

VAE-Based Synthetic EMG Generation with Mix-Consistency Loss for Recognizing Unseen Motion Combinations

Itsuki Yazawa¹ and Akira Furui¹

Abstract—Electromyogram (EMG)-based motion classification using machine learning has been widely employed in applications such as prosthesis control. While previous studies have explored generating synthetic patterns of combined motions to reduce training data requirements, these methods assume that combined motions can be represented as linear combinations of basic motions. However, this assumption often fails due to complex neuromuscular phenomena such as muscle co-contraction, resulting in low-fidelity synthetic signals and degraded classification performance. To address this limitation, we propose a novel method that learns to synthesize combined motion patterns in a structured latent space. Specifically, we employ a variational autoencoder (VAE) to encode EMG signals into a low-dimensional representation and introduce a mix-consistency loss that structures the latent space such that combined motions are embedded between their constituent basic motions. Synthetic patterns are then generated within this structured latent space and used to train classifiers for recognizing unseen combined motions. We validated our approach through upper-limb motion classification experiments with eight healthy participants. The results demonstrate that our method outperforms input-space synthesis approaches, achieving approximately 30% improvement in accuracy.

I. INTRODUCTION

Electromyogram (EMG) signals are electrical signals generated by muscle activity, measured non-invasively using surface electrodes. These signals contain rich information about human motor intentions and are commonly used for control interfaces, particularly prosthetic limbs [1]. Recent advances in deep learning have significantly improved the accuracy of EMG-based motion classification [2], [3].

However, EMG-based motion classification faces a fundamental challenge: the training data requirement scales with the number of target motions. This challenge becomes particularly acute for combined motions—simultaneous executions of multiple basic motions—which increase combinatorially. For instance, while four basic upper-limb motions require four data collection sessions, their pairwise combinations require an additional six sessions. This combinatorial growth in data collection leads to user fatigue and limits the practical deployment of multi-degree-of-freedom (DoF) control systems [4], [5].

To address this challenge, researchers have proposed representing combined motions as combinations of basic motion patterns, enabling multi-DoF control with limited training data [6]. To improve the robustness of such approaches,

recent methods generate synthetic combined motion data from basic motion measurements [7], [8]. Our previous work [8] demonstrated that convex combinations of basic motion patterns in the input signal space could improve classification accuracy for unseen combined motions. However, this approach relies on the assumption that combined motion patterns can be represented as linear combinations of basic motion patterns—an assumption that often fails to capture complex neuromuscular phenomena such as muscle co-contraction and motion-dependent modulation patterns.

In this paper, we propose a novel approach that learns to synthesize combined motion patterns in a structured latent space. We employ a variational autoencoder (VAE) [9] to encode EMG signals into a low-dimensional representation and introduce a mix-consistency loss that structures the latent space such that combined motions are embedded between their constituent basic motions. Synthetic combined motion data generated from this space are then used together with measured basic motion data to train a classifier, thereby enabling recognition of unseen combined motions without requiring their explicit measurement.

Our long-term vision is to develop a generalizable synthesis model that can be pre-trained on multiple subjects and transferred to new users, ultimately eliminating the need for combined motion data collection from individual users. As a first step toward this goal, we validate the feasibility of latent-space synthesis through intra-subject experiments.

II. PROPOSED METHOD

We address the problem of classifying EMG patterns $\mathbf{X} \in \mathbb{R}^{W \times D}$ measured from D electrodes with window size W . The target motion classes consist of M_b basic motions and M_c combined motions, where each combined motion represents a combination of K_m basic motion classes.

Our approach consists of two sequential training phases. In the VAE training phase, we collect EMG signals for all motion classes, including both basic and combined motions, to train a VAE capable of synthesizing user-specific EMG patterns for combined motions. Subsequently, in the classifier training phase, we measure EMG signals only for basic motion classes and input these into the pre-trained VAE to synthesize EMG patterns for the combined motions. The classifier is then trained using real EMG patterns for basic motions and synthetic EMG patterns for combined motions.

