Learning deep and shallow features for human activity recognition.

SANI, S., MASSIE, S., WIRATUNGA, N. and COOPER, K.

2017
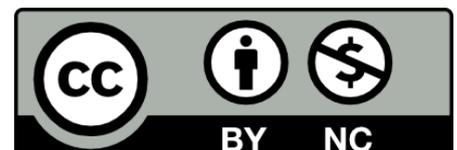
Learning Deep and Shallow Features for Human Activity Recognition

Sadiq Sani[1], Stewart Massie[1], Nirmalie Wiratunga[1], and Kay Cooper[2]

[1] School of Computing Science and Digital Media,
Robert Gordon University,
Aberdeen AB25 1HG, Scotland, UK
{s.sani|s.massie|n.wiratunga}@rgu.ac.uk

[2] School of Health Sciences,
Robert Gordon University, Aberdeen

Abstract. selfBACK is an mHealth decision support system used by patients for the self-management of Lower Back Pain. It uses Human Activity Recognition from wearable sensors to monitor user activity in order to measure their adherence to prescribed physical activity plans. Different feature representation approaches have been proposed for Human Activity Recognition, including shallow, such as with hand-crafted time domain features and frequency transformation features; or, more recently, deep with Convolutional Neural Net approaches. The different approaches have produced mixed results in previous work and a clear winner has not been identified. This is especially the case for wrist mounted accelerometer sensors which are more susceptible to random noise compared to data from sensors mounted at other body locations e.g. thigh, waist or lower back. In this paper, we compare 7 different feature representation approaches on accelerometer data collected from both the wrist and the thigh. In particular, we evaluate a Convolutional Neural Net hybrid approach that has been shown to be effective on image retrieval but not previously applied to Human Activity Recognition. Results show the hybrid approach is effective, producing the best results compared to both hand-crafted and frequency domain feature representations by a margin of over 1.4% on the wrist.

Keywords: Human activity recognition, feature representation, deep learning

1 Introduction

Human activity recognition (HAR) is the computational discovery of human activity from sensor data. It is receiving increasing interest in the areas of health care and fitness [9], largely motivated by the need to find creative ways to encourage physical activity. HAR is generally considered as a Machine Learning classification problem. However, there is little consistency in the research on which feature representation and classification models should be applied; to some extent this may reflect the differences that can exist between the detailed characteristics of different HAR problems. In particular, the sensor type and location may vary between applications, as can the activity types that need to be recognised.

The motivation for this work is to develop an effective HAR component for use in selfBACK [3], an EU funded project that is developing a self-management system for patients with Lower Back Pain (LBP). Recent published guidelines for the management of non-specific LBP advises patients to remain active and avoid excessive bed rest or extended periods of inactivity [1]. The HAR component monitors the patient activity using sensor data that is continuously polled from a wearable device. User activities are recognised in real time and if selfBACK detects continuous periods of sedentary behaviour, a notification is given to alert the user. In addition, a daily activity profile is generated and compared with a prescribed activity plan. The information in this daily profile includes the durations of activities and, for ambulation activities (e.g. walking, running, stair climbing), the intensity of activity, and the counts of steps taken.

Given the importance of accurate recognition of user activity for selfBACK, an effective HAR component becomes crucial. Different feature representation approaches for HAR have been proposed, from shallow hand-crafted features to frequency transformation features e.g. Fast Fourier Transform (FFT) and Discrete Cosine Transform (DCT) coefficients, and more recently, deep learning approaches [9, 17, 18]. All these approaches have had some degree of success and setbacks in performance [15]. However, previous work has presented mixed results and does not provide a clear answer on which feature representation approach to adopt. Given the intended real-world use case of selfBACK, it is important to determine the appropriate feature extraction approach to use for our preferred sensor configuration (single wrist-mounted accelerometer) and our selected activity classes (walking, jogging, up-stairs, downstairs, standing, sitting). Hence in this paper, we provide an empirical evaluation of 7 different feature representation approaches across the three different classes of features i.e. shallow hand-crafted, shallow frequency transformed, and deep CNN derived.

