UniMTS: Unified Pre-training for Motion Time Series

Xiyuan Zhang, Diyan Teng, Ranak Roy Chowdhury, Shuheng Li, Dezhi Hong, Rajesh K. Gupta, Jingbo Shang

UC San Diego, Amazon, Qualcomm

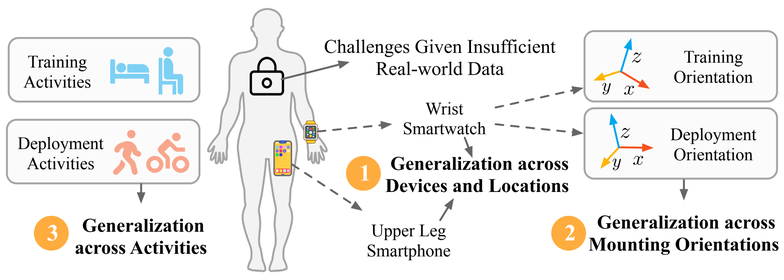
NEURAL INFORMATION PROCESSING SYSTEMS

TL; DR

The **first unified pre-training** framework for motion time series / human activity recognition that **generalizes across diverse device placements and activities**

Motivation

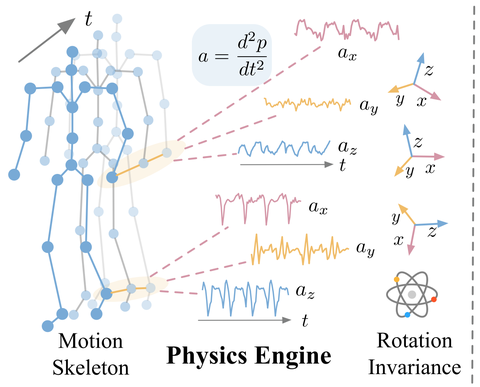
Task: Motion Time Series Classification / Human Activity Recognition (HAR) identifies human activities based on sensor time series collected from wearable devices (e.g., smartphone, smartwatch)

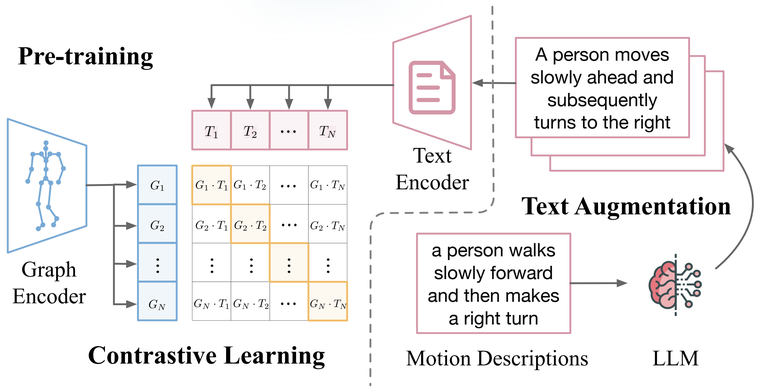


Challenge and our solution:

- ❓ **Generalization across devices and locations**: Smartwatch on the wrist vs smartphone near the upper leg
  - ✅ Synthesize time series from existing large-scale motion skeleton data with comprehensive joint coverage
  - ✅ Spatio-temporal graph neural network to model relationships across different device locations
- ❓ **Generalization across mounting orientations**: Smartphone facing towards or against the body in a pocket
  - ✅ Rotation-invariant augmentation
- ❓ **Generalization across activities**: Stationary activities such as lying or sitting, vs dynamic movements such as walking or cycling
  - ✅ Contrastive learning framework to align motion time series with LLM-enriched textual descriptions

Method

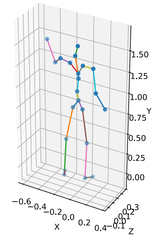




- **Physics engine for motion time series simulation**
- **Rotation-invariant augmentation**
- **Contrastive pre-training**
- **Text augmentation**

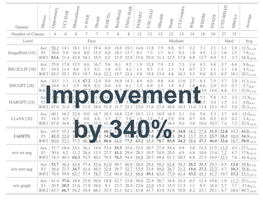
Experiments

Pre-training dataset: HumanML3D with 14,616 motion skeletons and ~3 textual descriptions for each motion skeleton. For example, the right figure shows the motion skeleton of “a person waves his hands”

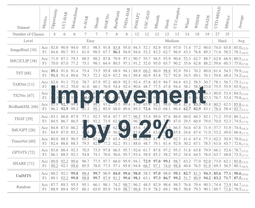


Evaluation dataset: 18 real-world datasets collected from various body locations, e.g., head, chest, back, arm, wrist, waist, leg, knee. 8 easy level datasets (< 10 activities), 8 medium level datasets (10 - 20 activities), 2 hard level datasets (> 20 activities)

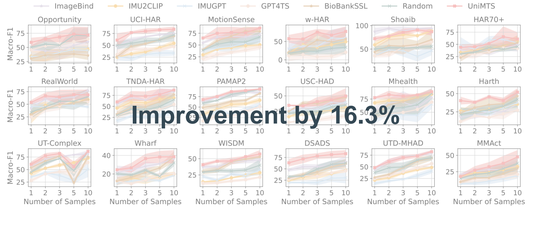
Baselines: (1) pre-trained HAR models (2) time-series classification models (3) traditional HAR models



Zero-shot setting

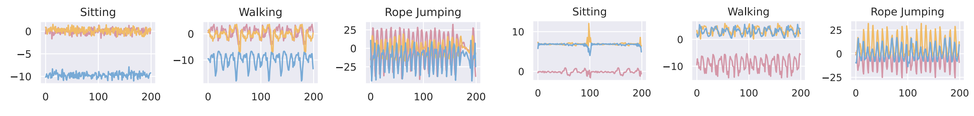


Full-shot setting



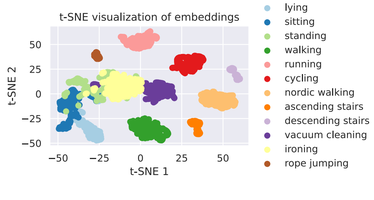
Few-shot setting

Simulated data have similar patterns as real data:

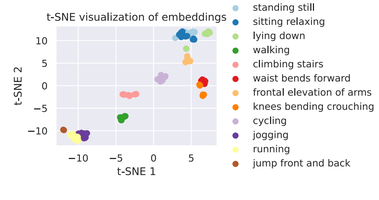


Simulated data | Real PAMAP2 data

Contrastive pre-training learns the semantics of time series:



PAMAP2



Mhealth

Paper

Code

Model