HATRNet: Human Activity/Transition Recognition using Deep Neural Networks

Nicholas Gaudio, Akash Levy, Jonas Messner, {nsgaudio, akashl, messnerj}@stanford.edu
Department of Electrical Engineering, Stanford University

Stanford | ENGINEERING Electrical Engineering
Stanford | ENGINEERING Computer Science

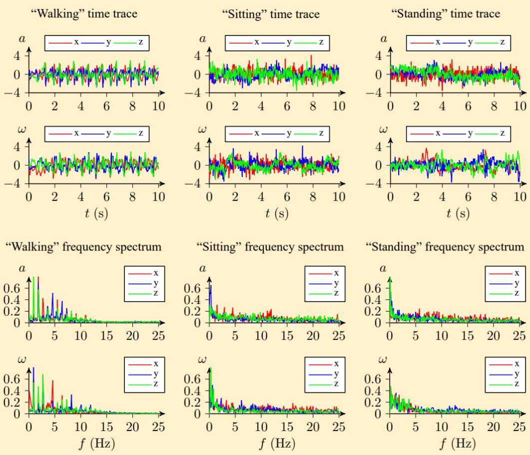
BACKGROUND and MOTIVATION

- Human Activity Recognition has great potential for customized healthcare
- Smartphones incorporate sensors (accelerometer, gyroscope etc.)
- Sensor data can be used to classify human activities and transitions

- Improvements compared to state of the art:
  1. Advanced preprocessing including data augmentation
  2. End-to-end deep learning solution (no feature extraction)
  3. Improved architecture enabling accurate classification of transitions

1. DATASET

- SBHAR dataset of 6 activities and 6 postural transitions from the Galaxy SII
- 3-axial linear acceleration and 3-axial angular velocities traces at 50 Hz
- Augmented to 3,640 examples with no feature selection (beyond FFT)

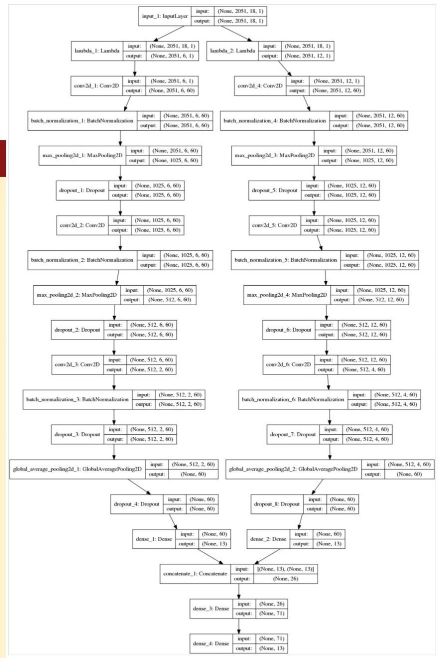


FUTURE WORK

- Model ensembling using data representations from sequence models
- Implement the model on a smartphone for real-time inference
- Incorporate frequency and phase traces into the sequence model

2. NEURAL NETWORK ARCHITECTURE

Siamese (non-weight sharing) CNN



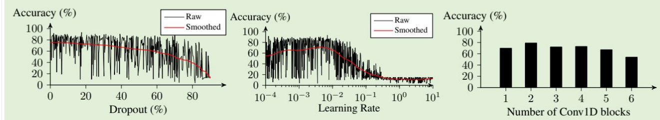
Siamese (non-weight sharing) CNN

- Left subnetwork takes time traces as input (6 zero-padded channels)
- Right subnetwork takes frequency and phase traces as input (12 interpolated channels)
- Late sensor fusion employed for encoded, efficient feature extraction
- Conv1D filter size: 1x14 Conv2D filter size: 3x42 Filter #: 60 learning rate: 0.0026

Sequence (LSTM) Model

- Takes time traces as input (6 channels of variable length)
- Two LSTM layers (128 to 32 output channels)

