technische universität
dortmund

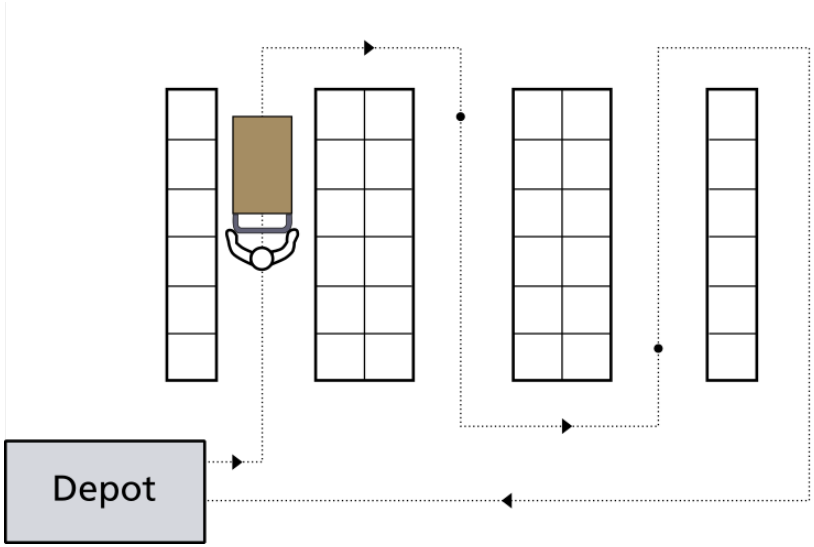
From Human Pose to On-Body Devices for Human-Activity Recognition

Fernando Moya Rueda, Gernot A. Fink

Pattern Recognition in Embedded Systems Group,
Technische Universität Dortmund

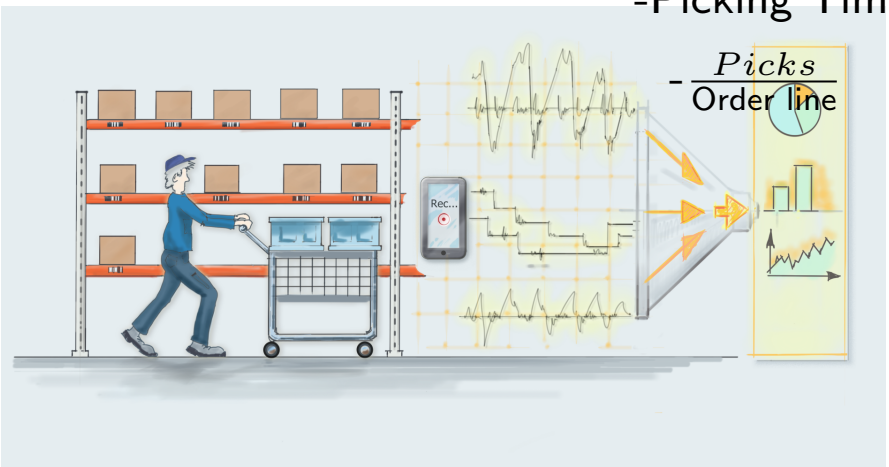
10. Dezember 2020

Human Activity Recognition HAR



. . . letting someone go shopping for you.

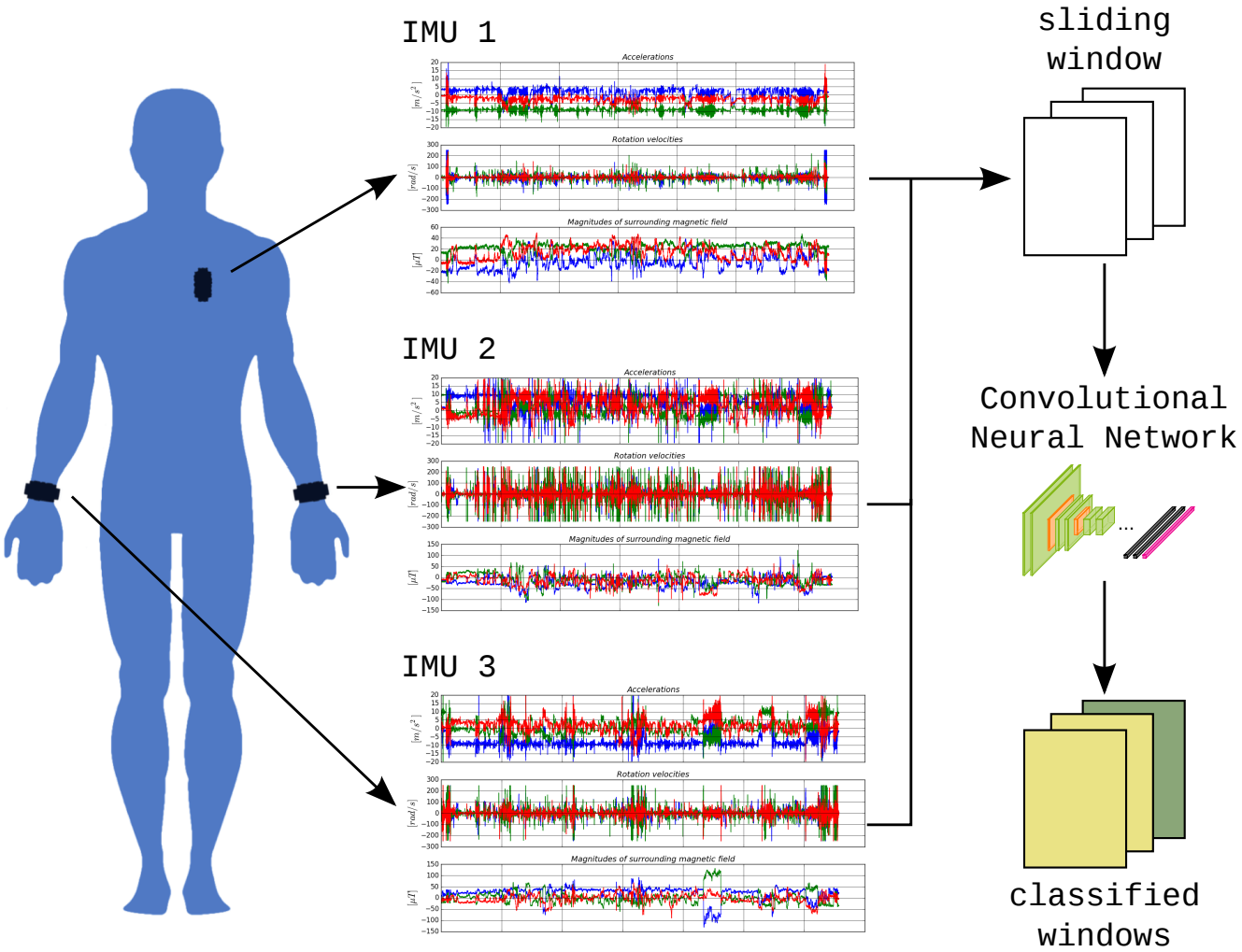
Measurements:
-Travel Time
-Picking Time
- $\frac{Picks}{\text{Order line}}$



