Novelty Assessment Report

Paper: EMBridge: Enhancing Gesture Generalization from EMG Signals Through Cross-modal Representation Learning

PDF URL: https://openreview.net/pdf?id=LqrWNdceum

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

Hand gesture classification using high-quality structured data such as videos, images, and hand skeletons is a well-explored problem in computer vision. Alternatively, leveraging low-power, cost-effective bio-signals, e.g. surface electromyography (sEMG), allows for continuous gesture prediction on wearable devices. In this work, we aim to enhance EMG representation quality by aligning it with embeddings obtained from structured, high-quality modalities that provide richer semantic guidance, ultimately enabling zero-shot gesture generalization. Specifically, we propose EMBridge, a cross-modal representation learning framework that bridges the modality gap between EMG and pose. EMBridge learns high-quality EMG representations by introducing a Querying Transformer (Q-Former), a masked pose reconstruction loss, and a community-aware soft contrastive learning objective that aligns the relative geometry of the embedding spaces. We evaluate EMBridge on both in-distribution and unseen gesture classification tasks and demonstrate consistent performance gains over all baselines. To the best of our knowledge, EMBridge is the first cross-modal representation learning framework to achieve zero-shot gesture classification from wearable EMG signals, showing potential toward real-world gesture recognition on wearable devices.

Core Task Landscape

This paper addresses: **Zero-Shot Gesture Classification from Wearable EMG Signals**

A total of **40 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Cross-User and Cross-Session Domain Adaptation**
- **Cross-Modal Representation Learning and Foundation Models**
- **Compositional and Disentangled Latent Representations**
- **Real-Time Intent Detection and Segmentation**
- **Supervised Classification Architectures**
- **Unsupervised Feature Learning and Clustering**
- **Wrist-Based and Compact Wearable Systems**
- **Robustness to Electrode Shift and Posture Variation**
- **Anomaly Detection and Out-of-Distribution Rejection**
- **Zero-Shot Learning with Semantic Attributes**
- ... and 2 more categories

Complete Taxonomy Tree

- Zero-Shot Gesture Classification from Wearable EMG Signals Survey Taxonomy
- Cross-User and Cross-Session Domain Adaptation
  - Unsupervised Domain Adaptation for Cross-User Transfer (7 papers)
  - [1] EMG-based Multi-User Hand Gesture Classification via Unsupervised Transfer Learning Using Unknown Calibration Gestures (Haojie Shi, 2024) View paper
  - [3] Linear non-conservative unsupervised domain adaptation for cross-subject EMG gesture recognition (Martin Colot, 2026) View paper
  - [7] Optimization of inter-subject sEMG-based hand gesture recognition tasks using unsupervised domain adaptation techniques (Zihao Wang, 2024) View paper
  - [14] Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition (Ulysse CÃ´tÃ©-Allard, 2020) View paper
  - [16] Unsupervised domain adaptation for gesture identification against electrode shift (Patrick P. K. Chan, 2022) View paper
  - [17] Emgsense: A low-effort self-supervised domain adaptation framework for emg sensing (Di Duan, 2023) View paper
  - [30] EMG-UP: Unsupervised Personalization in Cross-User EMG Gesture Recognition (Wang Nana, 2025) View paper
  - Domain Generalization Without Target Data (1 papers)
  - [29] Cross-subject EMG hand gesture recognition based on dynamic domain generalization (Yalan Ye, 2023) View paper
  - Supervised and Semi-Supervised Calibration (3 papers)
  - [32] Toward Highly Flexible Inter-User Calibration of Myoelectric Control Models With User-Defined Hand Gestures (Yangyang YUAN, 2024) View paper
  - [38] Feasibility of Using Wearable EMG Armbands combined with Unsupervised Transfer Learning for Seamless Myoelectric Control (M. H. Sohn, 2022) View paper
  - [39] Review of: "Feasibility of Using Wearable EMG Armbands combined with Unsupervised Transfer Learning for Seamless Myoelectric Control" (Yanjuan Geng, 2022) View paper