A. Training of VAE for Synthetic EMG Patterns

Fig. 1 illustrates an overview of the proposed method. We first train a VAE using dataset $\mathcal{D}_1 = \{(\mathbf{X}_i, \mathbf{y}_i)\}_{i=1}^{N_1}$, which

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

¹I. Yazawa and A. Furui are with Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-hiroshima, Japan (e-mail: {itsukiyazawa, akirafurui}@hiroshima-u.ac.jp)

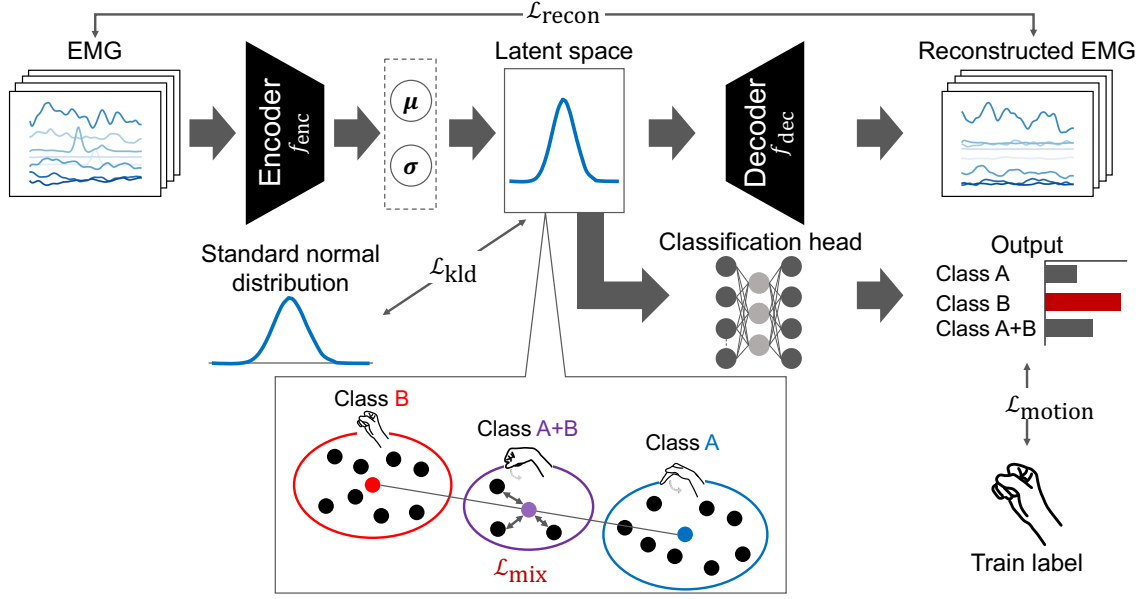


Fig. 1. Overview of the proposed VAE-based framework for synthetic EMG generation. The mix-consistency loss $\mathcal{L}_{\text{mix}}$ structures the latent space so that combined motions are positioned between their constituent basic motions, enabling realistic synthetic pattern generation.

contains EMG signals for both basic and combined motions. Here, $\mathbf{X}_i \in \mathbb{R}^{W \times D}$ represents the EMG pattern for the i -th sample, and $\mathbf{y}_i \in \{0, 1\}^{M_b + M_c}$ is a one-hot encoded label vector. The VAE encodes each EMG signal $\mathbf{X}_i$ into a Z -dimensional latent representation $\mathbf{z}_i \in \mathbb{R}^Z$.

The proposed VAE comprises three modules:

- **Encoder** extracts the mean $\mu_i \in \mathbb{R}^Z$ and standard deviation $\sigma_i \in \mathbb{R}^Z$ of the approximate posterior distribution $q(\mathbf{z}_i | \mathbf{X}_i)$ from input EMG signal $\mathbf{X}_i$. It consists of 1D convolutional layers followed by fully connected layers, incorporating batch normalization, max-pooling, and Leaky ReLU activations.
- **Decoder** samples a latent representation $\mathbf{z}_i \sim \mathcal{N}(\mu_i, \text{diag}(\sigma_i^2))$ and reconstructs EMG signal $\hat{\mathbf{X}}_i \in \mathbb{R}^{W \times D}$ from the latent vector $\mathbf{z}_i$. It comprises a fully connected layer followed by 1D transposed convolutional layers with progressive upsampling and nonlinear transformations using Leaky ReLU activations and a final hyperbolic tangent function.
- **Classification head** predicts class label $\hat{\mathbf{y}}_i \in \{0, 1\}^{M_b + M_c}$ from latent representation $\mathbf{z}_i$ using a multi-layer perceptron.