Goal: analyze this industrial process using data from wearables.

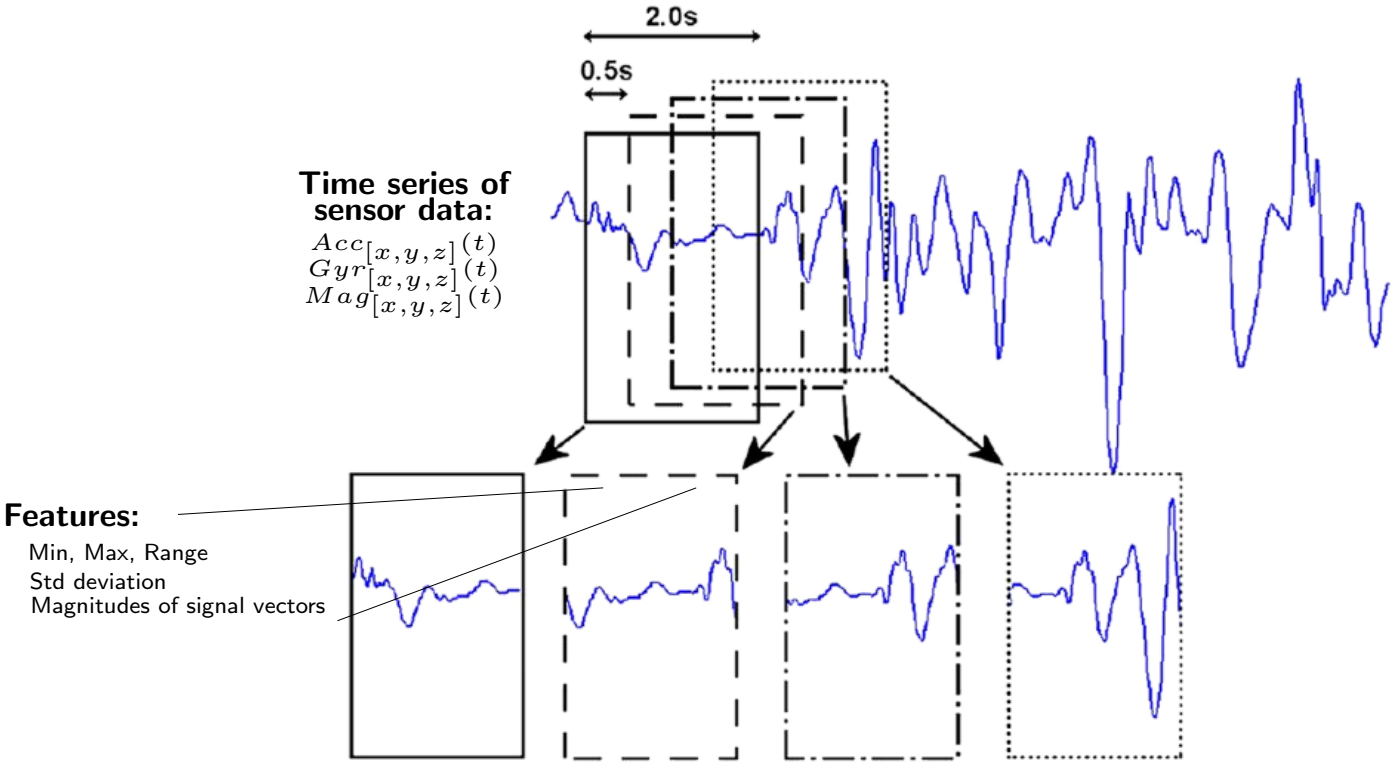
Data-driven HAR

- Human Activity Recognition using Wearables.



