ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/333662263

Improving CNN-based activity recognition by data augmentation and transfer learning

Conference Paper · June 2019

CITATIONS
0

READS
119

3 authors, including:

Evangelia I Zacharaki
University of Patras
82 PUBLICATIONS **1,391** CITATIONS
SEE PROFILE

Vasileios Megalooikonomou
University of Patras
245 PUBLICATIONS **1,811** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Frailsafe View project

All content following this page was uploaded by Evangelia I Zacharaki on 08 June 2019.

The user has requested enhancement of the downloaded file.

Improving CNN-based activity recognition by data augmentation and transfer learning

Gerasimos Kalouris
MDKAM Lab, Dep. of Computer Engineering and Informatics
University of Patras
Patras, Greece
kalouris@ceid.upatras.gr

Evangelia I. Zacharaki
VVR Group, Dep. Electrical and Computer Engineering
University of Patras
Patras, Greece
ezachar@upatras.gr

Vasileios Megalooikonomou
MDKAM Lab, Dep. of Computer Engineering and Informatics
University of Patras
Patras, Greece
vasilis@ceid.upatras.gr

***Abstract*—Activity classification is a challenging problem due to large signal dimensionality, high intra- and inter-subject variability in activity patterns, presence of transitional classes showing mixture of patterns, and dominance of the null class. Supervised learning has been the prevalent choice with deep neural networks (DNNs) showing some promising potential. Deep learning however requires a large number of labeled samples which is difficult to acquire, especially from vulnerable older people. In this paper we implement 3 different convolutional neural network architectures trained on data from older people, incorporating Bayesian optimization for efficient hyper-parameter tuning. We exploit various augmentation methods for time-series to make invariant predictions and also cross-utilize knowledge about physical activity of younger persons in order to improve generalization in our models designed for older adults.**

Keywords— activity recognition, deep learning, convolutional neural networks, transfer learning, data augmentation

I. INTRODUCTION

The assessment of the daily physical activity can provide insight on people's health condition and allow early detection of possible deterioration that can lead to undesired events, such as falls. Especially for the older population, adverse events increase the risk for disability, hospitalization, loss of autonomy and mortality. Therefore, a lot of interest has been shown in the last years for the development of wearable devices or smart environments monitoring activity [1]. Although many frameworks have been reported in the literature for activity recognition of older people, most of them have been tested on data from young and healthy participants [2, 3, 4], or the experiments were performed on laboratory conditions, e.g. in [5] a scaled model of a house was used along with a simulated sequence of activities. Those works report high classification accuracy, but results are not directly comparable with uncontrolled monitoring systems in real home environments.

The classification of time series has long time been explored with traditional methods performing either feature-based classification based on predefined feature extraction [2] [6] [7] [8] [9], embedding [10] and decomposition techniques [11], or evaluating similarity after some elastic matching that reduces time-scale variations, such as the dynamic time warping [12] [13]. With the domination of deep neural networks and the evidence that feature or metric learning approaches usually perform better than traditional techniques in many computer vision and text or speech recognition applications, lots of interest has been shown also for the use of deep learning architectures, such as convolutional neural networks (CNNs) [14] in activity recognition [15, 16, 17, 18]. CNNs were used for recognition of specific activities (actions in a kitchen) [15] monitored by a variety of body-worn, object-based and ambient sensors [19, 20], as well as for human activity recognition using data from smartphone sensors [21]. In [16] a shallow CNN was used with a weight sharing technique, on accelerometer signals showing improved classification over previous techniques, such as principal component analysis based on empirical cumulative distribution estimation and K-nearest neighbor classification. Evaluation was based on recordings collected using a cell phone in a controlled laboratory environment [22]. Deep 2D CNN were implemented in [18] by stacking the raw signals row-by-row such that every signal sequence becomes adjacent to every other sequence and extracting the magnitude of the 2D Discrete Fourier Transform of the created 2D image. Combination of convolutional and long short-term memory (LSTM) layers were used to model the temporal dependencies for multimodal activity recognition [17] [23]. A review on different deep, convolutional, and recurrent network architectures based on physiological activity signals recorded by wearable sensors can be found in [24], while extensive discussion on feature representation and learning approaches for human activity recognition systems can be found in the recent review [25].