- Cross-Modal Representation Learning and Foundation Models
  - Cross-Modal Alignment with Pose or Video ★ (2 papers)
  - [0] EMBridge: Enhancing Gesture Generalization from EMG Signals Through Cross-modal Representation Learning (Anon et al., 2026) View paper
  - [20] CPEP: Contrastive Pose-EMG Pre-training Enhances Gesture Generalization on EMG Signals (Cui Wen-hui, 2025) View paper
  - Foundation Models for EMG (3 papers)
  - [4] An EMG foundation model for neural decoding (Mihailidis, 2025) View paper
  - [11] Big data in myoelectric control: large multi-user models enable robust zero-shot EMG-based discrete gesture recognition (Ethan Eddy, 2024) View paper
  - [40] Foundation Models for EMG Human-Machine Interfaces (Fasulo, n.d.) View paper
- Compositional and Disentangled Latent Representations (1 papers)
  - [23] Disentangled EMG Representations Enable Zero-Shot Prediction of Multi-Finger Gestures via Linear and Compositional Latent Spaces (M Pizzi, 2025) View paper
- Real-Time Intent Detection and Segmentation (2 papers)
  - [22] ReactEMG: Stable, Low-Latency Intent Detection from sEMG via Masked Modeling (Wang Runsheng, 2025) View paper
- Supervised Classification Architectures
  - Convolutional and Recurrent Neural Networks (2 papers)
  - [2] EMG-based online classification of gestures with recurrent neural networks (Miguel SimÃ£o, 2019) View paper
  - [31] Convolution Neural Network for EMG-Based Finger Gesture Classification for Novel and Trained Gestures (Erik Lloyd, 2019) View paper
  - Transformer-Based Architectures (1 papers)
  - [9] ViT-HGR: Vision Transformer-based Hand Gesture Recognition from High Density Surface EMG Signals (Mansooreh Montazerin, 2022) View paper
  - Shallow and Hybrid Classifiers (3 papers)
  - [5] Surface EMG hand gesture recognition system based on PCA and GRNN (Jinxian Qi, 2020) View paper
  - [8] Classification of electromyographic hand gesture signals using modified fuzzy C-means clustering and two-step machine learning approach (Guangyu Jia, 2020) View paper
  - [10] Spectral Collaborative Representation based Classification for hand gestures recognition on electromyography signals (Boyali, 2016) View paper
  - Brain-Inspired and Neuromorphic Classifiers (1 papers)
  - [13] An EMG Gesture Recognition System with Flexible High-Density Sensors and Brain-Inspired High-Dimensional Classifier (A Moin, 2018) View paper
- Unsupervised Feature Learning and Clustering (2 papers)
  - [6] Hand gesture recognition using unsupervised learning (Birur, 2018) View paper
  - [28] Classifications of Dynamic EMG in Hand Gesture and Unsupervised Grasp Motion Segmentation (Mo Han, 2021) View paper
- Wrist-Based and Compact Wearable Systems (3 papers)
  - [15] A simplified wearable device powered by a generative EMG network for hand-gesture recognition and gait prediction (KK Kim, 2025) View paper
  - [18] A comprehensive analysis of wrist electromyography for hand gesture recognition: Advancing from forearm to wrist wearable devices (Botros, 2025) View paper
  - [21] From zero- to few-shot: deep temporal learning of wrist EMG enables scalable cross-user gesture recognition (Fady S. Botros, 2025) View paper
- Robustness to Electrode Shift and Posture Variation (1 papers)
  - [36] Posture-invariant myoelectric control with self-calibrating random forests (Jiang Xinyu, 2024) View paper
- Anomaly Detection and Out-of-Distribution Rejection (2 papers)
  - [12] Distraction Detection and Intention Recognition for Gesture-Controlled Unmanned Aerial Vehicle Operation (Datta, 2022) View paper
  - [37] Rejecting Unknown Gestures Based on Surface-Electromyography Using Variational Autoencoder. (Qingfeng Dai, 2024) View paper
- Zero-Shot Learning with Semantic Attributes (1 papers)
  - [26] Attributes' Importance for Zero-Shot Pose-Classification Based on Wearable Sensors (Hiroki Ohashi, 2022) View paper
- Methodological Reviews and Comparative Studies (5 papers)
  - [19] Deep Feature Learning from Electromyographic Signals for Gesture Recognition Systems (Wenjuan Zhong, 2025) View paper
  - [25] Optimization of EMG-based hand gesture recognition: Supervised vs. unsupervised data preprocessing on healthy subjects and transradial amputees (F. Riillo, 2014) View paper
  - [27] Artificial intelligence techniques for biosensor data analysis (Vishal Eswaran, 2026) View paper
  - [34] Unsupervised Learning of Wearable Sensor Data for Assessing Upper Limb Function (Janani S, 2024) View paper
  - [35] Supervised and unsupervised learning in processing myographic patterns (M. Shamsin, 2018) View paper
- Application-Specific Gesture Recognition (1 papers)
  - [24] Finger Motion Analysis for Interactive Applications Using Wearable and Wireless IoT Devices (Liu, 2024) View paper

Narrative

Core task: zero-shot gesture classification from wearable EMG signals. The field addresses the challenge of recognizing hand and finger gestures from electromyographic recordings without requiring labeled training data for every gesture class. The taxonomy reveals a rich landscape organized around several complementary strategies. Cross-user and cross-session domain adaptation methods tackle the variability that arises when models trained on one individual or recording session must generalize to new users or conditions. Cross-modal representation learning and foundation models leverage auxiliary modalities—such as video, pose, or kinematic data—to build shared embeddings that enable zero-shot transfer. Compositional and disentangled latent representations aim to factorize gesture signals into interpretable components, while real-time intent detection and segmentation focus on continuous recognition pipelines. Supervised classification architectures and unsupervised feature learning provide the backbone techniques, and specialized branches address wrist-based compact systems, robustness to electrode shift and posture variation, anomaly detection for out-of-distribution rejection, and zero-shot learning with semantic attributes. Methodological reviews and application-specific recognition round out the taxonomy, reflecting both foundational research and deployment concerns.