Random Coarse- to Fine-grain Hyperparameter Search



REFERENCES

[1] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomput.*, vol. 171, pp. 754–767, Jan. 2016.
[2] P. Kasnesis, C. Z. Patrikakis, and I. S. Venieris, "Perceptionnet: A deep convolutional neural network for late sensor fusion," *CoRR*, vol. abs/1811.00170, 2018.
[3] C. A. J. O. D. Fuentes, L. Gonzalez-Abril, "Online motion recognition using an accelerometer in a mobile device," *Expert systems with applications*, 2011.
[4] M. Kose, O. Incel, and C. Ersoy, "Online human activity recognition on smart phones," *Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data*, 01 2012.
[5] J. Qi, P. Yang, M. Hanneghan, and S. Tang, "Multiple density maps information fusion for effectively assessing intensity pattern of lifelogging physical activity," *Neurocomputing*, vol. 220, pp. 199–209, 2017.
[6] C. Reiff, K. Marlatt, and D. Dengel, "Difference in caloric expenditure in sitting versus standing desks," *Journal of physical activity & health*, vol. 9, pp. 1009–11, 09 2012.
[7] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*, 01 2013.
[8] A. L. Guennec, S. Malinowski, and R. Tavenard, "Data augmentation for time series classification using convolutional neural networks," *ECML/PKDD Workshop on Advanced Analytics andLearning on Temporal Data*, 2016.
[9] F. Chollet *et al.*, "Keras." https://keras.io, 2015.
[10] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.
[11] E. Jones, T. Oliphant, P. Peterson, et al., "SciPy: Open source scientific tools for Python," 2001.
[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python ," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[13] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "cudnn: Efficient primitives for deep learning," *arXiv preprint arXiv:1410.0759*, 2014.

3. RESULTS and DISCUSSION

Literature comparison (grouped postural transitions):

SVM, 561 features extracted [1]

	WA	WU	WD	SI	ST	LD	PT
WA	1834	64	5	3	2	0	1
WU	10	1743	51	5	5	0	16
WD	0	2	1671	1	7	0	1
SI	0	0	0	1875	**94**	6	3
ST	0	2	0	**109**	2049	0	1
LD	0	0	0	1	0	2148	2
PT	0	1	2	0	0	0	1036

CNN (2.5 second traces) [2]

	WA	WU	WD	SI	ST	LD
WA	487	0	9	0	0	0
WU	2	468	0	0	0	1
WD	0	0	420	0	0	0
SI	0	2	0	443	**46**	0
ST	0	0	0	**16**	516	0
LD	0	0	0	0	0	537

HATRNet Results (ungrouped postural transitions):

Siamese (non-weight sharing) CNN

	WA	WU	WD	SI	ST	LD	ST-SI	SI-ST	SI-L	L-SI	ST-L	L-ST
WA	21	0	1	0	0	0	0	0	0	0	0	0
WU	0	36	0	0	0	0	1	0	0	0	0	0
WD	0	0	33	0	0	0	0	0	0	0	0	0
SI	0	0	0	20	**0**	0	0	0	0	0	0	0
ST	0	0	0	**2**	27	0	0	0	0	0	0	0
LD	0	0	0	0	0	26	0	0	0	0	0	0
ST-SI	0	0	0	0	0	0	19	0	0	0	0	0
SI-ST	0	0	0	0	0	0	0	10	0	0	0	0
SI-L	0	0	0	0	0	0	0	0	11	0	0	0
L-SI	0	0	0	0	0	0	0	0	0	15	0	2
ST-L	0	0	0	0	0	0	1	0	1	0	10	0
L-ST	0	0	0	0	0	0	0	0	0	0	0	7

WA: Walking, WU: Walking-Upstairs, WD: Walking-Downstairs, SI: Sitting, ST: Standing, LD: Laying-Down, PT: Postural Transition, ST-SI: Stand-to-Sit, SI-ST: Sit-to-Stand, SI-L: Sit-to-Lie, L-SI: Lie-to-Sit, ST-L: Stand-to-Lie, L-ST: Lie-to-Stand

Architecture Comparison:

	CNN1	CNN2	LSTM	SVM [1]	Perceptionnet (CNN) [2]
Number of categories	12	7	12	7	6
Error Rate	3.29 %	0.82 %	18.11 %	3.22 %	2.75 %