The success of deep learning relies largely on the availability of large data amounts that allow to efficiently learn the multitude of network parameters without overfitting. When data are sparse, preprocessing techniques can be applied, such as data augmentation, to synthetically increase the variation of training samples, or transfer learning, to abridge the model complexity by constraining part of the parameters based on solutions found using other datasets from related domains.

Data augmentation is the process of simulating new data instances that maintain the correct labels, in order to increase the sample size when limited labeled data are available. It is more common in deep learning where the sample size is critical for model generalization. The aim is to cover different views of the

This research was partially supported by the FrailSafe Project (H2020-PHC-21-2015 - 690140) "Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions", co-funded by the European Commission under the Horizon 2020 research and innovation program. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used in this study and also thank the SigOpt team for the provided services.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

same target in order to make invariant predictions. Data augmentation usually relies on linear transformations in the spatial domain and has mainly been performed for image recognition. Label-preserving augmentation for time-series however is more challenging since the influence of any transformation is difficult to be determined without profound domain knowledge. Augmentation [26] [27] [28] is typically performed in data space or, as shown more recently, in feature space. Methods that are based on manifold learning for data synthesis [26] require the availability of an external corpus of annotated data, such as training samples with pose annotations. In the time series domain this is more rarely the case, therefore data augmentation usually relies on signal perturbations [26], warping or feature mapping techniques [27].

Transfer learning has also been proposed as semi-supervised technique to address the problem of limited training data. It aims to address settings in which training and test data come from different distributions. A transfer learning methodology for activity recognition without using new labeled data has previously been proposed in [29]. This approach aimed to exploit observations acquired from an old sensor for which trained models were available. A multi-view learning technique was presented which achieved high accuracy. The authors in [30] applied transfer learning to build personalized affective models without labeled target data. The latter is accomplished either by exploiting shared structure underlying source and target subjects, or by training multiple classifiers on source subjects and transfer the classification parameters to the target subjects. A semi-supervised clustering methodology is proposed in [31] for physical activity recognition. The approach is able to capture potential shifts in the subject's behavior, such as falls, while requiring a small number of labeled data.

In this paper, we propose a deep learning approach for recognizing activity of older people based on recordings from wearable sensors. Building upon previous work [32], which used only a small number of labeled recordings acquired under challenging conditions (mainly due to data variations and erroneous measurements), we now exploit data augmentation and domain adaptation techniques to make predictions more robust. Three advanced CNN architectures were implemented, each of them based on different underlying assumptions (such as existence of correlation across sensors or correlation across axes within the same sensor). Transfer learning techniques were applied based on available databases from younger population groups and compared against the standard random initialization of convolutional kernels. Moreover, we investigated various augmentation techniques for recordings from wearable sensors, as described in [28]. More details on the implemented deep network architectures are provided in Section II-A, while the data augmentation and transfer learning techniques are described the Sections II-B and II-C, respectively.

II. Methods

The purpose of the classification method was to discriminate basic dynamic movement or static activities, such as walking in several directions, sitting, standing, laying, using physiological signals of the senior adults acquired by wearable devices [33]. The current study is based only on movement and posture information, monitored by three 3-axial sensors (accelerometer, gyroscope, magnetometer), as described with more details in Section III. The data collection was part of the FrailSafe European research project [34] that aims at the development of a technological platform for early detection and prevention of frailty based on non-intrusive sensing (of physical, cognitive, psychological, social domains) for the ageing population [35] [36].