A particularly active line of work explores cross-modal alignment, where methods like EMBridge[0] and CPEP[20] align EMG features with visual or kinematic representations to enable zero-shot generalization. These approaches contrast with purely signal-driven strategies such as domain adaptation (Linear Domain Adaptation[3]) or unsupervised clustering (Fuzzy Clustering EMG[8]), which do not rely on auxiliary modalities. Foundation models (EMG Foundation Model[4]) represent an emerging direction that seeks to pretrain large-scale representations across diverse datasets, bridging multiple branches of the taxonomy. EMBridge[0] sits squarely within the cross-modal alignment cluster, emphasizing the use of pose or video as a supervisory signal to learn transferable EMG embeddings. Compared to CPEP[20], which also pursues cross-modal learning, EMBridge[0] may differ in the specific alignment objective or the choice of auxiliary modality, while both share the goal of enabling recognition of unseen gestures through learned correspondences rather than direct labeled supervision.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. CPEP: Contrastive Pose-EMG Pre-training Enhances Gesture Generalization on EMG Signals

Authors: Cui Wen-hui, Wenhui Cui, Pouransari, Hadi, Chris Sandino, et al. (19 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Hand gesture classification using high-quality structured data such as videos, images, and hand skeletons is a well-explored problem in computer vision. Leveraging low-power, cost-effective biosignals, e.g. surface electromyography (sEMG), allows for continuous gesture prediction on wearables. In this paper, we demonstrate that learning representations from weak-modality data that are aligned with those from structured, high-quality data can improve representation quality and enables zero-shot c...

Relationship Analysis

Both papers belong to the Cross-Modal Alignment with Pose or Video category, using contrastive learning to align EMG signals with pose embeddings for zero-shot gesture classification. They share the same core approach of freezing a pre-trained pose encoder and training an EMG encoder via contrastive objectives on the emg2pose dataset. The key differences are that EMBridge introduces a Q-Former with learnable queries, masked pose reconstruction loss, and community-aware soft contrastive learning (CASCLe) to capture neighborhood structures, while CPEP uses a simpler projection head with standard InfoNCE loss and focuses on symmetric contrastive alignment without the additional reconstruction and soft-target objectives.

Contributions Analysis

Overall novelty summary. The paper proposes EMBridge, a cross-modal representation learning framework that aligns EMG signals with pose embeddings to enable zero-shot gesture classification. It resides in the 'Cross-Modal Alignment with Pose or Video' leaf, which contains only one other sibling paper (CPEP). This leaf sits within the broader 'Cross-Modal Representation Learning and Foundation Models' branch, indicating a relatively sparse but emerging research direction. The taxonomy shows that cross-modal alignment represents one of several complementary strategies in the field, alongside domain adaptation, compositional representations, and foundation models.

The taxonomy reveals neighboring research directions that pursue zero-shot generalization through different mechanisms. The sibling branch 'Foundation Models for EMG' contains three papers building large-scale pre-trained models without cross-modal supervision. Adjacent branches include 'Cross-User and Cross-Session Domain Adaptation' (ten papers across three leaves) addressing user variability through transfer learning, and 'Zero-Shot Learning with Semantic Attributes' (one paper) using class descriptions rather than visual modalities. EMBridge diverges from these by leveraging structured pose data as supervisory signal, rather than relying solely on EMG-domain techniques or semantic metadata.

Among the three contributions analyzed, the framework and contrastive objective appear relatively novel within the limited search scope. The 'EMBridge framework' examined ten candidates with zero refutations, and the 'CASCLe objective' similarly found no overlapping prior work among ten candidates. However, the 'first zero-shot gesture classification' claim examined nine candidates and identified three potentially refutable papers, suggesting that zero-shot EMG gesture recognition has been explored previously. The analysis is based on twenty-nine total candidates from semantic search, not an exhaustive literature review, so these findings reflect the most semantically similar work rather than complete field coverage.

The limited search scope (twenty-nine candidates) and sparse taxonomy leaf (two papers total) suggest this work occupies a relatively unexplored intersection of cross-modal learning and EMG-based gesture recognition. The refutation of the 'first zero-shot' claim indicates that while the specific framework may be novel, the broader goal has precedent. The analysis cannot determine whether EMBridge's technical approach—combining Q-Former, masked reconstruction, and community-aware contrastive learning—represents a significant departure from the one identified sibling paper or the three refuting candidates.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: EMBridge cross-modal representation learning framework

Description: The authors introduce EMBridge, a framework that enhances EMG representation quality by aligning it with pose embeddings through three components: a Querying Transformer (Q-Former), a masked pose reconstruction loss, and a community-aware soft contrastive learning objective that aligns the relative geometry of embedding spaces.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Alignment-enhanced interactive fusion model for complete and incomplete multimodal hand gesture recognition

URL: View paper

Brief Assessment

Alignment-Enhanced Fusion[57] focuses on fusing sEMG and accelerometer signals for hand gesture recognition using hierarchical fusion and supervised contrastive learning, not on aligning EMG with pose embeddings through transformers for zero-shot generalization.