To train the VAE and structure the latent space, we minimize the following total loss function:

$$\mathcal{L}_{\text{vae}} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{kld}} + \gamma \mathcal{L}_{\text{motion}} + \mathcal{L}_{\text{mix}}, \quad (1)$$

where β and γ are hyperparameters controlling the respective loss terms.

The total loss consists of four components:

- **Reconstruction loss** encourages accurate reconstruction of input EMG signals $\mathbf{X}_i$ from the reconstructed outputs $\hat{\mathbf{X}}_i$. This loss comprises time-domain mean squared error (MSE) and frequency-domain MSE based on the

fast Fourier transform (FFT):

$$\mathcal{L}_{\text{recon}} = \mathcal{L}_{\text{time}} + \delta \mathcal{L}_{\text{freq}}, \quad (2)$$

$$\mathcal{L}_{\text{time}} = \frac{1}{N_1 D W} \sum_{i=1}^{N_1} \left\| \hat{\mathbf{X}}_i - \mathbf{X}_i \right\|_2^2, \quad (3)$$

$$\mathcal{L}_{\text{freq}} = \frac{1}{N_1 D} \sum_{i=1}^{N_1} \sum_{d=1}^D \left\| \text{FFT}(\mathbf{x}_{i,d}) - \text{FFT}(\hat{\mathbf{x}}_{i,d}) \right\|_2^2. \quad (4)$$

Here, $\mathbf{x}_{i,d} \in \mathbb{R}^W$ represents the one-dimensional time series corresponding to the d -th channel of the i -th sample, where $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,D}]^\top$. The FFT is computed for each channel over the entire time window, with weighting parameter δ controlling the contribution of the frequency-domain loss.

- **Kullback-Leibler divergence** encourages the posterior distribution to match a standard normal distribution:

$$\mathcal{L}_{\text{kld}} = \frac{1}{2N_1} \sum_{i=1}^{N_1} \sum_{j=1}^Z (\mu_{ij}^2 + \sigma_{ij}^2 - \log \sigma_{ij}^2 - 1). \quad (5)$$

- **Motion classification loss** promotes class-discriminative structure in the latent space:

$$\mathcal{L}_{\text{motion}} = \frac{1}{N_1} \sum_{i=1}^{N_1} \ell(\hat{\mathbf{y}}_i, \mathbf{y}_i), \quad (6)$$

where $\hat{\mathbf{y}}_i$ is the predicted class probability from the classification head given latent vector $\mathbf{z}_i$, and $\ell(\cdot, \cdot)$ denotes the cross-entropy loss.

- **Mix-consistency loss** encourages combined motions to be embedded near the centroid of their constituent basic

motions:

$$\mathcal{L}_{\text{mix}} = \frac{1}{|M_c|} \sum_{c \in M_c} \frac{1}{|I_c|} \sum_{i \in I_c} \left\| \mathbf{z}_i^{(c)} - \frac{1}{K_m} \sum_{b=1}^{K_m} \bar{\mathbf{z}}^{(b)} \right\|_2^2. \quad (7)$$

Here, I_c denotes the set of sample indices for combined motion class c , $\mathbf{z}_i^{(c)}$ is the latent vector of the i -th sample in class c , and $\bar{\mathbf{z}}^{(b)}$ is the mean latent vector of basic motion class b computed over all samples in $\mathcal{D}_1$.

B. Generation of Synthetic EMG and Classifier Training

Using the pre-trained VAE, we perform data generation and classifier training based on a newly measured dataset $\mathcal{D}_2 = \{(\mathbf{X}_i, \mathbf{y}_i)\}_{i=1}^{N_2}$, which contains only basic motion classes. Each label $\mathbf{y}_i \in \{0, 1\}^{M_b}$ is a one-hot vector corresponding to a basic motion.

