FROM HUMAN POSE TO ON-BODY DEVICES FOR HUMAN-ACTIVITY RECOGNITION

Fernando Moya Rueda and Gernot A. Fink

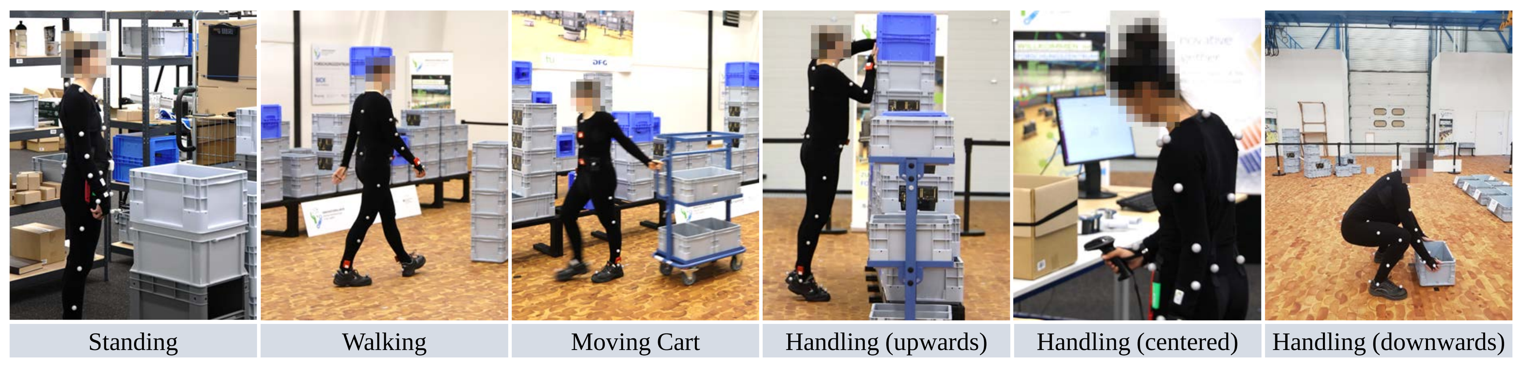
Department of Computer Science, TU Dortmund University, Germany

International Conference on Pattern Recognition, ICPR, Milan 2020.

Abstract

Human Activity Recognition (HAR), using inertial measurements from on-body devices, has not seen a great advantage from deep architectures. This drawback is mainly due to the lack of annotated data, diversity of on-body device configurations, the class-unbalance problem, and non-standard human activity definitions. Approaches for improving the performance of such architectures, e.g., transfer learning, are therefore difficult to apply. This paper introduces a method for transfer learning from human-pose estimations as a source for improving HAR using inertial measurements obtained from on-body devices. We propose to fine-tune deep architectures, trained using sequences of human poses from a large dataset and their derivatives, for solving HAR on inertial measurements from on-body devices. Derivatives of human poses will be considered as a sort of synthetic data for HAR. We deploy two different temporal-convolutional architectures as classifiers. An evaluation of the method is carried out on three benchmark datasets improving the classification performance.