---

2. sEMG-vision Tra: A Gesture Recognition Method Based on Surface EMG Signal-Vision Fusion

URL: View paper

Brief Assessment

sEMG-Vision Fusion[65] uses simple feature concatenation and transformer encoding for supervised gesture classification, whereas EMBridge introduces specialized cross-modal alignment components (Q-Former, masked pose reconstruction, community-aware soft contrastive learning) for zero-shot generalization through representation learning.

---

3. Smooth Multiscale Convolutional Attention Transformer Network for Continuous Motion Estimation of Hand Knuckle Angle using Surface EMG Signals

URL: View paper

Brief Assessment

Smooth Multiscale Transformer[61] focuses on continuous hand motion estimation from EMG signals using convolutional attention and transformer architectures for regression tasks, not cross-modal representation learning between EMG and pose embeddings for zero-shot gesture classification.

4. PREDICTING CONTINUOUS HAND POSE FROM WEARABLE EMG SENSOR DATA USING TRANSFORMER-BASED DEEP-LEARNING MODELS

URL: View paper

Brief Assessment

Continuous Hand Pose[63] focuses on direct regression of hand pose from EMG using transformer architectures, not cross-modal representation learning or alignment between EMG and pose embeddings through contrastive learning objectives.

5. Efficient transformer for sEMG-based hand pose estimation using inter-channel relationship

URL: View paper

Brief Assessment

Inter-Channel Transformer[62] focuses on exploiting inter-channel relationships in sEMG for hand pose estimation using transformers, not on cross-modal representation learning that aligns EMG with pose embeddings through Q-Former, masked reconstruction, or contrastive objectives.

6. A multimodal transformer framework with biomechanical constraints for injury prediction and human motion analysis

URL: View paper

Brief Assessment

Biomechanical Transformer Injury[56] focuses on multimodal sensor integration (IMUs, EMG, plantar pressure) for injury prediction with biomechanical constraints, not on cross-modal representation learning that aligns EMG with pose embeddings using transformers and contrastive objectives.

7. Multimodal pose estimation and simulation modelling for real-time human motion analysis

URL: View paper

Brief Assessment

Multimodal Pose Simulation[60] focuses on fusing visual and inertial data for pose estimation and simulation modeling, not on aligning EMG signals with pose embeddings using transformers for gesture recognition.

8. From Wrist to Finger: Hand Pose Tracking Using Ring-Watch Wearables

URL: View paper

Brief Assessment

Ring-Watch Wearables[58] focuses on sensor fusion between IMU rings and EMG wristbands for hand pose tracking, not cross-modal representation learning that aligns EMG with pose embeddings through transformers, Q-Former, or contrastive objectives.

9. FEASIBILITY OF AN INTEGRATED MULTI-MODEL APPROACH FOR DYNAMIC MUSCULOSKELETAL DISORDER RISK ASSESSMENT

URL: View paper

Brief Assessment

Integrated Multi-Model Risk[64] focuses on musculoskeletal disorder risk assessment using multi-model integration for biomechanical analysis, not on cross-modal representation learning between EMG signals and pose embeddings using transformers.

10. Multimodal Transformer Models for Human Action Classification

URL: View paper

Brief Assessment

Multimodal Transformer Action[59] focuses on multimodal fusion for human action classification in kitchen scenarios using transformers, not on cross-modal representation learning between EMG signals and pose embeddings for gesture recognition.

Contribution 2: Community-aware soft contrastive learning (CASCLe) objective

Description: The authors propose CASCLe, a novel contrastive learning objective that constructs soft targets based on community-level structural similarities in the pose embedding space rather than treating all non-matching samples as equally distant negatives, thereby capturing semantic relationships between poses.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Open-world semantic segmentation via contrasting and clustering vision-language embedding

URL: View paper

Brief Assessment

Open-World Semantic Segmentation[48] focuses on vision-language alignment for semantic segmentation using image-caption pairs, not on EMG-pose alignment or gesture recognition. The soft contrastive learning approaches differ fundamentally in domain and application.

2. Structure-Enhanced Contrastive Learning for Graph Clustering

URL: View paper

Brief Assessment

Structure-Enhanced Clustering[49] focuses on graph clustering using community structure in network topology, while the original paper addresses cross-modal EMG-pose alignment using geometric proximity in embedding spaces. These are fundamentally different application domains and technical approaches.

3. Multi-Graph Contrastive Learning for Community Detection in Multi-Layer Networks

URL: View paper

Brief Assessment

Multi-Graph Contrastive[41] focuses on multi-layer network clustering using graph-based contrastive learning with consistency-diversity decomposition, not on cross-modal alignment or soft targets based on pose embedding geometry for gesture recognition.

4. Graph-text contrastive learning of inorganic crystal structure toward a foundation model of inorganic materials

URL: View paper

Brief Assessment

Graph-Text Crystal Structure[50] focuses on contrastive learning between crystal graph representations and text descriptions of inorganic materials. It does not address soft contrastive learning using community structure or geometric proximity in embedding spaces as proposed in CASCLe.