Wrist data presents additional challenges as it has different characteristics to data collected from other body locations (e.g. thigh). It is more prone to random noise due to increased variations in movement and posture possible with the hand during activities. As a result, the performance of HAR with a wrist-mounted sensor may suffer. However, long-term usage of an application may be improved with a less obtrusive wrist-worn sensor. It is therefore important to compare both locations, in order to assess the trade-off between performance and usability. Hence, we provide an empirical evaluation of wrist-mounted and thigh-mounted accelerometer data.

The main contributions of this work are:

- the evaluation of deep and shallow features for HAR for the specific activities required for the management of LBP, with sensors located on both the wrist and thigh;

- a CNN hybrid approach has been introduced for HAR in which deep features are generated from the raw accelerometer data and then fed to a traditional classifiers; and

- a new public dataset has been made available to the research community to support future work in this area.

The rest of this paper is organised as follows: in Section 2, we highlight important related work on HAR. Section 3 discusses the different feature representation approaches. Evaluation, describing our dataset and experiments, is presented in Section 4 with our conclusions in Section 5.

---

[3] http://www.selfback.net/

2 Related Work on HAR

In this work, we focus on sensor input from a single tri-axial accelerometer mounted on a person's wrist or thigh. A tri-axial accelerometer sensor measures changes in acceleration in 3 dimensional space [9]. Other types of wearable sensors have also been proposed e.g gyroscope. A recent study compared the use of accelerometer, gyroscope and magnetometer for activity recognition [19]. The study found the gyroscope alone was effective for activity recognition while the magnetometer alone was less useful. However, the accelerometer still produced the best activity recognition accuracy. Other sensors that have been used include heart rate monitor [20], light and temperature sensors [13]. These sensors are however typically used in combination with the accelerometer rather than independently.

Some studies have proposed the use of a multiplicity of accelerometers [4, 12] or combination of accelerometer and other sensor types placed at different locations on the body. These configurations however have limited practical use outside of a laboratory setting. In addition, limited improvements have been reported from using multiple sensors for recognising every day activities [6] which may not justify the inconvenience, especially as this may hinder the real-world adoption of the activity recognition system. For these reasons, some studies e.g. [11] have limited themselves to using single accelerometers which is also the case for selfBACK [3].

Another important consideration is the placement location of the sensor. Several body locations have been proposed e.g. thigh, hip, back, wrist and ankle. Many comparative studies exist that compare activity recognition performance at these different locations [4]. The wrist is considered the least intrusive location and has been shown to produce high accuracy especially for ambulation and upper-body activities [11]. Hence, this is the chosen sensor location for our system.

Many different feature extraction approaches have been proposed for accelerometer data for the purpose of activity recognition [9]. Most of these approaches involve extracting statistics e.g. mean, standard deviation, percentiles etc. on the raw accelerometer data (time domain features). FFTs applied to the raw data have been shown to be beneficial. Typically this requires a further preprocessing step applied to the resulting FFT coefficients in order to extract features that measure characteristics such as spectral energy, spectral entropy and dominant frequency [5]. Another approach is to use coefficients obtained from applying DCT on the raw accelerometer data as features [17]. This is particularly attractive as it avoids the need for further preprocessing of the data to extract features to generate instances for the classifiers.

Recently, deep learning approaches have been applied to the task of HAR due to their ability to extract features in an unsupervised manner. In particular, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Deep Belief Networks (DBNs), have all been applied with promising results [7]. One of the early approaches to apply deep belief networks for feature learning is [14]. Here the authors use stacked Restricted Boltzmann Machines (RBMs) to learn higher level features from raw accelerometer data. RBMs are fully connected bipartite graphs that generatively model input data by training a set of stochastic binary hidden input units which function as low-level feature detectors. A comparative evaluation with other representation approaches,

e.g. FFT and Principal Component Analysis (PCA), did not present RBM as a clear winner. Four different datasets were used in the evaluation, of which only one (Skoda) used a single wrist mounted accelerometer. However, this dataset was collected from a single person and the activity classes in the dataset are car assembly activities which are very different from the usage scenario with selfBACK.