To generate combined motion samples, we employ Mixup-like convex combination in the latent space of the VAE. Fig. 2 illustrates the synthetic data generation method. We randomly sample K_m different basic motion samples and linearly combine their latent representations to generate synthetic data for combined motion class m :

$$\boldsymbol{\mu}_i^{(k)}, \boldsymbol{\sigma}_i^{(k)} = f_{\text{enc}}(\mathbf{X}_i^{(k)}), \quad (8)$$

$$\mathbf{z}_i^{(k)} \sim \mathcal{N}(\boldsymbol{\mu}_i^{(k)}, \text{diag}(\boldsymbol{\sigma}_i^{(k)^2})), \quad (9)$$

$$\tilde{\mathbf{z}}_i^{(m)} = \sum_{k=1}^{K_m} \lambda_k \mathbf{z}_i^{(k)}, \quad (10)$$

$$\tilde{\mathbf{X}}_i^{(m)} = f_{\text{dec}}(\tilde{\mathbf{z}}_i^{(m)}). \quad (11)$$

Here, $\mathbf{X}_i^{(k)}$ is the EMG signal for the i -th sample from basic motion class k , and $f_{\text{enc}}(\cdot), f_{\text{dec}}(\cdot)$ denote the encoder and decoder of the VAE, respectively. The coefficients $\lambda_k \in [0, 1]$ represent mixing ratios that satisfy $\sum_{k=1}^{K_m} \lambda_k = 1$. These coefficients are sampled from a symmetric Dirichlet distribution to incorporate randomness while preserving the convex combination constraint:

$$p(\boldsymbol{\lambda} | \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K_m} \lambda_k^{\alpha-1}, \quad (12)$$

$$B(\boldsymbol{\alpha}) = \frac{\Gamma(\boldsymbol{\alpha})^{K_m}}{\Gamma(K_m \boldsymbol{\alpha})}, \quad (13)$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_{K_m}]^\top$ and $\boldsymbol{\alpha} > 0$ controls the shape of the Dirichlet distribution. The label $\hat{\mathbf{y}} \in \{0, 1\}^{M_c}$ for each generated sample $\tilde{\mathbf{X}}_i$ is set as a one-hot vector corresponding to the target combined motion class.

After generating the synthetic data, we combine the measured basic motion dataset $\mathcal{D}_2$ with the synthetic combined motion dataset $\tilde{\mathcal{D}}_2 = \{(\tilde{\mathbf{X}}_j, \tilde{\mathbf{y}}_j)\}_{j=1}^{\tilde{N}_2}$ to form the full training set $\mathcal{D} = \mathcal{D}_2 \cup \tilde{\mathcal{D}}_2$. The classifier is trained by minimizing the following loss:

$$\mathcal{L}_{\text{cls}} = \frac{1}{N_2} \sum_{i=1}^{N_2} \ell(f_{\text{cls}}(\mathbf{X}_i), \mathbf{y}_i) + \frac{1}{\tilde{N}_2} \sum_{j=1}^{\tilde{N}_2} \ell(f_{\text{cls}}(\tilde{\mathbf{X}}_j), \tilde{\mathbf{y}}_j), \quad (14)$$

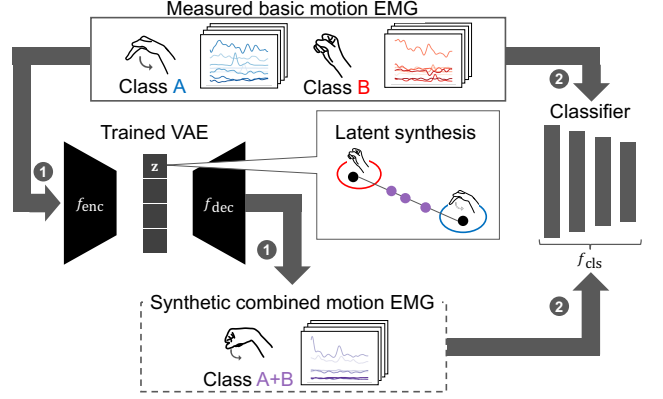


Fig. 2. Pipeline for synthetic combined motion generation and classifier training using measured basic motion EMG and pre-trained VAE latent synthesis.

where $f_{\text{cls}}(\cdot)$ represents the classifier, $\tilde{N}_2$ is the number of synthetic combined motion samples.