---

5. Adaptive graph contrastive learning for community detection

URL: View paper

Brief Assessment

Adaptive Graph Contrastive[42] focuses on graph-based community detection using contrastive learning on graph structures, not on cross-modal alignment between EMG and pose embeddings in continuous gesture spaces.

---

6. Rcoco: contrastive collective link prediction across multiplex network in Riemannian space

URL: View paper

Brief Assessment

Rcoco Multiplex[44] focuses on multiplex network alignment using community structure in Riemannian space for link prediction, not gesture recognition from EMG signals with pose embeddings.

---

7. Motif-Based Contrastive Learning for Community Detection

URL: View paper

Brief Assessment

Motif-Based Contrastive[43] applies contrastive learning to fuse higher-order motif structures with lower-order network information for community detection in graphs, not for cross-modal alignment between EMG and pose embeddings in continuous gesture spaces.

---

8. Single-View Graph Contrastive Learning with Soft Neighborhood Awareness

URL: View paper

Brief Assessment

Soft Neighborhood Awareness[47] focuses on single-view graph contrastive learning with soft neighborhood awareness for graph neural networks, not cross-modal EMG-pose alignment with community-based soft targets in embedding spaces.

---

9. Contrastive learning for multi-layer network community detection via learnable network augmentation

URL: View paper

Brief Assessment

Learnable Network Augmentation[45] focuses on multi-layer network community detection using contrastive learning between network layers, not on soft contrastive learning with community-level structural similarities in pose embedding spaces for gesture recognition.

---

10. Supporting clustering with contrastive learning

URL: View paper

Brief Assessment

Clustering with Contrastive[46] focuses on supporting clustering tasks in NLP using soft targets derived from instance-level similarities, not on cross-modal EMG-pose alignment or community-level structural similarities in pose embedding spaces.

Contribution 3: First zero-shot gesture classification from wearable EMG signals

Description: The authors claim that EMBridge is the first framework to achieve zero-shot gesture classification from wearable EMG signals, demonstrating the ability to recognize novel gestures without requiring training samples for those gestures.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Gesture Recognition of EMG Signals Based on Migration Learning

URL: View paper

Brief Assessment

Migration Learning Gesture[52] focuses on transfer learning strategies to improve cross-user generalization using pre-trained LSTM models, not zero-shot classification. The candidate requires training data from target gestures (even if reduced to 2%), whereas the original paper claims zero-shot capability without any training samples for novel gestures.

---

2. Brain-inspired self-organization with cellular neuromorphic computing for multimodal unsupervised learning

URL: View paper

Brief Assessment

Brain-Inspired Self-Organization[53] focuses on multimodal unsupervised learning using DVS/EMG data with self-organizing maps, not cross-modal representation learning for zero-shot gesture classification from wearable EMG signals.

---

3. From zero- to few-shot: deep temporal learning of wrist EMG enables scalable cross-user gesture recognition

URL: View paper

Prior Art Analysis

Zero to Few-Shot[21] demonstrates that zero-shot gesture classification from wearable EMG signals was achieved prior to the ORIGINAL paper's submission. The candidate explicitly reports achieving 78.2% zero-shot performance using wrist EMG for cross-user hand gesture recognition, directly contradicting the ORIGINAL paper's claim of being 'the first' to achieve this capability. Both papers address the same core problem: recognizing gestures from wearable EMG without requiring training samples from the target user. The candidate's work establishes that zero-shot classification from wearable EMG was already demonstrated in the literature.

Evidence

Evidence 1 - **Rationale**: This pair directly refutes the novelty claim. The ORIGINAL paper claims to be 'the first' to achieve zero-shot gesture classification from wearable EMG, while Zero to Few-Shot[21] explicitly reports achieving 78.2% zero-shot performance using wrist EMG, demonstrating that this capability existed prior to the ORIGINAL paper. - **Original**: to the best of our knowledge, embridge is the first cross-modal representation learning framework to achieve zero-shot gesture classification from wearable emg signals, showing potential toward real-world gesture recognition on wearable devices. - **Candidate**: in cross-user models, wrist emg demonstrated a zero-shot performance of 78.2%, compared to 71.6% for forearm emg (pâ<â0.05).

Evidence 2 - **Rationale**: Zero to Few-Shot[21] explicitly describes evaluating cross-user models that work with zero data from the end user (zero-shot), directly addressing the same problem space that the ORIGINAL paper claims to be first in solving. - **Original**: • to the best of our knowledge, embridge is the first cross-modal framework enabling zeroshot classification of unseen gestures for emg signals from wearable devices. - **Candidate**: this study therefore evaluated various cross-user models to reduce the calibration burden and compared wristand forearm-based models. eight different machine learning architectures were evaluated across 33 users, using varying amounts of data from the end user.