More recently, CNNs have been popularly applied to the task of HAR. One of the main advantages of CNNs is their ability to model local dependencies that may exist between adjacent data points in the accelerometer data [21]. In [16], CNNs are used for representation of both accelerometer and gyroscope data from the Samsung mobile phone dataset [2]. The dataset also provides a set of 561 hand-crafted features. The evaluation compares a CNN classifier architecture (with 3 hidden layers) with four other classifiers, including SVM and Naive Bayes that use the hand-crafted features. Results showed deep features to be on par with hand-crafted features using SVM.

Zeng et al. used CNNs with a modified weight sharing technique called partial weight sharing [21]. Evaluation compared performance of standard CNNs, partial weight sharing CNNs, time-domain hand crafted features, PCA and RBMS on three datasets using accuracy. Results show the CNN with partial weight sharing to perform best. However, only the Skoda datasets used a wrist mounted accelerometer and performance was reported as accuracy which can be misleading given that the datasets contained unbalanced class distributions. Furthermore the comparative study considered only kNN on just the raw features and did not consider state-of-the-classifiers such as SVM.

Huang et al. found that while CNNs are good at learning invariant features, SVMs may be better at producing decision surfaces for classification [8]. For image classification they show that a hybrid approach, with a SVM trained on the features learned by the convolutional network, outperforms standard CNNs. Our work applies a similar approach but instead of image classification our focus is on studying comparative performance of deep hybrid features on HAR.

3 Feature Representation

The feature representations considered in this work can be divided into three categories: hand-crafted features, frequency transformation features and deep features.

3.1 Hand-crafted Features

Shallow hand-crafted features are the most common used in HAR and involve the computation of a number of defined measures on either the raw accelerometer data (time-domain) or the frequency transformation of the data (frequency domain). These measures are designed to capture the characteristics of the signal that are useful for distinguishing different classes of activities by bridging the semantic gap from low level sensor features to richer representations. Hand-crafted features can be extracted both from the original time series accelerometer values (time-domain features) or from a frequency transformation of the accelerometer data (frequency-domain features). For time and frequency domains, the input is a vector of real values $\overrightarrow{v} = v_1, v_2, ...v_n$ for

Table 1. Hand-crafted features for both time and frequency domains.

Time Domain Features
Mean
Standard deviation
Inter-quartile range
Lag-one-autocorrelation
Percentiles (10,25,50,75,90)
Peak-to-peak amplitude
Power
Skewness
Kurtosis
Log-energy
Zero crossings
Root squared mean

Frequency Domain Features
Dominant frequency
Spectral centroid
Maximum
Mean
Median
Standard deviation

each axis x, y and z. A function θ_i is then applied to each vector to compute a single feature value. The time-domain and frequency domain features used in this work are presented in Table 1, see [17] for more details on these features.

While hand-crafted features have worked well for HAR, a significant disadvantage is that they are sensor-type specific. A different set of features need to be defined for each different type of input data i.e. accelerometer, gyroscope, time-domain and frequency domain. Hence, some understanding of the characteristics of the data is required. In addition, it is not always clear which features are likely to work best. The choice of features is usually made through empirical evaluation of different combinations of features or with the aid of feature selection algorithms.

3.2 Frequency Transform Features

Feature extraction from the frequency domain involves first applying a single function f on the raw accelerometer data to map into an alternative representation space (i.e. the frequency domain), where it is expected that distinctions between different activities will be more emphasised. The intuition is that some activities have particular characteristics, such as repetitive actions, which will become more evident. The main difference between frequency transform and hand-crafted features is that the coefficients of transformation are directly used for feature representation without taking further measurements. Common transformations include FFTs and DCTs.