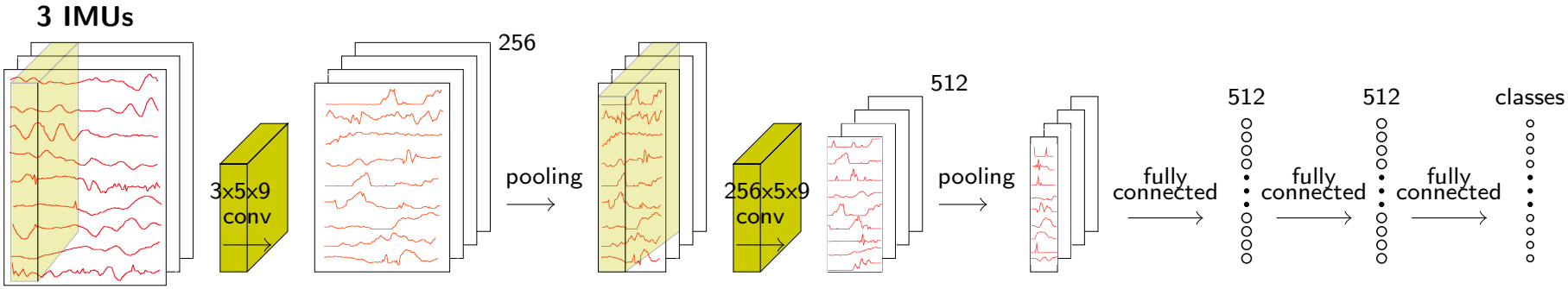
Related Work- Traditional approach

- Traditional pipeline using statistical features: min, max, average and standard deviation for analyzing the order picking task.



[1] S. Feldhorst, M. Masoudinejad, M. ten Hompel und G. A. Fink: *Motion Classification for Analyzing the Order Picking Process Using Mobile Sensors*. In: *Proc. Int. Conf. Pattern Recognition Applications and Methods (ICPRAM)*, Rome, Italy, 2016

Related Work- Temporal convolution networks



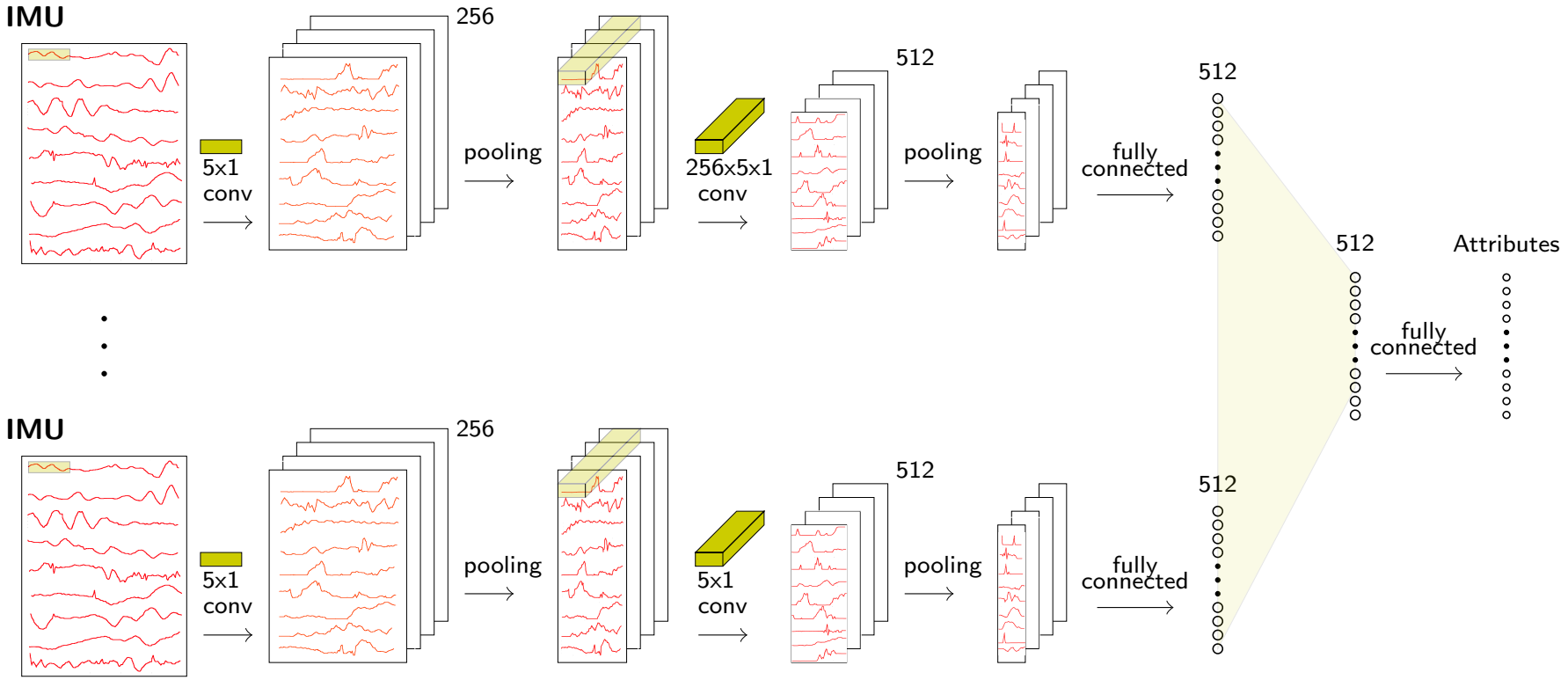
Deep convolutional neural network using temporal-convolution layers for multichannel time series HAR [1][2].