Evidence 3 - **Rationale**: Both papers address the same application domain of wearable EMG for hand gesture recognition, establishing that Zero to Few-Shot[21] is working on the identical problem that the ORIGINAL paper claims novelty for. - **Original**: hand gesture recognition on wearable devices has recently attracted significant interest (pyun et al., 2024; moin et al., 2021) and demonstrated potential across diverse applications such as rehabilitation (marcos-ant ´on et al., 2023), human-computer interaction (jarque-bou et al., 2021), and prost... - **Candidate**: wrist electromyography (emg) is emerging as an enticing wearable input modality for humanâmachine interaction. traditionally recorded from the forearm for use in transradial prostheses, wrist-based emg sensors are now being integrated into devices such as watches and wristbands for hand gesture reco...

Evidence 4 - **Rationale**: Both papers identify the same core challenge (generalizing without user-specific training/calibration) and Zero to Few-Shot[21] demonstrates solving it through cross-user models that achieve zero-shot performance, refuting the claim that the ORIGINAL paper is first to address this. - **Original**: however, predicting hand gestures from wearable emg, especially generalizing to unseen gestures without task-specific training, remains challenging (laput & harrison, 2019). - **Candidate**: consumer familiarity with wrist-worn devices makes wrist emg a compelling option, but the need for individualized user calibration remains a challenge. approach. this study therefore evaluated various cross-user models to reduce the calibration burden

---

4. Cross-subject EMG hand gesture recognition based on dynamic domain generalization

URL: View paper

Brief Assessment

Dynamic Domain Generalization[29] focuses on cross-subject generalization without calibration data but does not claim to be the first zero-shot gesture classification framework from wearable EMG signals, nor does it employ cross-modal representation learning with pose data as guidance.

---

5. Transfer-Modal Extraction of Surface EMG Features for Upper Limb Motor Classification

URL: View paper

Brief Assessment

Transfer-Modal Extraction[51] focuses on synthesizing EMG features from accelerometer signals for supervised classification, not zero-shot gesture recognition. It does not address cross-modal representation learning or zero-shot generalization capabilities.

---

6. Big data in myoelectric control: large multi-user models enable robust zero-shot EMG-based discrete gesture recognition

URL: View paper

Prior Art Analysis

Big Data Zero-Shot[11] demonstrates that zero-shot cross-user myoelectric control (gesture classification without user-specific training) was already achieved prior to the original paper. The candidate explicitly states achieving 93.0% classification accuracy on 306 completely unseen users who provided no training data, directly contradicting the novelty claim. Both papers address the same core problem of recognizing gestures from EMG signals without requiring training samples from test users, though they employ different technical approaches (discrete classification vs. cross-modal learning).

Evidence

Evidence 1 - **Rationale**: This pair demonstrates that Big Data Zero-Shot[11] already achieved zero-shot gesture classification from wearable EMG signals before the original paper. The candidate explicitly describes achieving classification on unseen users without any training data from those users, which is the definition of zero-shot classification that the original paper claims as a first. - **Original**: to the best of our knowledge, embridge is the first cross-modal representation learning framework to achieve zero-shot gesture classification from wearable emg signals - **Candidate**: this work dispels this notion, showing that true zero-shot cross-user myoelectric control is achievable without user-specific training. by taking a discrete approach to classification (i.e., recognizing the entire dynamic gesture as a single event), a classification accuracy of 93.0% for six gesture...

Evidence 2 - **Rationale**: This evidence pair shows that the candidate paper addresses the exact same problem space—achieving gesture classification from wearable EMG without user-specific training—and demonstrates it was already solved. The candidate's achievement of 'true zero-shot cross-user myoelectric control' directly refutes the claim of being 'first' to enable zero-shot classification for EMG signals from wearable devices. - **Original**: to the best of our knowledge, embridge is the first cross-modal framework enabling zeroshot classification of unseen gestures for emg signals from wearable devices. - **Candidate**: myoelectric control, the use of electromyogram (emg) signals generated during muscle contractions to control a system or device, is a promising modality for enabling always-available control of emerging ubiquitous computing applications. however, its widespread use has historically been limited by t...

---

7. From Pose to Muscle: Multimodal Learning for Piano Hand Muscle Electromyography

URL: View paper

Brief Assessment

Pose to Muscle[54] focuses on piano hand muscle EMG estimation from pose data, not zero-shot gesture classification. The candidate performs pose-to-EMG inference for piano performance, which is a different task from the original paper's zero-shot gesture recognition framework.

---

8. Multimodal Gesture Recognition using CNN-GCN-LSTM with RGB, Depth, and Skeleton Data

URL: View paper

Brief Assessment

Multimodal CNN-GCN-LSTM[55] focuses on multimodal gesture recognition using RGB, depth, and skeleton data, not wearable EMG signals or cross-modal learning frameworks for EMG-based zero-shot classification.