A. CNN architectures

We studied the performance of convolutional neural network architectures in recognizing movement patterns of older people. The aim was to examine an alternative way of activity classification that circumvents the need for dedicated feature extraction. We investigated three CNNs and customized them using advanced Bayesian optimization techniques for discriminating the basic movement patterns of senior adults. The sliding window technique was applied to extract overlapping time windows of the multi-channel recordings. Depending on the dimensionality of the convolutional filter and the multi-dimensional signal representation, various architectures could be defined that process the multi-channel information in a different way, thereby affecting the networks' performance. The common aspect in all three architectures is the extraction of translation-invariant patterns by applying the processing units of the CNNs along the temporal dimension and by sharing the units among multiple sensors [15].

The time window used to extract the samples for classification should be large enough to contain sufficient information for the discrimination of the activity, but not too large in order to avoid mixture of activities within its time span, and also to keep the latency (buffering time) of real-time recognition in data streams limited. The amount of overlap between the extracted time windows also affects the performance. High overlap corresponds to dense sampling allowing to continuously monitor the transition of activities and reducing the likelihood to miss short activity patterns or fail to recognize them due to partial effects. However, high overlap increases the computational cost significantly without any substantial benefit as a result of processing related instances. For each experiment a time window of 68 time points (corresponding to $2.72sec$) with 50% overlap was used, in accordance to previous work [3]. Since multiple activity patterns can be present in any sliding window, the label for each instance is determined by the class majority.

The three deep network architectures are relatively shallow since the available training samples were not enough to efficiently train deep architectures without jeopardizing their generalization ability. The layers of a CNN have neurons arranged in three dimensions, that are denoted as *width*, *height* and *depth* ($W \times H \times D$). For all three networks the *width* corresponds to the temporal dimension, therefore $W = 68$ in the input layer. The main difference of the networks is reflected in the arrangement of the other two dimensions, which affects the number of computed feature maps, as detailed next. It should also be noted that the number of sensors (N) might vary across experiments according to the availability of magnetometer recordings, as described later in Section II-C. In all cases however, the same sensors were used across the different architectures (CNN1, CNN2, CNN3).

CNN1: 1D convolution is performed on the input data along the temporal dimension, with a convolutional kernel of size 5×1. All channels corresponding to recordings of the N tri-axial sensors, are arranged in the depth dimension, therefore $D = 3 \cdot N$.

CNN2: 2D convolution is performed with a $5 \times N$ convolutional kernel along the temporal dimension and along the sensing modality by stacking the N sensors (accelerometer, gyroscope, and optionally magnetometer) row-by-row and arranging the x, y and z axes in the depth dimension ($D = 3$). Since the height of the input data equals the height of the convolutional kernel ($H = N$), the 2D kernel slides only along the temporal dimension.

CNN3: Following the idea of Jiang and Yin [18], we created a 2D signal by stacking the input channels row-by-row with repetition, such that every sensor becomes adjacent to every other sensor. Specifically, in the case of 3 sensors we arranged the recordings of the accelerometer, gyroscope and magnetometer in x, y, z, and introduced again the accelerometer in x, y, z, thereby creating a 2D signal of height $H = 12$. By using a 5×6 convolutional kernel, all different sensor combinations were possible (accelerometer with gyroscope, gyroscope with magnetometer and magnetometer with accelerometer). The convolutional kernel this time slides along both axes (over time and over sensors). By using a stride of 1×3 the bundles of x, y, and z channels were kept together. The depth dimension is vanished in this architecture ($D = 1$).

The conceptual architecture is illustrated in Fig. 1 and consist of three convolutional layers. The first two convolutional layers are followed by a max pooling layer to reduce input dimensionality, while each layer is followed by a ReLU activation unit (not shown in Fig. 1 due to space limitations) and a batch normalization layer. Dropout is applied before the final (fully connected) layer. Details on parameters used in each layer, such as number of filters, can be found in [32].

As an extension to these baseline architectures [32], in this work we incorporate data augmentation to synthetically increase the data instances, and transfer learning techniques to restrict the model parameters, aiming to reduce overfitting. The implemented techniques affect only the training phase, while inference is performed on the original (unchanged) test set to assess the recognition performance.