[1] J. Yang, M. N. Nguyen, P. P. San, X. Li und S. Krishnaswamy: *Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition.*. In: *IJCAI*, S. 3995–4001, 2015

[2] N. Y. Hammerla, S. Halloran und T. Plötz: *Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables*. In: *Proc. Int. Joint Conference on Artificial Intelligence (IJCAI)*, 2016

Related Work - Temporal Convolution Network

- CNN architecture processing IMUs independently and late fusion [1].



[1] R. Grzeszick, J. M. Lenk, F. M. Rueda, S. Feldhorst, M. ten Hompel und G. A. Fink: *Deep Neural Network based Human Activity Recognition for the Order Picking Process*. In: *Proc. Int. Workshop on Sensor-based Activity Recognition and Interaction (iWOAR)*, Rostock, Germany, 2017

Related Work - Annotation

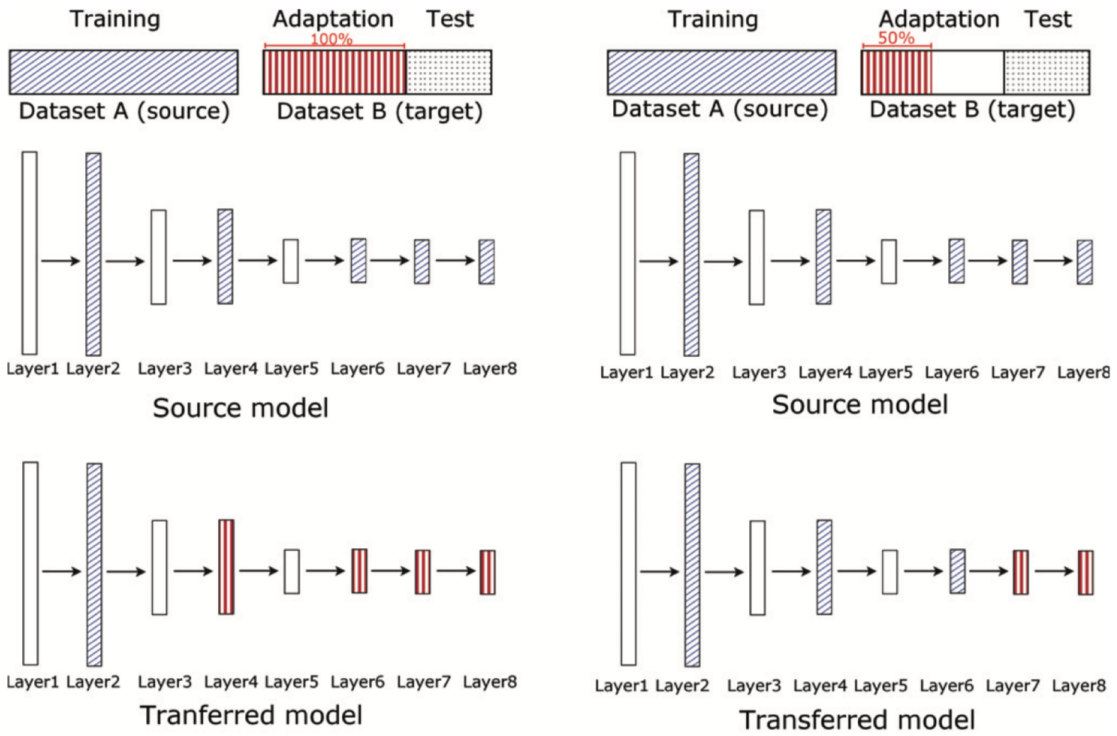
- Graphical representation of the HAR using attribute representations as an intermediate layer.

$N = 1$	Effort Assessment		
	Rec.	Total time[h]	Time per Rec.[h]
A1	38	$49:18$	$01:17 \pm 00:24$
A2	37	$72:06$	$01:56 \pm 00:34$
A3	36	$48:55$	$01:21 \pm 00:30$
A4	28	$37:19$	$01:19 \pm 00:37$
A5	30	$84:18$	$02:48 \pm 01:19$
A6	38	$23:11$	$00:36 \pm 00:17$
All	207	$303:27$	$01:25 \pm 00:57$

[1] F. Niemann, C. Reining, F. Moya Rueda, E. Altermann, N. R. Nair, J. A. Steffens, G. A. Fink und M. ten Hompel: *Logistic Activity Recognition Challenge (LARa) – A Motion Capture and Inertial Measurement Dataset*, 2020

