 2014. Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features. *Neural Networks* 55, 42–58. doi:10.1016/j.neunet.2014.03.010.
- Kim, H., Mnih, A., 2018. Disentangling by factorising, in: *Proceedings of the International Conference on Machine Learning (ICML)*, PMLR. pp. 2649–2658.
- Kingma, D.P., Welling, M., 2014. Auto-encoding variational Bayes, in: *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Lee, S., Cho, S., Im, S., 2021. DRANet: Disentangling representation and adaptation networks for unsupervised cross-domain adaptation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15252–15261. doi:10.1109/CVPR46437.2021.01504.
- Leone, F., Gentile, C., Ciancio, A.L., Gruppioni, E., Davalli, A., Sacchetti, R., Guglielmelli, E., Zollo, L., 2019. Simultaneous sEMG classification of hand/wrist gestures and forces. *Frontiers in Neurobotics* 13, 42. doi:10.3389/fnbot.2019.00042.
- Leone, F., Mereu, F., Gentile, C., Cordella, F., Gruppioni, E., Zollo, L., 2023. Hierarchical strategy for sEMG classification of the hand/wrist gestures and forces of transradial amputees. *Frontiers in Neurobotics* 17, 1092006. doi:10.3389/fnbot.2023.1092006.
- Li, G., Li, Y., Yu, L., Geng, Y., 2011. Conditioning and sampling issues of EMG signals in motion recognition of multifunctional myoelectric prostheses. *Annals of Biomedical Engineering* 39, 1779–1787. doi:10.1007/s10439-011-0265-x.
- Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., Bachem, O., 2019. Challenging common assumptions in the unsupervised learning of disentangled representations, in: *Proceedings of the International Conference on Machine Learning (ICML)*, PMLR. pp. 4114–4124.
- Locatello, F., Poole, B., Rätsch, G., Schölkopf, B., Bachem, O., Tschannen, M., 2020. Weakly-supervised disentanglement without compromises, in: *Proceedings of the International Conference on Machine Learning (ICML)*, PMLR. pp. 6348–6359.
- Niu, X., Furui, A., 2025. Towards cross-subject emg pattern recognition via dual-branch adversarial feature disentanglement, in: *Proceedings of the 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE. pp. 1–6. doi:10.1109/EMBC58623.2025.11253475.
- Pearson, K., 1901. Liii. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, 559–572. doi:10.1080/14786440109462720.
- Pelaez Murcigo, L., Henrich, M.C., Spaich, E.G., Dosen, S., 2022. Reducing the number of EMG electrodes during online hand gesture classification with changing wrist positions. *Journal of NeuroEngineering and Rehabilitation* 19, 78. doi:10.1186/s12984-022-01065-9.
- Peng, L., Hou, Z., Chen, Y., Wang, W., Tong, L., Li, P., 2013. Combined use of sEMG and accelerometer in hand motion classification considering forearm rotation, in: *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE. pp. 4227–4230. doi:10.1109/EMBC.2013.6610475.
- Phinyomark, A., Scheme, E., 2018. EMG pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing* 2, 21. doi:10.3390/bdcc2030021.
- Scheme, E., Englehart, K., 2011. Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *Journal of Rehabilitation Research and Development* 48. doi:10.1682/JRRD.2010.09.0177.
- Scheme, E., Englehart, K., 2013. Training strategies for mitigating the effect of proportional control on classification in pattern recognition-based myoelectric control. *Journal of Prosthetics and Orthotics* 25, 76–83. doi:10.1097/JPO.0b013e318290ce35.
- Schölkopf, B., Locatello, F., Bauer, S., Ke, N.R., Kalchbrenner, N., Goyal, A., Bengio, Y., 2021. Toward causal representation learning. *Proceedings of the IEEE* 109, 612–634. doi:10.1109/JPR0C.2021.3058954.
- Tkach, D., Huang, H., Kuiken, T.A., 2010. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of NeuroEngineering and Rehabilitation* 7, 21. doi:10.1186/1743-0003-7-21.
- Tokmakov, P., Wang, Y.X., Hebert, M., 2019. Learning compositional representations for few-shot recognition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6372–6381. doi:10.1109/ICCV.2019.00646.
- Wang, X., Chen, H., Tang, S., Wu, Z., Zhu, W., 2024. Disentangled representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 9677–9696. doi:10.1109/TPAMI.2024.3436797.
- Xiong, B., Chen, W., Li, H., Niu, Y., Zeng, N., Gan, Z., Xu, Y., 2024. PatchEMG: Few-shot emg signal generation with diffusion models for data augmentation to improve classification performance. *IEEE Transactions on Instrumentation and Measurement* 73. doi:10.1109/TIM.2024.3398096.
- Xiong, D., Zhang, D., Zhao, X., Zhao, Y., 2021. Deep learning for EMG-based human-machine interaction: A review. *IEEE/CAA Journal of Automatica Sinica* 8, 512–533. doi:10.1109/JAS.2021.1003876.
- Xu, J., Le, H., Huang, M., Athar, S., Samaras, D., 2021. Variational feature disentanglement for fine-grained few-shot classification, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8812–8821. doi:10.1109/ICCV48922.2021.00870.
- Yan, M., Chen, Z., Li, J., Li, J., Pan, L., 2025. Simultaneous estimation of wrist joint angle and torque during isokinetic contraction based on HD-sEMG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 33, 3137–3145. doi:10.1109/TNSRE.2025.3546989.
- Yazawa, I., Furui, A., 2026. VAE-Based Synthetic EMG Generation with Mix-Consistency Loss for Recognizing Unseen Motion Combinations, in: *Proceedings of the IEEE/SICE International Symposium on System Integration (SII)*, IEEE. pp. 1285–1290. doi:10.1109/SII64115.2026.11404666.
- Zhang, Z., Ming, Y., Shen, Q., Wang, Y., Zhang, Y., 2025. An extended variational autoencoder for cross-subject electromyograph gesture recognition. *Biomedical Signal Processing and Control* 99, 106828. doi:10.1016/j.bspc.2024.106828.