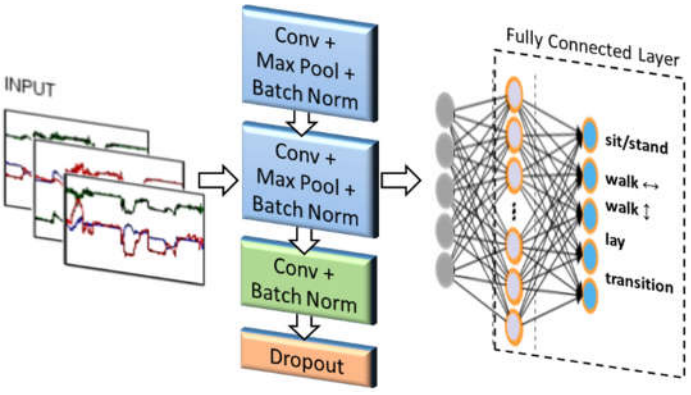


Fig. 1. General architecture for the 3 implemented CNNs. Each convolutional layer (*Conv*) is followed by a batch normalization layer (*Batch Norm*) and a rectified linear unit (ReLU) activation unit, which is not illustrated in the figure due to space limitations. Max pooling (*Max Pool*) is performed only on the first 2 blocks. Detailed parameters for each individual CNN can be found in [32].

B. *Data augmentation*

We implemented data transformation techniques mainly as described in [28], but also our own, in order to capture variation that was possibly not originally sampled. We excluded techniques that modify the window length or perform subsampling or cropping of the instances. The following transformations were applied:

1) *Rotate*: The order of channels is permute randomly to make the model more robust to possible differences in sensor placement between participants.
2) *3D-Rotate*: Rotation is performed around the anterior-posterior axis (which is prone to sensor misplacement) preserving the relative position of channels. This is a special case of the *Rotate* transformation, suitable to the FrailSafe sensors which are located at the sternum.
3) *Scale*: Scaling of the signal is performed by multiplication with a random scalar to simulate multiplicative noise in the signals.
4) *Jitter*: Gaussian noise is added to the signals to make predictions more resistant to additive sensor noise that is common in wearable sensors.
5) *Permute*: Permutation changes the order of small scale patterns inside the time window and is performed by splitting the window in 2-10 equal-sized segments and re-ordering them. This technique is more suited to participants having tremor that can cause fluctuating signals.

The above techniques can be applied either individually or in combination. In [28] the best results were obtained by a combination of transformations. Our experiments with several combinations showed that the two combinations *3D-Rotate & Scale*, and *Rotate & Permute* performed best for the FrailSafe dataset. Accordingly, we created 800 new instances for each one of the two combinations, thus augmenting the training set with 1600 new samples.

C. *Knowledge transfer from younger population groups*

Transfer learning leverages knowledge from one or multiple domains to improve predictions on new, related tasks. Our aim was to exploit available human activity recordings of accelerometer, gyroscope and magnetometer sensors, that have been annotated as part of other studies. For this purpose, we exploited two datasets, PAMAP2 [6] and UCI HAR [7]. The PAMAP2 dataset is collected from 9 young subjects (1 female, 8 male) with average age 27.2±3.3 years old. Three Inertial Measurement Units (IMUs) sensory data containing 3-axial accelerometer, 3-axial gyroscope and 3-axial magnetometer, were placed on the chest, and had a sampling frequency of 100 Hz. The majority of the activities had a duration of 3 minutes except of the most intense ones, like ascending and descending stairs which lasted only 1 minute, while a break was introduced in regular intervals.

The UCI HAR dataset has been acquired from 30 volunteers within an age bracket of 19-48 years while carrying a waist-

mounted smartphone with embedded inertial sensors (3-axial accelerometer and 3-axial gyroscope) operating at a constant rate of 50 Hz. Since this dataset does not include a magnetometer, the CNN architectures used for training and fine-tuning during the transfer learning experiments included only the accelerometer and gyroscope's channels. Each person performed six activities, same with the ones studied in this paper. The class distribution for both datasets is illustrated in Fig. 2.