Related Work - Transfer Learning

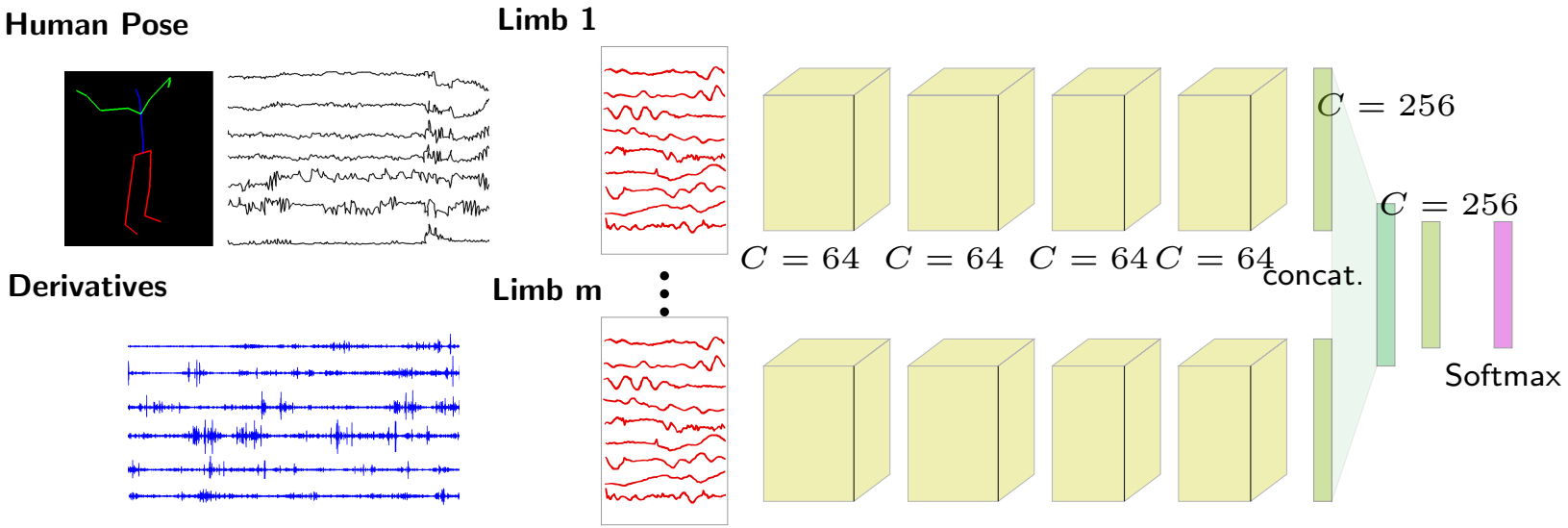
- Across subjects within the same dataset; different datasets; and across datasets, device positions, and sensor types



[1] F. J. Ordóñez Morales und D. Roggen: *Deep Convolutional Feature Transfer across Mobile Activity Recognition Domains, Sensor Modalities and Locations*. In: *Proc. of ACM ISWC*, S. 92–99, New York, NY, USA, 2016. Association for Computing Machinery

Approach

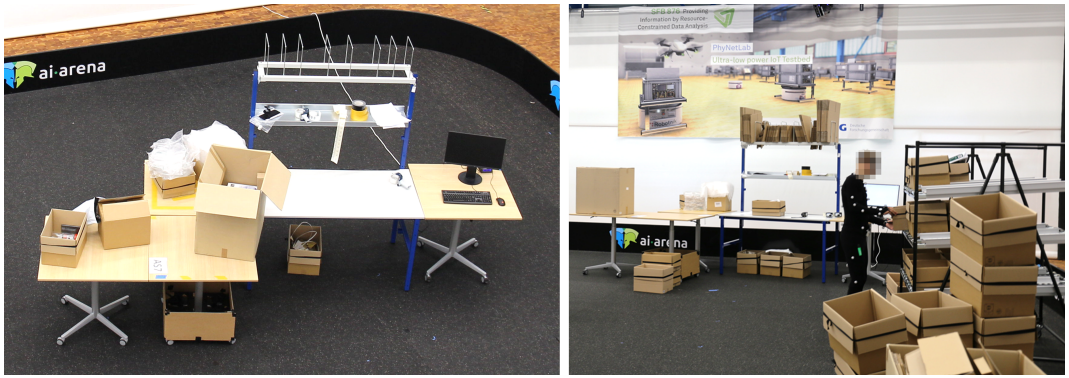
- Graphical representation of the HAR using attribute representations as an intermediate layer.



Dataset

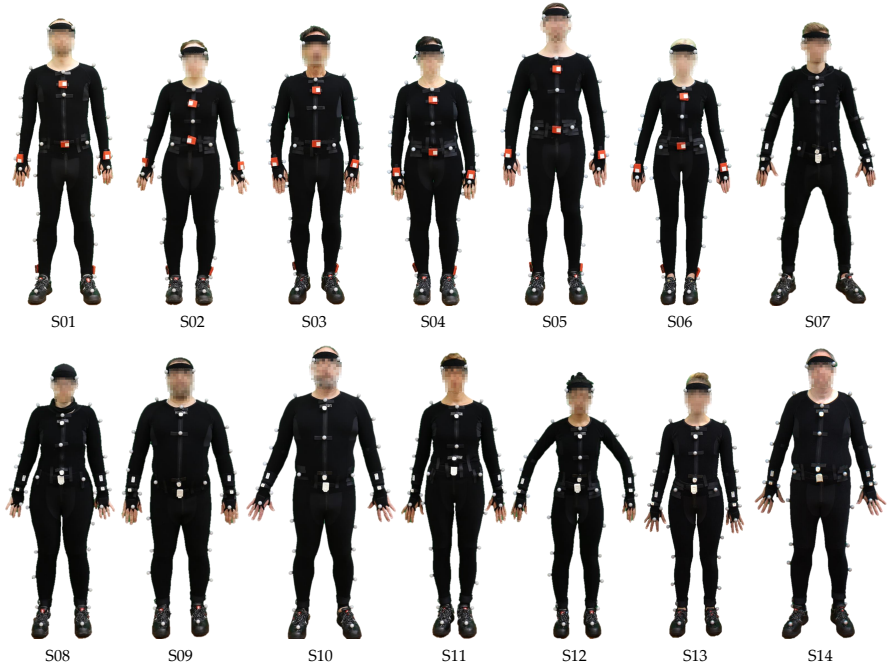
- LARa is a large dataset of recordings of 714[min] from subjects in the Intralogistics. It consists of measurements from a marker-based MoCap system and On-Body Devices (OBs). A synthetic set is created by derivating sequences of joint poses, LARA SOBs.