In summary, our approach first learns a structured latent space by training a VAE on both basic and combined motion classes. Combined motion samples are then generated in the latent space by convexly combining the latent vectors of K_m basic motions. Finally, a classifier is trained using both measured basic motions and synthetic combined motions, enabling recognition of previously unseen combined motions.

III. EXPERIMENTS

A. Experimental Setup

To evaluate the effectiveness of the proposed method, we conducted an upper-limb motion classification experiment using an EMG dataset from our previous work [8]. The dataset includes recordings from eight healthy adult participants (mean age: 23 ± 0.76 years). For each participant, eight surface electrodes were placed circumferentially around the right forearm near the elbow ($D = 8$). EMG signals were acquired using a wireless acquisition system (Trigno, Delsys Inc.) at a sampling rate of 2,000 Hz.

During the experiment, participants remained seated with their right elbow resting on a desk and performed a total of 18 distinct upper-limb motions. These included 6 basic motions: hand opening (S1), hand grasping (S2), wrist extension (S3), wrist flexion (S4), pronation (S5), and supination (S6) and 12 combined motions, each constructed by combining two basic motions ($K_m = 2$): opening and pronation (C1), grasping and pronation (C2), extension and pronation (C3), flexion and pronation (C4), opening and supination (C5), grasping and supination (C6), extension and supination (C7), flexion and supination (C8), opening and extension (C9), grasping and extension (C10), opening and flexion (C11), and grasping and flexion (C12). Thus, the total number of classes was $M_b = 6$ and $M_c = 12$.

Each motion was performed for 4 seconds continuously, and the task was repeated across 6 trials. Participants rested for 4 seconds between motions and 40 seconds between

trials while maintaining a relaxed posture. Prior to the study, participants provided written consent after being briefed about the research objectives. The study protocol received approval from the Ethics Committee at Hiroshima University (approval number: E-840).

B. Data Processing

The raw EMG signals were preprocessed as follows. First, a fourth-order band-stop filter (60–62 Hz) was applied to eliminate power line noise. Subsequently, full-wave rectification was performed, followed by smoothing using a second-order Butterworth low-pass filter with a cutoff frequency of 2.0 Hz. To remove transitional artifacts at motion onset, the initial 5% of each signal segment was discarded. The smoothed signals were then clipped between the 1st and 99th percentiles to suppress outliers, then normalized to the range $[-1, 1]$. The resulting EMG signals were used as the input patterns $\mathbf{X}$.

The experimental protocol employed a cross-validation approach using the 6 available trials per participant. For each fold, trials were allocated as follows:

- Two trials containing both basic and combined motions were used to train the VAE.
- Two different trials containing only basic motions were used to generate synthetic combined motion data and train the classifier.
- The remaining 2 trials containing both basic and combined motions served as test data.

This allocation ensures that the classifier never sees actual combined motion data during training, thereby validating the method’s ability to recognize unseen combined motions. Classification accuracy was computed by averaging results across all possible trial combinations.

C. Evaluation Methods

1) *Training configuration*: We used a convolutional neural network (CNN) as the classifier. Both the VAE and CNN classifier were optimized using Adam with a batch size of 64 and a learning rate of 1.0×10^{-4} . Training proceeded for 200 epochs for the VAE and 10 epochs for the CNN classifier. Detailed model architectures are provided in Table. I. In our model, the parameters α of all Leaky ReLU functions were set to 0.2. Moreover, the size of latent vector was set to 6 ($Z = 6$).

The loss function hyperparameters were configured as follows. The β parameter was scaled by the latent-to-input dimensionality ratio to maintain consistent loss scales across different dimensions [10]:

$$\beta = \beta_0 \cdot \frac{Z}{WD}, \quad (15)$$

where β_0 represents the base value that was gradually increased during training to stabilize VAE optimization, with a maximum value of 0.3. The γ parameter was fixed at 2 throughout training, while δ was set to 0 for the first 100 epochs and changed to 1 for the remaining training. For synthetic data generation, the α parameter of the Dirichlet distribution was set to 50.