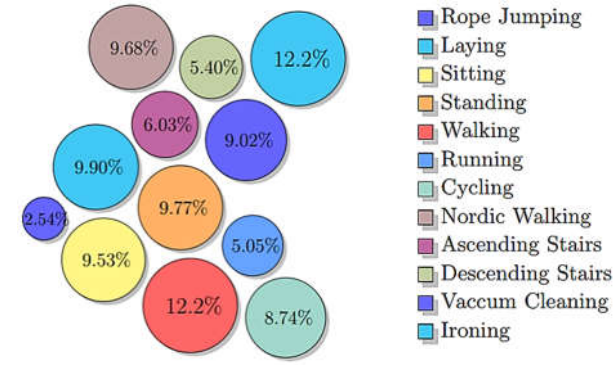
D. *Implementation details*

Bayesian optimization lately became very popular for tuning the hyperparameters of deep networks [37] due to its potential to handle multi-parametric problems with costly objective functions when first- or second-order derivatives are not available. We optimized the networks hyperparameters by the Bayesian hyperparameter optimization platform SigOpt [38], which is a standardized and scalable platform, accompanied by an API facilitating the generation of well performing models. It also allows parallelization for faster evaluation. The CNNs were implemented in TensorFlow [39] using Keras [40] and CUDA. The experiments were executed on an Intel(R) Xeon(R) @ 3.70GHz processor with 8GB RAM and a GPU (NVIDIA(R) Titan Xp 12GB VRAM).

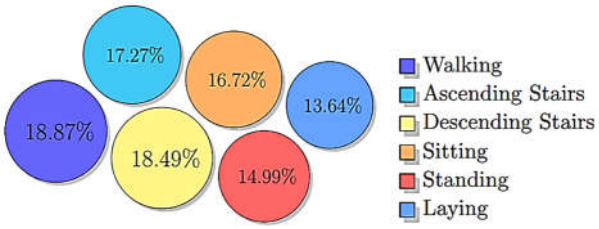
III. Data Description

The recordings were obtained from older people (age: 70-92 years) who participated in the FrailSafe project [34] using a WBAN (Wireless Body Area Network) system developed by SMARTEX [33]. The WBAN is composed by a sensorized garment collecting a variety of physiological signals, an electronic device and a software tool. The participants were monitored at home during standard day-time activities. For activity recognition we used only the signals from the 3-axial accelerometer, 3-axial gyroscope and 3-axial magnetometer. In cases of possible sensor misplacement which caused rotation of axes, the channels were accordingly interchanged. The reference orientation was defined by the recordings of the subjects selected for building the classification models (training subjects) [32]. The sampling rate of the recordings was at 25 Hz. To ensure consistency of all incorporated datasets, the PAMAP2 and UCI HAR were resampled to this rate too. Also, the measurement units of the FrailSafe and UCI HAR sensor recordings were converted to coincide with the units of PAMAP2.

For the collection of the annotated data some of the project's participants (the less frail) volunteered to perform a set of activities while wearing the FrailSafe's wearable device. The protocol for data annotation was performed in clinical centers of three different countries (Greece, France, Cyprus) to capture the data variability and included the following actions: 1 *min* standing, 1 *min* sitting, 1 *min* walking, then ascending/descending stairs for 30 *sec* each, and finally laying for 30 *sec*. Each participant performed all activities except of ascending/descending stair which was only performed if stairs were accessible by the older person in his/her residence. The timing was recorded by medical instructors while the participant followed the protocol. After data collection, the time intervals corresponding to the first five seconds of the beginning of each



(a) PAMAP2



(b) UCI HAR

Fig. 2. Distribution of samples across classes for the PAMAP2 dataset (top) and the UCI HAR dataset (bottom) used for transfer learning. Only the common (with the FrailSafe dataset) subset of activities was used.

activity were labeled as *transition* state, to indicate the time required to switch between two different activities. The distribution of the different activities in the FrailSafe annotated data is illustrated in Fig. 3, where the NULL class represents the transition state.