Standing	Walking	Moving Cart	Handling (upwards)	Handling (centered)	Handling (downwards)



[1] F. Niemann, C. Reining, F. Moya Rueda, E. Altermann, N. R. Nair, J. A. Steffens, G. A. Fink und M. ten Hompel: *Logistic Activity Recognition Challenge (LARa) – A Motion Capture and Inertial Measurement Dataset*, 2020

LARa Dataset



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

[1] F. Niemann, C. Reining, F. Moya Rueda, E. Altermann, N. R. Nair, J. A. Steffens, G. A. Fink und M. ten Hompel: *Logistic Activity Recognition Challenge (LARa) – A Motion Capture and Inertial Measurement Dataset*, 2020

Evaluation

- Baseline IMU-CNN using Softmax as a classifier on LARa
- Deep learning architectures for predicting Attributes
  - Opportunity [1] contains recordings from 7 OBs with a recording rate of 30HZ from 4 Subjects.
    - Locomotion
    - Gestures
  - Pamap2 [2] contains recordings from 3 OBs and a HR monitor with a recording rate of 100HZ from 9 Subjects.

[1] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán und D. Roggen: *The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition*. Pattern Recognition Letters, 34(15):2033–2042, 2013

[2] A. Reiss und D. Stricker: *Introducing a new benchmarked dataset for activity monitoring*. In: *ISWC*, S. 108–109. IEEE, 2012

Evaluation - HAR on LARa

- Datasets on Activities of Daily Living

Dataset	tCNN	IMU-tCNN
LARa MoCap	74.86 ± 0.7	75.47 ± 0.6
LARa MoCap100	76.07 ± 0.5	76.32 ± 0.2
LARa MoCap30	75.60 ± 0.3	76.53 ± 0.0
LARa SOBs	55.46 ± 0.6	60.06 ± 0.3
LARa SOBs30	56.16 ± 0.3	56.10 ± 0.4
LARa OBs	75.75 ± 0.4	75.65 ± 0.4
LARa OBs30	76.02 ± 0.0	75.63 ± 0.0

Evaluation - Transfer on Locomotion

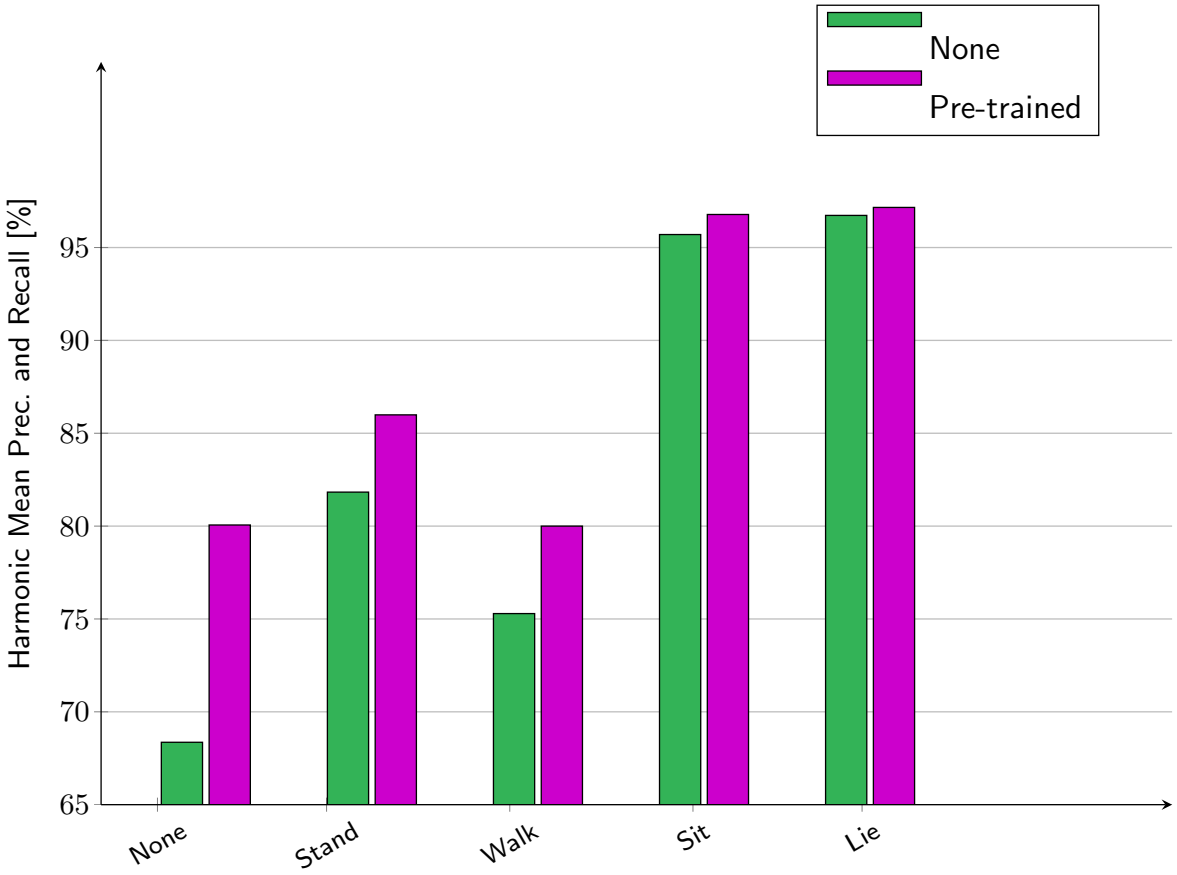
- Locomotion dataset using the tCNN and the IMU-tCNN. The networks are trained from scratch (None) or pretrained on LARa datasets. Mean and std from the $wF1\,[\%]$ are given as training method is repeated five times.