---

9. CPEP: Contrastive Pose-EMG Pre-training Enhances Gesture Generalization on EMG Signals

URL: View paper

Prior Art Analysis

CPEP[20] demonstrates that it achieved zero-shot gesture classification from wearable EMG signals before the ORIGINAL paper. Both papers use the same EMG2Pose dataset and similar cross-modal contrastive learning approaches to align EMG with pose representations. CPEP[20] explicitly claims to be 'the first framework enabling zero-shot classification of unseen gestures from emg signals' and presents experimental results showing zero-shot classification performance on both in-distribution and unseen gestures. The ORIGINAL paper's claim of being 'the first' is directly contradicted by CPEP[20]'s prior publication and identical capability.

Evidence

Evidence 1 - **Rationale**: Both papers explicitly claim to be 'the first' to achieve zero-shot gesture classification from EMG signals. CPEP[20]'s claim directly refutes the ORIGINAL paper's novelty claim, as CPEP[20] was published earlier and demonstrates the same capability. - **Original**: to the best of our knowledge, embridge is the first cross-modal representation learning framework to achieve zero-shot gesture classification from wearable emg signals, showing potential toward real-world gesture recognition on wearable devices. - **Candidate**: to the best of our knowledge, cpep is the first framework enabling zero-shot classification of unseen gestures from emg signals.

Evidence 2 - **Rationale**: Both frameworks use cross-modal alignment between EMG and pose to enable zero-shot classification. CPEP[20] demonstrates this approach was already established before the ORIGINAL paper's submission. - **Original**: we propose embridge, a cross-modal representation learning framework that bridges the modality gap between emg and pose. embridge learns highquality emg representations by introducing a querying transformer (q-former), a masked pose reconstruction loss, and a community-aware soft contrastive learnin... - **Candidate**: we introduce acontrastivepose-emgpretraining (cpep) framework to align emg and pose representations, where we learn an emg encoder that produces high-quality and pose-informative representations.

Evidence 3 - **Rationale**: Both papers use identical evaluation protocols (zero-shot classification and linear probing) on the same dataset (EMG2Pose) to demonstrate zero-shot gesture classification capabilities, showing CPEP[20] achieved this first. - **Original**: following clip evaluation protocol (radford et al., 2021), we validate the learned emg representations through zero-shot classification and linear probing, demonstrating superior performance on both in-distribution and unseen gestures compared to benchmark models. - **Candidate**: following clip evaluation protocol [16], we validate the learned emg representations through zero-shot classification and linear probing, demonstrating superior performance on both in-distribution and unseen gestures compared toemg2posebenchmark models.

Evidence 4 - **Rationale**: Both papers use the same EMG2Pose dataset with paired EMG-pose recordings for pre-training and evaluation, demonstrating that CPEP[20] already established this approach for zero-shot gesture classification. - **Original**: we utilize large-scale public emg datasets (salter et al., 2024; atzori et al., 2014), which provides simultaneous paired emg and pose recordings to pre-train our embridge model and perform downstream evaluations. - **Candidate**: we utilize the large-scale public emg datasetemg2pose[17], which provides simultaneous paired emg and pose recordings to pre-train our cpep model and perform downstream evaluations.

Appendix: Text Similarity Detection

Textual similarity detection checked 30 papers and found 3 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. CPEP: Contrastive Pose-EMG Pre-training Enhances Gesture Generalization on EMG Signals

Detected in: Core Task (sibling), Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] EMBridge: Enhancing Gesture Generalization from EMG Signals Through Cross-modal Representation Learning View paper
- [1] EMG-based Multi-User Hand Gesture Classification via Unsupervised Transfer Learning Using Unknown Calibration Gestures View paper
- [2] EMG-based online classification of gestures with recurrent neural networks View paper
- [3] Linear non-conservative unsupervised domain adaptation for cross-subject EMG gesture recognition View paper
- [4] An EMG foundation model for neural decoding View paper
- [5] Surface EMG hand gesture recognition system based on PCA and GRNN View paper
- [6] Hand gesture recognition using unsupervised learning View paper
- [7] Optimization of inter-subject sEMG-based hand gesture recognition tasks using unsupervised domain adaptation techniques View paper
- [8] Classification of electromyographic hand gesture signals using modified fuzzy C-means clustering and two-step machine learning approach View paper
- [9] ViT-HGR: Vision Transformer-based Hand Gesture Recognition from High Density Surface EMG Signals View paper
- [10] Spectral Collaborative Representation based Classification for hand gestures recognition on electromyography signals View paper
- [11] Big data in myoelectric control: large multi-user models enable robust zero-shot EMG-based discrete gesture recognition View paper
- [12] Distraction Detection and Intention Recognition for Gesture-Controlled Unmanned Aerial Vehicle Operation View paper
- [13] An EMG Gesture Recognition System with Flexible High-Density Sensors and Brain-Inspired High-Dimensional Classifier View paper
- [14] Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition View paper
- [15] A simplified wearable device powered by a generative EMG network for hand-gesture recognition and gait prediction View paper
- [16] Unsupervised domain adaptation for gesture identification against electrode shift View paper
- [17] Emgsense: A low-effort self-supervised domain adaptation framework for emg sensing View paper
- [18] A comprehensive analysis of wrist electromyography for hand gesture recognition: Advancing from forearm to wrist wearable devices View paper
- [19] Deep Feature Learning from Electromyographic Signals for Gesture Recognition Systems View paper
- [20] CPEP: Contrastive Pose-EMG Pre-training Enhances Gesture Generalization on EMG Signals View paper
- [21] From zero- to few-shot: deep temporal learning of wrist EMG enables scalable cross-user gesture recognition View paper
- [22] ReactEMG: Stable, Low-Latency Intent Detection from sEMG via Masked Modeling View paper
- [23] Disentangled EMG Representations Enable Zero-Shot Prediction of Multi-Finger Gestures via Linear and Compositional Latent Spaces View paper
- [24] Finger Motion Analysis for Interactive Applications Using Wearable and Wireless IoT Devices View paper
- [25] Optimization of EMG-based hand gesture recognition: Supervised vs. unsupervised data preprocessing on healthy subjects and transradial amputees View paper