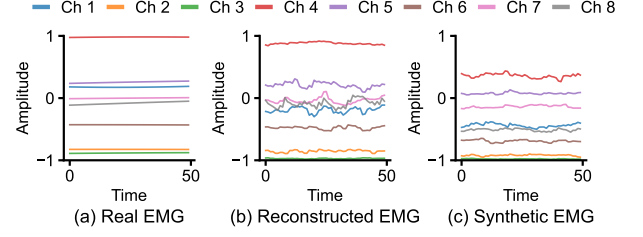


Fig. 3. Example of EMG signals for combined motion C2, participant 1: (a) measured, (b) VAE-reconstructed, and (c) synthetic patterns showing comparable amplitude trends across 8 channels.

2) *Baseline comparisons*: We evaluated the proposed method against three baseline approaches:

- **Fully-supervised**: Trained on measured data from 2 trials of all 18 motion classes (basic and combined). This represents the ideal scenario with complete supervision.
- **Basic-only [6]**: Trained exclusively on measured data from 2 trials of the 6 basic motion classes, providing a baseline for methods without combined motion data.
- **Mixup-based synthesis [8]**: Applied data augmentation method by mixing input signals directly based Mixup concept [11] without using the VAE, as proposed in our previous work. This represents a simpler synthesis approach for comparison.

All baseline methods used identical training configurations to ensure fair comparison.

IV. RESULTS

A. Evaluation of Synthetic EMG Pattern Generation

Fig. 3 presents a comparison of EMG signals for a representative combined motion class, showing (a) the original measured EMG signal, (b) its VAE reconstructed, and (c) the reconstruction generated from synthetic data. The horizontal and vertical axes represent time and signal amplitude, respectively.

The comparison between Fig. 3(a) and (b) demonstrates that the VAE successfully captures the overall amplitude trends of the original signal. However, minor reconstruction errors are evident in localized amplitude fluctuations and noise characteristics. The comparison between Fig. 3(b) and (c) reveals that synthetic combined data produce reconstructions similar to those of actual measured data, indicating that the VAE effectively captures realistic and class-consistent features of combined motions.

B. Structure of Latent Space

Fig. 4 shows a comparison between the measured input EMG signals and their corresponding latent representations obtained from the VAE, visualized in two dimensions using principal component analysis.

While all motion classes are included in Fig. 4, we highlight the combined class “Hand Opening + Pronation” (C1), which is composed of two basic motions: “Opening” (S1) and “Pronation” (S5). In Fig. 4(a), the input EMG

TABLE I
MODEL ARCHITECTURE CONFIGURATIONS

Component	Layer Configuration	Output Channels	Kernel/Stride	Activation	Additional Features
VAE	<i>Encoder:</i>				
	3 × 1D Conv	32, 64, 128	3/1, 3/1, 4/2	Leaky ReLU	BatchNorm; MaxPool ^a
	Fully connected	128	—	Leaky ReLU	—
	→ μ branch	6	—	—	—
	→ $\log \sigma^2$ branch	6	—	—	—
	<i>Decoder:</i>				
	Fully connected	768	—	Leaky ReLU	—
CNN Classifier	3 × 1D TransConv	64, 32, 8	4/2, 3/2, 2/2	Leaky ReLU; Tanh ^c	BatchNorm ^a
	<i>Classification head:</i>				
	3 × Fully connected	64, 32, 18	—	ReLU	—
CNN Classifier	4 × 1D Conv	16, 32, 64, 128	3/1	PReLU	BatchNorm; MaxPool; GAP
	2 × Fully connected	64, 18	—	—	BatchNorm ^b

^aApplied to first two layers only. ^bApplied to first FC layer only. ^cLeaky ReLU for first two layers, Tanh for final layer. The abbreviations Conv, TransConv, BatchNorm, MaxPool, and GAP denote convolutional, transposed convolutional, batch normalization, max pooling, global average pooling, respectively.