FFT is an efficient algorithm optimised for computing the discrete Fourier transform of a digital input. Fourier transforms decompose an input signal into its constituent sine waves. In contrast, DCT, a similar algorithm to FFT, decomposes a given signal into it's constituent cosine waves. Also, DCT returns an ordered sequence of coefficients such that the most significant information is concentrated at the lower indices of the sequence. This means that higher DCT coefficients can be discarded without losing information, making DCT better for compression.

For frequency transform features, a transformation function (DCT or FFT) ϕ is applied to time-series accelerometer vector $\overrightarrow{v}$ of each axis. The output of ϕ is a vector of

coefficients which describe the sinusoidal wave forms that constitute the original signal. Accordingly the transformed vector representations, $\mathbf{x}' = \phi(\mathbf{x})$, $\mathbf{y}' = \phi(\mathbf{y})$ and $\mathbf{z}' = \phi(\mathbf{z})$, are obtained for each axis of a given instance. Additionally we derive a further magnitude vector, $\mathbf{m} = \{m_{i1}, ..., m_{il}\}$ of the accelerometer data for each instance as a separate axis, where m_{ij} is defined in equation 1.

$$m_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2} \tag{1}$$

As with $\mathbf{x}'$, $\mathbf{y}'$ and $\mathbf{z}'$, we also apply ϕ to $\mathbf{m}$ to obtain $\mathbf{m}' = \phi(\mathbf{m})$. This means that our representation of a training instance consists of the pair $(\{\mathbf{x}', \mathbf{y}', \mathbf{z}', \mathbf{m}'\}, c)$, where c is the corresponding activity class label. Note that the coefficients returned after applying ϕ are combinations of negative and positive real values for DCT, and a combination of real and complex values for FFT. For the purpose of feature representation, we are only interested in the magnitude of the coefficients, irrespective of (positive or negative) sign. Accordingly for each coefficient e.g. x_{ij}', we maintain its absolute value $|x_{ij}'|$ and for complex coefficients, we convert them to real numbers. In our approach we retain a subset of the l coefficients. The final feature representation is obtained by concatenating the absolute values of the first l coefficients of $\mathbf{x}'$, $\mathbf{y}'$, $\mathbf{z}'$ and $\mathbf{m}'$ to produce a single feature vector of length $4xl$. An illustration of this feature selection and concatenation appears in Figure 1.
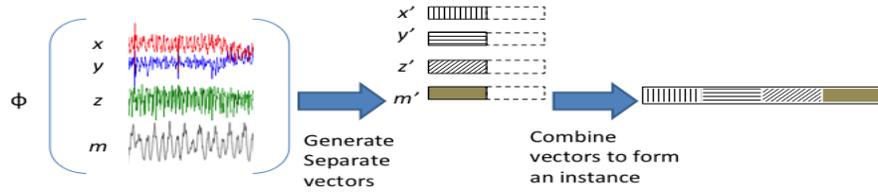
Fig. 1. Feature extraction and vector generation using frequency transforms.