- [26] Attributes' Importance for Zero-Shot Pose-Classification Based on Wearable Sensors View paper
- [27] Artificial intelligence techniques for biosensor data analysis View paper
- [28] Classifications of Dynamic EMG in Hand Gesture and Unsupervised Grasp Motion Segmentation View paper
- [29] Cross-subject EMG hand gesture recognition based on dynamic domain generalization View paper
- [30] EMG-UP: Unsupervised Personalization in Cross-User EMG Gesture Recognition View paper
- [31] Convolution Neural Network for EMG-Based Finger Gesture Classification for Novel and Trained Gestures View paper
- [32] Toward Highly Flexible Inter-User Calibration of Myoelectric Control Models With User-Defined Hand Gestures View paper
- [33] ReactEMG: Zero-Shot, Low-Latency Intent Detection via sEMG View paper
- [34] Unsupervised Learning of Wearable Sensor Data for Assessing Upper Limb Function View paper
- [35] Supervised and unsupervised learning in processing myographic patterns View paper
- [36] Posture-invariant myoelectric control with self-calibrating random forests View paper
- [37] Rejecting Unknown Gestures Based on Surface-Electromyography Using Variational Autoencoder. View paper
- [38] Feasibility of Using Wearable EMG Armbands combined with Unsupervised Transfer Learning for Seamless Myoelectric Control View paper
- [39] Review of: "Feasibility of Using Wearable EMG Armbands combined with Unsupervised Transfer Learning for Seamless Myoelectric Control" View paper
- [40] Foundation Models for EMG Human-Machine Interfaces View paper
- [41] Multi-Graph Contrastive Learning for Community Detection in Multi-Layer Networks View paper
- [42] Adaptive graph contrastive learning for community detection View paper
- [43] Motif-Based Contrastive Learning for Community Detection View paper
- [44] Rcoco: contrastive collective link prediction across multiplex network in Riemannian space View paper
- [45] Contrastive learning for multi-layer network community detection via learnable network augmentation View paper
- [46] Supporting clustering with contrastive learning View paper
- [47] Single-View Graph Contrastive Learning with Soft Neighborhood Awareness View paper
- [48] Open-world semantic segmentation via contrasting and clustering vision-language embedding View paper
- [49] Structure-Enhanced Contrastive Learning for Graph Clustering View paper
- [50] Graph-text contrastive learning of inorganic crystal structure toward a foundation model of inorganic materials View paper
- [51] Transfer-Modal Extraction of Surface EMG Features for Upper Limb Motor Classification View paper
- [52] Gesture Recognition of EMG Signals Based on Migration Learning View paper
- [53] Brain-inspired self-organization with cellular neuromorphic computing for multimodal unsupervised learning View paper
- [54] From Pose to Muscle: Multimodal Learning for Piano Hand Muscle Electromyography View paper
- [55] Multimodal Gesture Recognition using CNN-GCN-LSTM with RGB, Depth, and Skeleton Data View paper
- [56] A multimodal transformer framework with biomechanical constraints for injury prediction and human motion analysis View paper
- [57] Alignment-enhanced interactive fusion model for complete and incomplete multimodal hand gesture recognition View paper
- [58] From Wrist to Finger: Hand Pose Tracking Using Ring-Watch Wearables View paper
- [59] Multimodal Transformer Models for Human Action Classification View paper
- [60] Multimodal pose estimation and simulation modelling for real-time human motion analysis View paper
- [61] Smooth Multiscale Convolutional Attention Transformer Network for Continuous Motion Estimation of Hand Knuckle Angle using Surface EMG Signals View paper
- [62] Efficient transformer for sEMG-based hand pose estimation using inter-channel relationship View paper
- [63] PREDICTING CONTINUOUS HAND POSE FROM WEARABLE EMG SENSOR DATA USING TRANSFORMER-BASED DEEP-LEARNING MODELS View paper
- [64] FEASIBILITY OF AN INTEGRATED MULTI-MODEL APPROACH FOR DYNAMIC MUSCULOSKELETAL DISORDER RISK ASSESSMENT View paper
- [65] sEMG-vision Tra: A Gesture Recognition Method Based on Surface EMG Signal-Vision Fusion View paper