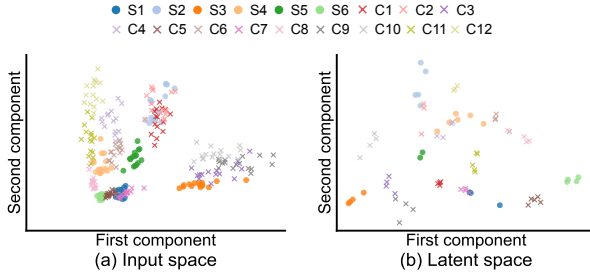


Fig. 4. Principal component analysis of EMG patterns in (a) input space and (b) VAE latent space, showing structured embedding of combined motions between basic motions (Participant 1).

signals corresponding to the combined motion form a distribution that is clearly distinct from those of the individual basic motions. This suggests that, in the original signal space, the combined motion exhibits a unique pattern not easily interpreted as a simple combination of the constituent basic motions. In contrast, Fig. 4(b) shows the latent space, where the combined motion is distributed near the midpoint between the two corresponding basic motion clusters. This indicates that the latent space has been structured such that combined motions are positioned in a meaningful location—approximately interpolated between the latent representations of the basic motions involved.

C. Classification Performance

Table II presents the classification accuracy results for all evaluated method. The fully-supervised method, trained on actual EMG patterns for all motion classes, achieved a high average classification accuracy exceeding 80%. This represents an upper-bound performance under ideal conditions where complete training data are available. In contrast, the Mixup method, which generates synthetic combined motion data in the input space, achieved lower accuracy of approximately 40%. Our proposed method, utilizing a structured latent space for synthetic data generation, substantially improved classification accuracy to around 75%, demonstrating clear superiority over the input-space synthesis approach.

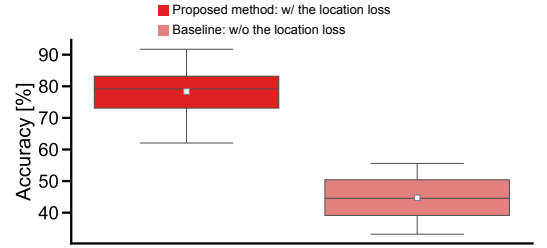


Fig. 5. Ablation study of mix-consistency loss on classification accuracy.

Fig. 5 presents an ablation study evaluating the contribution of the mix-consistency loss $\mathcal{L}_{\text{mix}}$. The medial line and white square denote the median and mean values across participants, respectively. The baseline model without this loss component achieved approximately 40% accuracy, similar to the Mixup method. Incorporating the mix-consistency loss resulted in notable accuracy improvement, confirming the effectiveness of structuring the latent space for combined motion synthesis.

V. DISCUSSION

This study aimed to enable multi-DoF motion classification without explicit training data for combined motions by learning to generate such data synthetically in advance. The experimental results demonstrate the effectiveness of our VAE-based approach over conventional linear synthesis methods.

The conventional Mixup method, which linearly combines basic motion data in the input space, resulted in significant degradation in classification accuracy (Table II). This performance gap can be attributed to the fundamental limitation that EMG patterns of combined motions do not necessarily exhibit linear relationships with their constituent basic motions. The complex neuromuscular interactions, including co-contraction patterns and synergistic activations, create nonlinear dependencies that cannot be captured through simple linear interpolation in the input space.

TABLE II
CLASSIFICATION ACCURACY (MEAN $\pm$ STANDARD DEVIATION) FOR PROPOSED METHOD AND BASELINES

Method	Training data configuration		Accuracy (%)		
	Basic	Combined	Basic	Combined	Overall
Fully-supervised	Real	Real	83.15 $\pm$ 6.91	86.47 $\pm$ 5.44	85.36 $\pm$ 5.10
Basic-only	Real	—	90.65 $\pm$ 7.64	16.42 $\pm$ 3.52	41.17 $\pm$ 4.34
Mixup-based synthesis [8]	Real	Synthetic (input space)	86.97 $\pm$ 8.94	23.21 $\pm$ 5.85	44.46 $\pm$ 6.38
Ours	Real	Synthetic (latent space)	82.43 $\pm$ 10.24	76.29 $\pm$ 9.40	78.34 $\pm$ 9.14

Bold values indicate the best performance among methods that do not use real combined motion data for training.