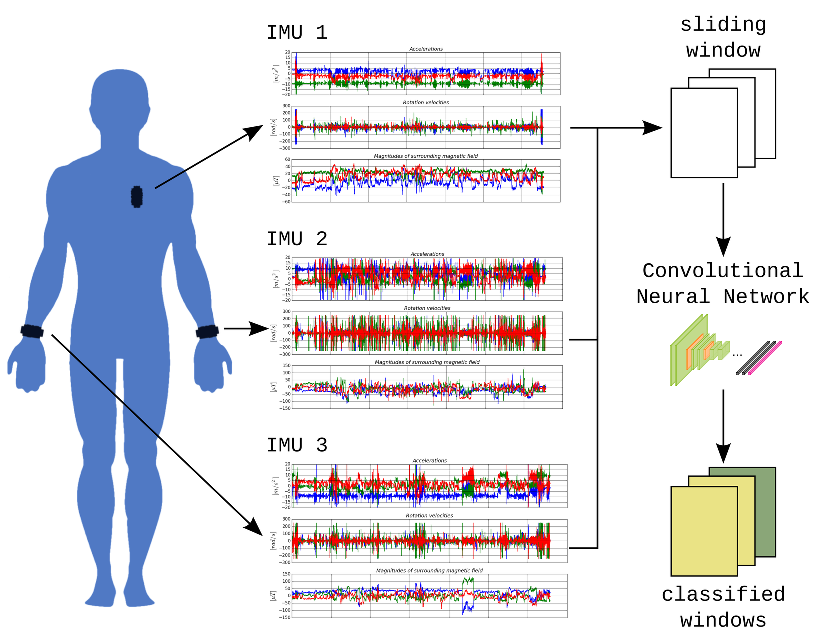
Source Dataset



Lara-Set	Record. Rate	Measurements	Dimensions	Channels
LARa-MoCap	200Hz	22 Joint Poses	$[Pos_{x,y,z}, Rot_{x,y,z}]$	132 Channels
LARa-OBs	100Hz	5 On-Body Devices	$[Acc_{x,y,z}, AngAcc_{x,y,z}]$	30 Channels
LARa-SOBS	200/100Hz	22 Joint Poses	$[Pos_{x,y,z}, Rot_{x,y,z}]$	132 Channels

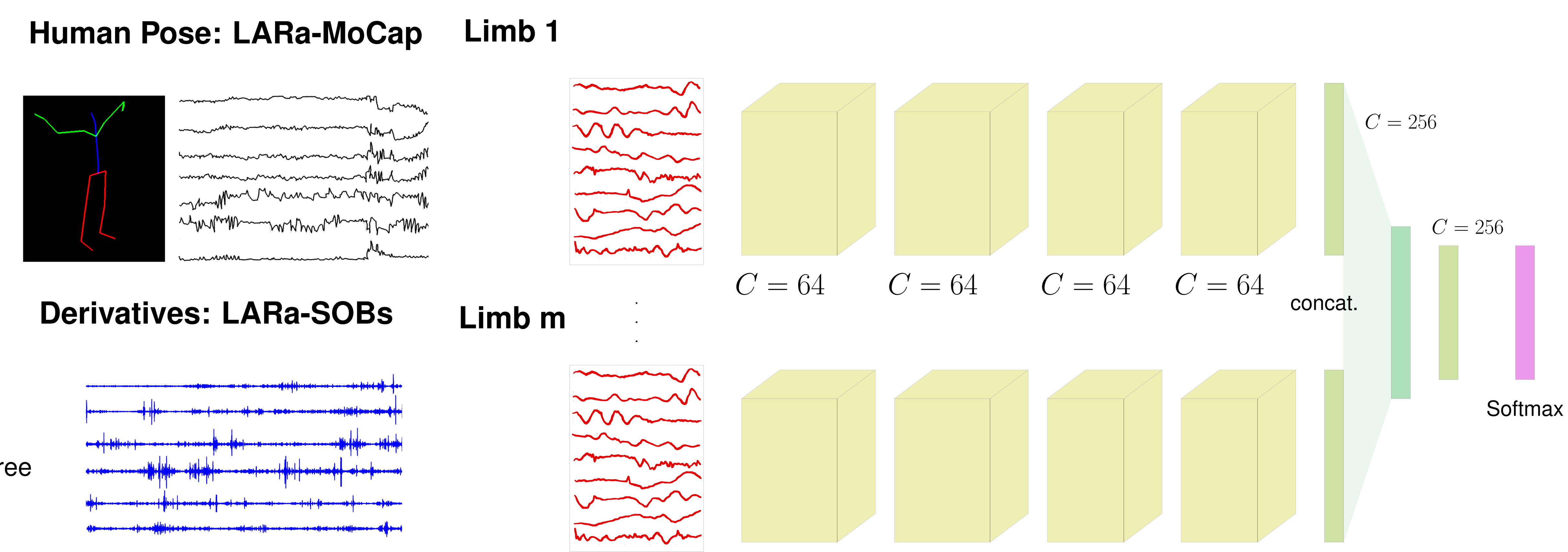
LARa is a large dataset of recordings of 714[min] from subjects in the Intralogistics. It consists of measurements from a marker-based MoCap system and On-Body Devices (OBs). A synthetic set is created by derivating sequences of joint poses, LARA SOBs.