Datasets	tCNN			
	None	SOBs30	MoCap30	OBs30
Locomotion-20	73.93 ± 0.8	83.79 ± 0.6	83.86 ± 0.4	$\mathbf{84.18 \pm 0.1}$
Locomotion-50	79.92 ± 0.5	84.03 ± 0.0	84.28 ± 0.0	$\mathbf{85.26 \pm 0.0}$
Locomotion	84.53 ± 0.2	88.43 ± 0.3	$\mathbf{88.74 \pm 0.0}$	87.75 ± 0.3

Datasets	IMU-tCNN			
	None	SOBs30	MoCap30	OBs30
Locomotion-20	78.85 ± 1.0	80.53 ± 0.3	$\mathbf{81.45 \pm 0.4}$	76.92 ± 0.3
Locomotion-50	$77.01 \pm 0.$	80.80 ± 0.0	$\mathbf{83.52 \pm 0.3}$	77.92 ± 0.5
Locomotion	79.35 ± 0.5	86.26 ± 0.7	$\mathbf{87.61 \pm 0.4}$	83.52 ± 0.3

Evaluation - Transfer on Locomotion

- Harmonic Mean of Precision and Recall [%] per class activity for the **Locomotion** dataset using the tCNN, trained from scratch and using the LARa-MoCap.



Evaluation - Transfer on Gestures

- Gestures dataset using the tCNN and the IMU-tCNN. The networks are trained from scratch (None) or pretrained on LARa datasets. Mean and std from the $wF1\,[\%]$ are given as training method is repeated five times.

Datasets	tCNN			
	None	SOBs30	MoCap30	OBs30
Gestures-20	85.20 ± 0.3	84.60 ± 0.0	$\mathbf{85.60 \pm 0.1}$	84.45 ± 0.2
Gestures-50	$\mathbf{87.79 \pm 0.3}$	87.20 ± 0.1	87.45 ± 0.6	86.26 ± 0.2
Gestures	90.90 ± 0.1	$\mathbf{91.13 \pm 0.1}$	90.92 ± 0.3	90.64 ± 0.1

Datasets	IMU-tCNN			
	None	SOBs30	MoCap30	OBs30
Gestures-20	85.20 ± 0.3	84.60 ± 0.0	$\mathbf{85.60 \pm 0.1}$	84.45 ± 0.2
Gestures-50	$\mathbf{87.79 \pm 0.3}$	87.20 ± 0.1	87.45 ± 0.6	86.26 ± 0.2
Gestures	90.90 ± 0.1	$\mathbf{91.13 \pm 0.1}$	90.92 ± 0.3	90.64 ± 0.1

Evaluation - Transfer on Pamap2

- The $wF1\,[\%]$ of the tCNN and the IMU-tCNN on the **Pamap2** dataset. The networks are trained from scratch (None) or pretrained on LARa datasets. Mean and std from the $wF1\,[\%]$ are given as training method is repeated five times.

Datasets	tCNN			
	None	SOBs	MoCap100	OBs
Pamap2-33	33.8 ± 22.7	$\mathbf{65.4 \pm 3.2}$	59.4 ± 2.4	61.3 ± 2.1
Pamap2-50	88.5 ± 0.4	89.2 ± 0.2	91.5 ± 0.3	$\mathbf{91.5 \pm 0.7}$
Pamap2	87.0 ± 0.4	91.0 ± 0.4	89.7 ± 1.5	$\mathbf{91.2 \pm 0.6}$

Datasets	IMU-tCNN			
	None	SOBs	MoCap100	OBs
Pamap2-33	62.9 ± 0.8	60.4 ± 1.1	$\mathbf{62.1 \pm 4.0}$	49.7 ± 1.1
Pamap2-50	88.5 ± 0.9	88.9 ± 1.3	$\mathbf{91.0 \pm 0.3}$	84.7 ± 2.1
Pamap2	86.7 ± 0.0	87.3 ± 0.1	$\mathbf{90.4 \pm 0.2}$	88.5 ± 1.3

Evaluation - Comparison with the State of the Art

- The $wF1\,[\%]$ of the best tCNN and IMU-tCNN pretrained with LARa-MoCap and -SOBs on the three datatsets. Results are compared with the benchmark networks on the datasets.

Architecture	Pamap2	Locomotion	Gestures
tCNN[1]	87.37	87.8	85.1
tCNN[2]	87.2	-	90.8
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**