Our proposed method addresses this limitation by performing synthesis in the structured latent space of a VAE. The substantial improvement in classification accuracy (approximately 30% over the Mixup method) demonstrates that the learned latent representation better captures the underlying motion patterns. As illustrated in Fig. 3, the synthetic EMG patterns generated in the latent space closely resemble actual measured patterns, enabling the classifier to effectively learn discriminative features for combined motions.

The latent space visualization in Fig. 4 provides crucial insights into why our method succeeds. The structured embedding, where combined motions are positioned near the intermediate regions between their constituent basic motions, enables reliable interpolation for synthetic data generation. This spatial organization results from the explicit regularization imposed by our mix-consistency loss $\mathcal{L}_{\text{mix}}$. The ablation study in Fig. 5 confirms the critical role of this loss component, as the baseline model without $\mathcal{L}_{\text{mix}}$ achieved only 40% accuracy, similar to the Mixup method. This indicates that the VAE alone is insufficient to create the desired latent space structure, and the mix-consistency loss is essential for ensuring that learned representations respect the compositional nature of combined motions.

Despite these improvements, our method exhibits an approximately 10% accuracy gap compared to the fully-supervised baseline (Table II). This gap primarily stems from imperfect signal reconstruction, as evidenced by noise and amplitude variations in the synthetic patterns (Fig. 3), and the current requirement for some combined motion data during VAE training. Future work should focus on cross-subject transfer learning, where VAE models pre-trained on multiple subjects could enable synthetic data generation for new users with only basic motion data, thereby eliminating the need for individual combined motion collection and enhancing practical applicability.

VI. CONCLUSION

This study proposed a VAE-based method for recognizing unseen combined motions from EMG signals through synthetic data generation in a structured latent space. By introducing the mix-consistency loss, our approach ensures that combined motions are embedded between their constituent basic motions, enabling realistic synthetic pattern generation for classifier training.

Experimental validation with eight participants demonstrated substantial improvements in combined motion clas-

sification accuracy, achieving 78.34% compared to 44.46% for conventional input-space synthesis methods. The mix-consistency loss proved essential for structuring the latent space, with ablation studies confirming its critical role in performance enhancement.

While this proof-of-concept demonstrates the feasibility of latent space synthesis, the current requirement for same-subject combined motion data during VAE training limits immediate burden reduction benefits. Future work should focus on cross-subject transfer learning to enable practical deployment with minimal per-user calibration.

REFERENCES

- [1] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, and O. C. Aszmann, "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, pp. 797–809, 2014.
- [2] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, pp. 760–771, 2019.
- [3] T. Triwiyanto, I. P. A. Pawana, and M. H. Purnomo, "An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1678–1688, 2020.
- [4] K. Nagata and K. Magatani, "Basic study on combined motion estimation using multichannel surface EMG signals," in *Proc. 33rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*. IEEE, 2011, pp. 7865–7868.
- [5] A. E. Olsson, P. Sager, E. Andersson, A. Björkman, N. Malešević, and C. Antfolk, "Extraction of multi-labelled movement information from the raw HD-sEMG image with time-domain depth," *Sci. Rep.*, vol. 9, no. 1, p. 7244, 2019.
- [6] K. Shima and T. Tsuji, "Classification of combined motions in human joints through learning of individual motions based on muscle synergy theory," in *Proc. IEEE/SICE Int. Symp. Sys. Intg. (SII)*. IEEE, 2010, pp. 323–328.
- [7] A. E. Olsson, N. Malešević, A. Björkman, and C. Antfolk, "Can deeep synthesis of EMG overcome the geometric growth of training data required to recognize multiarticulate motions?" in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*. IEEE, 2021, pp. 6380–6383.
- [8] I. Yazawa, S. Yoneda, and A. Furui, "Recognition of unseen combined motions via convex combination-based EMG pattern synthesis for myoelectric control," in *Proc. 47th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*. IEEE, 2025.
- [9] D. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2013.
- [10] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, " β -VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. Int. Conf. Learn. Rep.*, 2017.
- [11] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Rep. (ICLR)*, 2018.