Approach

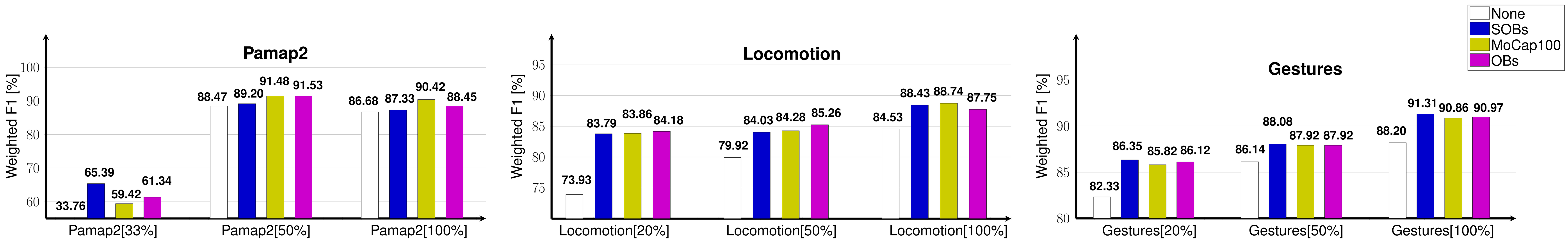


Overview of the proposed approach using an example of three IMUs worn by a worker at both wrists and the torso.

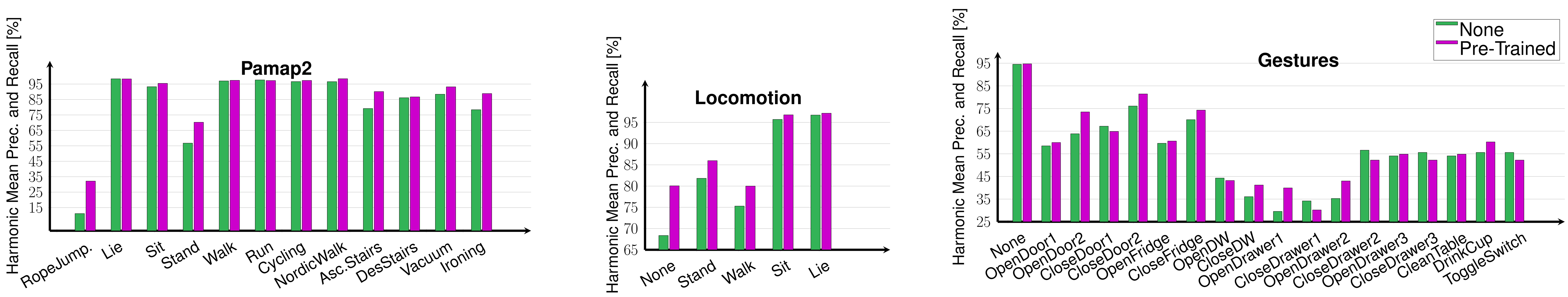
IMU-CNN Architecture



Experiments on Target Datasets



Weighted $F1[\%]$ of the datasets. The networks are trained from scratch (None) or pretrained on LARa datasets. Weighted $F1[\%]$ correspond to the mean of training/deploying the network five times. Three proportions of the training datasets are considered.



Harmonic mean of the precision and recall per class activity for the three test datasets, trained from scratch and using LARa sets for pre-training.

Architecture	Pamap2	Locomotion	Gestures
tCNN [1]	87.37*	87.80	85.10
CNN [2]	87.20*	-	90.80
IMU-tCNN [3]	89.01	88.23*	**92.15**
tCNN-SOBs	90.95	88.43	91.31
tCNN-MoCap	**91.48**	**88.74**	90.86
tCNN-OBs	**91.53**	87.75	**90.97**

Weighted $F1[\%]$ of the tCNN pretrained with LARa-MoCap and SOBs on three datasets. Results are compared with the benchmark networks of the dataset.

References

[1] F. J. Ordoñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," Sensors, vol.16, no. 1, p. 115, 2016

[2] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," arXiv preprint arXiv:1604.08880, 2016.

[3] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. ten Hompel, "Deep neural network based human activity recognition for the order picking process," in In Proc. of the 4th Int. Workshop on Sensor-based Activity Recognition and Interaction. ACM, 2017.