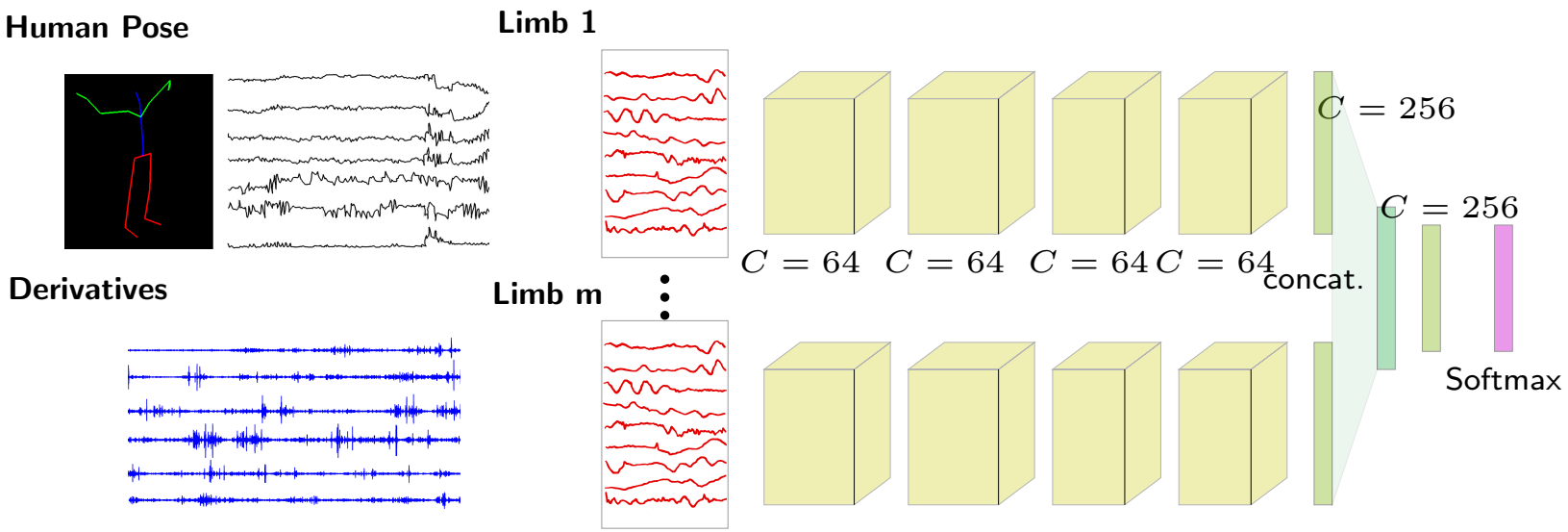
[1] F. J. Ordóñez und D. Roggen: *Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition*. Sensors, 16(1):115, 2016

[2] N. Y. Hammerla, S. Halloran und T. Plötz: *Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables*. In: *Proc. Int. Joint Conference on Artificial Intelligence (IJCAI)*, 2016

[3] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst und 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

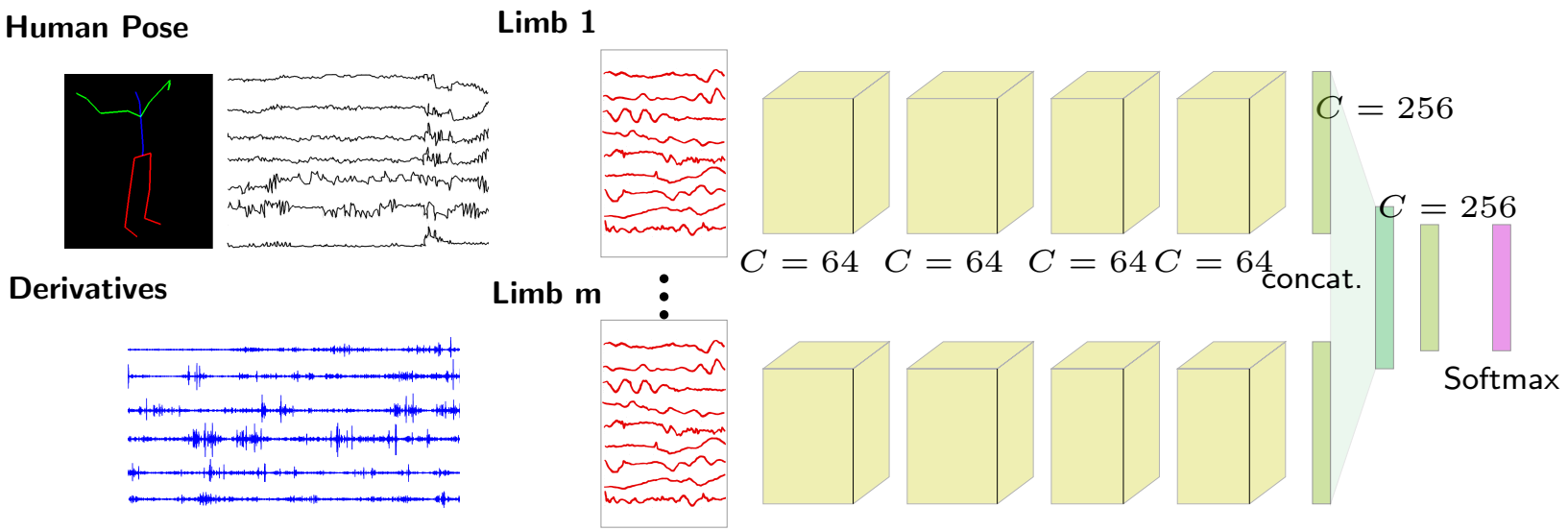
Conclusion

- We have presented an approach for solving multichannel time-series HAR using transfer learning across different data sources, namely, from human joint-poses to inertial measurements.
- We proposed to use synthetic-inertial measurements as an additional source. The synthetic measurements are derived from human joint-poses.



Conclusion

- This approach is a sort of transfer learning across three target domains with different physical, but related, measurements, different number of on-body devices, and recording rates.
- Architectures' performances improved for all three target datasets, even when deploying a proportion of them.



Conclusion

Thank you for your kind attention