3.3 Deep Features

Deep learning approaches have been applied to the task of HAR due to their ability to extract features in an unsupervised manner. They are able to stack multiple layers of differentiable functions to create a hierarchy of increasingly more abstract features [10]. Recent applications have used more of CNNs due to their ability to model local dependencies that may exist between adjacent data points in the accelerometer data [21]. CNNs are a type of Deep Neural Network that is able to extract increasingly more abstract feature representations by passing the input data through a stack of multiple convolutional layers [10], where each layer in the stack takes as input, the output of the previous layer of convolutional operators. An example of a CNN is shown in Figure 2.
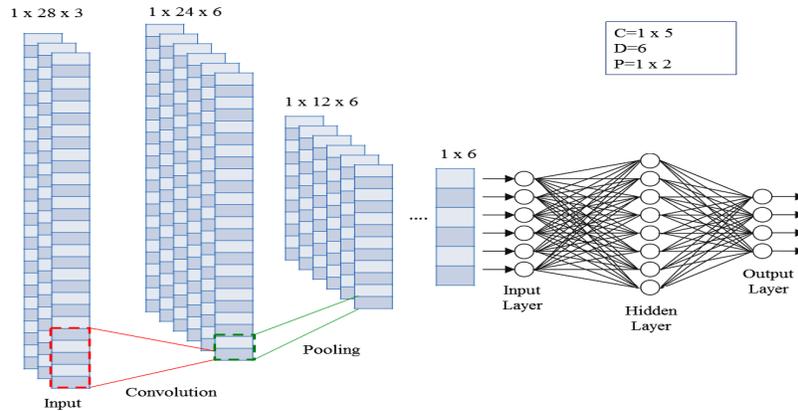
Fig. 2. Illustration of CNN

The input into the CNN in Figure 2 is a 3-dimensional matrix representation with dimensions $1 \times 28 \times 3$ representing the width, length and depth respectively. Tri-axial acceleromter data typically have a width of 1, a length l and a depth of 3 representing the x, y and z axes. A convolution operation is then applied by passing a convolution filter over the input which exploits local relationships between adjacent data points. This operation is defined by two parameters, D representing the number of convolution filters to apply and C, the dimensions of each filter. For this example, $D = 6$ and $C = 1 \times 5$. The output of the convolution operation is a matrix with dimensions $1 \times 24 \times 6$, these dimensions being determined by the dimension of the input and the parameters of the convolution operation applied. This output is then passed through a Pooling operation which basically performs dimensionality reduction. The parameter P determines the dimensions of the pooling operator which in this example is 1×2, which results in a reduction of the width of its input by half. The output of the pooling layer can be passed through additional Convolution and Pooling layers. The output of the final Pooling layer is then flattened into a 1-dimensional representation and then fed into a fully connected neural network. The entire network (including convolution layers) is trained through back propagation over a number of generations until some convergence criteria is reached. Detailed description of CNNs can be obtained in [10].

Note that once the CNN is fully trained, it can be used to provide feature representations for use with other types of classifiers e.g. SVM, or kNN. This is achieved by cutting off the trained network after the final pooling layer and just before the fully-connected neural network. Each training example is then passed through the convolutional network in order to obtain an abstract representation which is used to train the SVM or kNN classifier. A similar operation is performed for each test example to obtain an abstract representation which is passed to the SVM or kNN for classification. We refer to this as a CNN hybrid approach.

4 Evaluation

A custom dataset was created to reflect the requirements of selfBACK. A group of 34 volunteer participants were used for data collection. The age range of participants is 18 to 54 years and the gender distribution is 52% Female and 48% Male. Data collection captured data for the six activities shown in Table 2.

Table 2. Details of activity classes in our dataset.

Activity	Description
Walking	Walking at normal pace
Jogging	Jogging on a treadmill at moderate speed
Up Stairs	Walking up 4 - 6 flights of stairs
Down Stairs	Walking down 4 - 6 a flights of stairs
Standing	Standing relatively still
Sitting	Sitting still with hands on desk or thighs

This set of activities was selected by the health partners of selfBACK as the set of activities of normal daily living that they would like to monitor for the management of LBP. Data was collected using the Axivity Ax3 tri-axial accelerometer [4], at a sampling rate of 100Hz with a range of ± 4g. Accelerometers were mounted on the right-hand wrist and thigh of the participants using specially designed wristbands provided by Axivity for the wrist, and tape for the thigh. The activities have a balanced distribution between classes, as the participants were asked to carry out each activity for the same period of time (3 minutes). Each dataset consists of a set of accelerometer readings for each user on each activity, with each reading containing a timestamp and the x, y, z-axis readings for a specific poll of the accelerometer sensor. The wrist and thigh dataset are available on GitHub [5] as an open source dataset.

