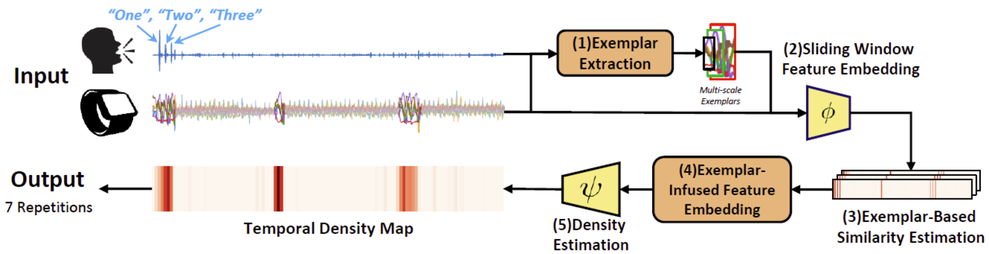
Count What You Want: Exemplar Identification and Few-shot Counting of Human Actions in the Wild

Yifeng Huang[1*], Duc Duy Nguyen[2*], Lam Nguyen[2], Cuong Pham[2,3] Minh Hoai[1,2]
*Equal contribution [1]Stony Brook University [2]VinAI
[3]Posts & Telecommunications Institute of Technology

Introduction

Our method focus on counting the number of repetitive actions with user-specified exemplars on wearable device. This exemplar is provided with audio cues("one", "two", "three").

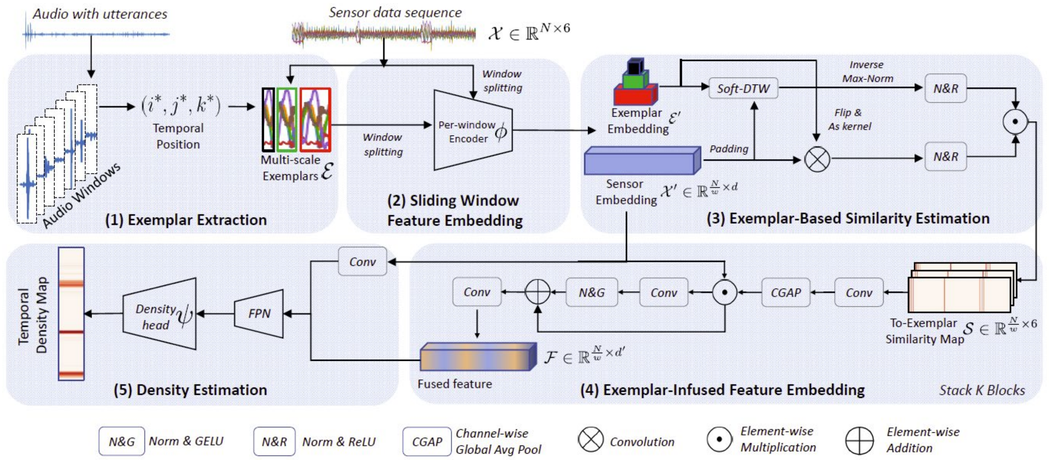


Motivation: Previous methodologies relying heavily on temporal self-similarity, encounter a significant limitation that they tend to focus solely on the most repetitive temporal patterns. This approach overlooks less frequent but potentially crucial actions.

Main Contributions:

- ❑ A novel strategy for using audio prompts to specify exemplars of what needs to be counted.
- ❑ A novel counting method that utilizes exemplars, incorporating a distance-preserving loss and an exemplar-based data synthesis pipeline.
- ❑ An unique dataset with multiple data modalities to develop a practical counting method for real-world scenarios.

Method



Pretraining with Synthesized Data :

- ❑ Fragment extraction with audio cues
- ❑ Augmentation with duration scaling, time shifting, amplitude scaling and random noise addition.

Training Objective :

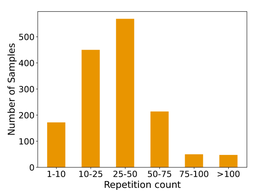
- ❑ Counting Loss
- ❑ Embedding Distance Preserving Loss

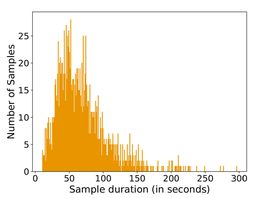
$$\mathfrak{L}_c = (\mathrm{sum}(\mathcal{T}) - \hat{c})^2$$

$$\mathfrak{L}_{pl} = \mathcal{X}'^T \mathcal{L} \mathcal{X}' \quad \mathcal{L} = \mathcal{D} - \mathcal{W}$$

$$\mathcal{W}_{ij} = \exp(-\frac{||\mathcal{X}_i - \hat{\mathcal{X}}_j||^2}{2\sigma^2})$$

DWC Dataset



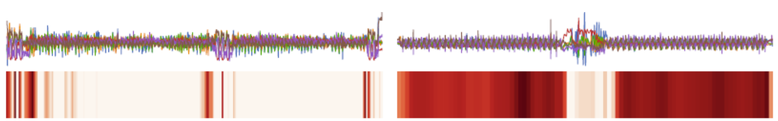


We introduce a new action counting dataset named DWC (Diverse Wearable Counting). This dataset, comprising 1502 entries from 37 subjects, spans seven categories — kitchen activities, household chores, physical exercises, factory tasks, daily routines, instrument-involved activities, and rehabilitation training. Encompassing 50 distinct action classes, DWC offers significantly greater diversity than existing datasets in this area.

Result

Method	Val Set		Test Set	
	MAE	RMSE	MAE	RMSE
Mean	17.18	21.91	14.80	17.49
Frequency-based	28.10	45.31	28.65	45.39
RepNet	11.95	17.33	10.82	14.75
TransRAC	14.51	20.40	12.97	16.82
Proposed	**7.66**	**12.25**	**7.47**	**13.09**

Components	Combinations				
Pretrain	✗	✗	✗	✗	✓
Dist. Preserving Loss	✗	✗	✗	✓	✓
Constrained Detection	✗	✗	✓	✓	✓
Similarity Estimation	✗	✓	✓	✓	✓
MAE	11.30	10.87	10.32	10.05	**7.66**
RMSE	16.15	15.23	14.96	14.72	**12.25**



(a) Predict: **9.7**, GT: **8**

(b) Predict: **60.7**, GT: **60**

Code & Data