4.1 Experiment Design

Experiments are reported using our dataset of 34 users. Evaluations are conducted using a leave-one-person-out methodology in which each user is selected in turn for testing and the remaining 33 are used for training. In this way, we are testing the general applicability of the system to users whose data is not included in the trained model. Performance is reported using micro-averaged F1. We compare 7 different feature representations for both wrist and thigh mounted sensors. The 7 representations we use consist of 4 shallow feature representations and 3 deep feature representations:
Shallow Features use a Support Vector Machine (SVM) for classification as it provides state-of-the-art performance and was found to outperform kNN and Naive Bayes in experimentation. The number of coefficients for the frequency domain representations

---

[4] http://axivity.com/product/ax3
[5] https://github.com/rgu-selfback/activity-recognition

was also selected by initial experimentation. The hand-crafted and frequency domain representations reported are:

- *Time*: hand-crafted time domain features, detailed in Table 1;
- *Freq*: hand-crafted frequency domain features, detailed in Table 1;
- *FFT*: applies Fast Fourier Transform and feeds the first 80 coefficients from the x, y, z axis along with the magnitude to an SVM classifier;
- *DCT*: applies Discreet Cosine Transform and feeds the first 80 coefficients from the x, y, z axis along with the magnitude to an SVM classifier;

Deep Features have a vector of accelerometer values as an input into the CNN. After experimenting with different parameter settings, the final configuration used for the wrist data is 5 convolution layers with 150, 100, 80, 60, and 40 convolution filters respectively, created with a filter of size 10 and pooling size of 2. The last pooling layer is connected to a fully connected network with 2 hidden layers, with 900 units in the first layer and 300 units in the second layer and dropout probability of 0.5, and a final output layer with 6 units representing the 6 activity classes in our dataset and softmax regression. A shallower configuration is used for the thigh data with 3 convolution layers with 150, 100, and 80 filters respectively. For the deep features, we use both a neural network for classification as well as the CNN hybrid approach with traditional classifiers. For the hybrid approach, we cut off the network, after it has been trained, at the last pooling layer and feed the output of this as vector to the SVM or kNN classifier. The purpose of this is to compare the performance of the representation learned by the CNN using the same classifier as was used for the other representations. The deep features reported include a standard CNN and two hybrid approaches:

- *CNN*: uses CNN features with a Neural Network classifier;
- *CNN-SVM*: a CNN hybrid approach using SVM as the classifier with features extracted from the CNN; and
- *CNN-kNN*: a further CNN hybrid approach using kNN as the classifier with features extracted from the CNN.

4.2 Results

Wrist: The first set of results we present is a comparison of the F1 results for the different representations with the accelerometer positioned on the wrist, as shown in Figure 3. The best results are achieved using the deep features, second best using frequency transform features, with the hand-crafted features performing poorly. Within the deep features, the hybrid approaches of *CNN-SVM* and *CNN-kNN* deliver the highest F1 scores at 0.850 and 0.845 respectively, slightly outperforming *CNN* which has a score of 0.839 although the difference is not significant[6]. *CNN-SVM* significantly outperform both *FFT* and the hand-crafted features. Among the remaining features *DCT* performs well with an F1 score of 0.836 which significantly outperforms *FFT* (0.797), *Time* (0.769) and *Freq* (0.705). Generally data collected from the wrist tends to be more noisy than data collected from the thigh; the deep features with the 5-layer architecture

---

[6] significance is tested with a two-tailed student's t-test at p=0.05
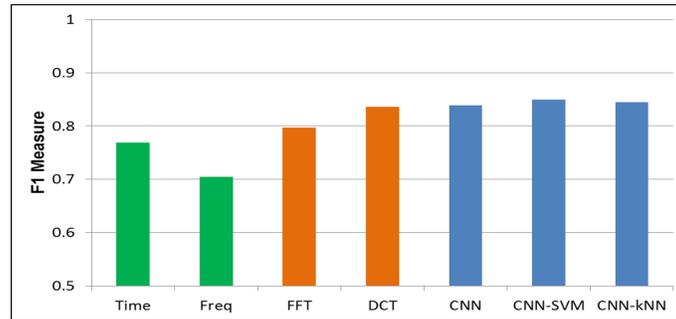
Fig. 3. F1 results for different feature representations for the wrist

appear to be more effective at building noise-resistant models. Similarly *DCT*, which disregard higher frequency coefficients, may be more effective than other shallow features at eliminating noise.

Thigh: The second set of results compares the F1 scores for the accelerometer positioned on the thigh (see Figure 4). The best result is achieved by *DCT* with an F1 score of 0.967, which is slightly better than achieved by the deep learning features of *CNN*, *CNN-SVM* and *CNN-kNN* with scores of 0.959, 0.957 and 0.949 respectively. The differences are not significant, however, *DCT* is significantly better than the other shallow features. Among the deep features CNN and CNN-SVM are on a par, both slightly outperforming CNN-kNN, and significantly better than FFT (0.927), Time (0.936), and Freq (0.820). It is interesting that Time outperforms FFT for this location and may explain hand-crafted time domain features are a popular choice for HAR with data collected from the thigh.

Discussion: HAR with data collected from the thigh for the 6 classes evaluated here is excellent (0.967 for DCT) and superior to collection from the wrist by over 11%. This difference reflects that HAR from the thigh is an easier problem because there is less variation in the data; typically less noise; and data from the thigh is better able to identify leg orientation to distinguish between standing and sitting. However, the 0.850 obtained with deep features is still good performance, particularly if distinguishing between standing/sitting is less important than between active/inactive classes for selfBACK. Wrist movement alone appears insufficient to differentiate between standing/sitting. There is still a clear trade-off between the accuracy of classification from the thigh and the ease of wear on the wrist.

Deep feature representations, particularly using the hybrid approach, perform well on both locations. They provide the best results with data from the wrist; the deeper architecture seems particularly effective at modelling the more complex and noisy data associated with the wrist. The hybrid combination (CNN-SVM) appears to take advantage of both the CNNs ability to learn effective invariant features and SVMs ability to generate good decision boundaries. Deep features also have an efficiency advantage in real applications because, once the model has been built they take raw accelerometer values as inputs.
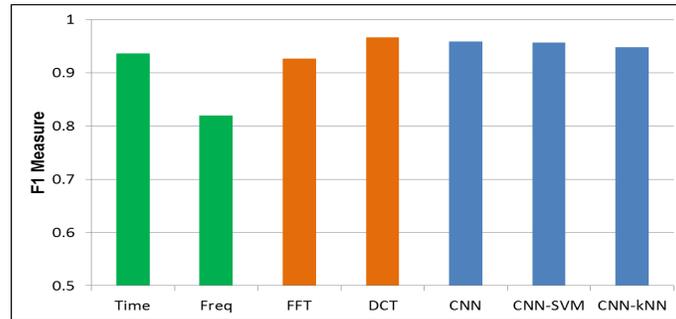
Fig. 4. F1 results for different feature representations for the thigh

Shallow features generally have mixed results. However, DCT performs well on both locations, perhaps because frequency models are good at distinguishing between different active activity classes and by selecting lower frequency coefficients DCT is good at filtering out noise. DCT is also attractive for real time applications because by directly using coefficients obtained from the raw accelerometer data as features extensive pre-processing is not required. With hand-crafted features, *Time* performs adequately on the thigh data but is not competitive for wrist data.

5 Conclusion

In this paper, we have presented an analysis of deep and shallow feature representations for accelerometer data on HAR with data collected from the wrist and thigh. Three types of representations are considered: hand-crafted, frequency transform and deep features which includes two hybrid approaches. For the purpose of selfBACK, it is important to determine which representation approach is best suited for HAR using a single wrist-mounted accelerometer on the required activity classes. However, wrist data is more prone to random noise due to the variation in movement and posture of the hand. We evaluate the difference in classification performance with data collected form both locations and find data from the thigh to outperform data from the wrist by 11%. However, classification with data from the wrist still achieves good performance with an F1 score reaching 85%. Comparative evaluation of the seven representation approaches shows deep features to work best on data form the wrist. In particular CNN-SVM, a hybrid approach, that uses CNN to learn features for an SVM classifier outperformed both hand-crafted and frequency transform by a margin of over 1.4%.

References

1. O. Airaksinen, J. Brox, C. o. Cedraschi, J. Hildebrandt, J. Klaber-Moffett, F. Kovacs, A. Mannion, S. Reis, J. Staal, H. Ursin, et al. Chapter 4 european guidelines for the management of chronic nonspecific low back pain. *European spine journal*, 15:s192–s300, 2006.
2. D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *ESANN*, 2013.

3. K. Bach, T. Szczepanski, A. Aamodt, OE. Gundersen, and P. J. Mork. Case Representation and similarity assessment in the selfBACK decision support system. In *Proc of Int Conf on Case-Based Reasoning)*, pages 3246. Springer, 2016.

4. L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive computing*, pages 1–17. Springer, 2004.

5. D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Comp*, 14(7):645–662, 2010.

6. L. Gao, A. Bourke, and J. Nelson. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical eng & physics*, 36(6):779–785, 2014.

7. N. Y. Hammerla, S. Halloran, and T. Ploetz. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proc of the 25th Int Joint Conf on AI*, 2016.

8. F. J. Huang, and Y. Lecun. Large-scale learning with SVM and Convolutional for generic object categorization. In *Proc IEEE Comp Soc Conf on Computer Vision and Pattern Recognition*, pages 284291, 2016.

9. O. D. Lara and M. A. Labrador. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, 15(3):1192–1209, 2013.

10. Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 255–258. MIT Press, 1998.

11. A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise*, 45(11):2193, 2013.

12. J. Mäntyjärvi, J. Himberg, and T. Seppänen. Recognizing human motion with multiple acceleration sensors. In *2001 IEEE Int Conf on Systems, Man, and Cybernetics*, volume 2, pages 747–752. IEEE, 2001.

13. U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *BSN Int Workshop on Wearable and Implantable Body Sensor Networks, 2006.* . IEEE, 2006.

14. T. Plötz, N. Y. Hammerla, and P. Olivier. Feature learning for activity recognition in ubiquitous computing. In *Proc of the 22nd Int Joint Conf on Artificial Intelligence*, pages 1729–1734. AAAI Press, 2011.

15. D. Ravi, C. Wong, B. Lo, and G. Z. Yang. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, 21(1):56–64, 2017.

16. C. A. Ronao and S.-B. Cho. Deep convolutional neural networks for human activity recognition with smartphone sensors. In *22nd Intl Conf on Neural Information Processing*, pages 46–53. Springer, 2015.

17. S. Sani, N. Wiratunga, S. Massie, and K. Cooper. Selfback-activity recognition for self-management of low back pain. In *Research and Development in Intelligent Systems XXXIII*, pages 281–294. Springer, 2016.

18. S. Sani, N. Wiratunga, S. Massie, and K. Cooper. kNN Sampling for Personalised Human ActivityRecognition. In *Proc of 25th Int Conf on Case-Based Reasoning*. Springer, 2017

19. M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga. Fusion of smartphone motion sensors for physical activity recognition. *Sensors*, 14(6):10146–10176, 2014.

20. E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Proc of 11th IEEE Int Symp on Wearable Computers*, pages 37–40. 2007.

21. M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *Proc of 6th Int Conf on Mobile Computing, Applications and Services*, pages 197–205. 2014.