For the experiments, although six activity classes were initially defined, the classes *sitting* and *standing* were merged into one class, as well as *ascending stairs* and *descending stairs*. This was performed based on previously reported work that suggests that these classes are not easily separable [3]. To that end, the final classes were: *sitting/standing*, *walking*, *ascending/descending stairs*, *laying*, *transition*.

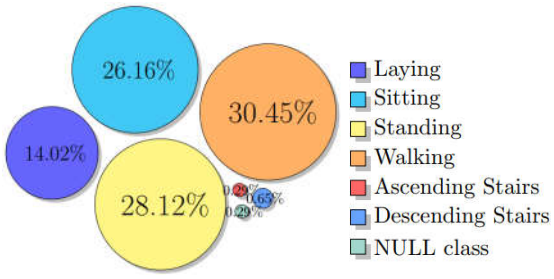


Fig. 3. Samples distribution across activities for the FrailSafe data acquired from older people.

IV. Results

The three architectures were assessed on a dataset of 23 older subjects. Evaluation of the methods was based on data-split, i.e. we randomly selected 16 of them for training, in order to approximate the deep learning requirements of large number of

annotations, and 7 subjects for testing. The hyper-parameters were tuned by performing 4-fold stratified cross-validation on the training set using Bayesian optimization. For each fold, the optimized CNN was then assessed on the independent test set. The obtained hyper-parameter values and their relative contribution in each classification model can be found in [32] for the baseline architectures, and in TABLE I after data augmentation.

For the transfer learning experiments, the hyper-parameters for each CNN were optimized using the PAMAP2 and UCI HAR recordings, as auxiliary datasets. Since these datasets are similar to the target dataset (FrailSafe), we used the convolutional networks as fixed feature extractors, as in [41]. For CNN1 the best results were obtained when only the fully connected layer and output layer were trained on PAMAP2 and UCI HAR datasets, whereas for CNN2 and CNN3 the transfer learning performance was poor. The tuned parameters using transfer learning with each of the two incorporated datasets are presented in TABLE II. An estimate of the methods' performance can be obtained by the percentage of (re)trained parameters shown in TABLE II, for which lower values are better, indicating higher utilization of the acquired weights (features).

All results obtained from the baseline architectures (with random initialization on the FrailSafe dataset) and the incorporated techniques are included in TABLE III. It can be observed that the difference in performance for the three investigated architectures is small, but the highest test accuracy (84.89%) is observed for CNN1 using transfer learning from PAMAP2 dataset. Moreover, the deviation of the results across folds reduces, implying that the additional information helps the stability of predictions.

The average (across folds) confusion matrices showing the sensitivity of the networks for each individual class are illustrated in Fig. 4. It can be observed that the all networks fail in detecting the transitional class and also in differentiating walking in stairs from walking on the floor. The former might be explained by the variability of transitional patterns which actually form more than one cluster, while the latter was expected due to the very small number of samples (only 2 subjects were available; one was used for training and one for testing). The intermix of walking classes might also be attributed to the commonalities of the (forward/backward and upstairs/downstairs) walking patterns of the elderly, such as slow speed and possible instability.

TABLE I. OPTIMIZED HYPER-PARAMETERS USING SIGOPT FOR THE FRAILSAFE DATASET (*SGD*: STOCHASTIC GRADIENT DESCENT; m= 10^{-3})

	Hyper-parameters	**Values**			**Contribution (%) in the model**		
		CNN1	***CNN2***	***CNN3***	***CNN1***	***CNN2***	***CNN3***
FrailSafe dataset (N=3)	Batch	42	100	71	1.70	1.71	3.73
	Dense Layer Size	497	1000	750	1.37	1.70	2.07
	Dropout prob.	0.58	0.54	0.60	2.21	1.69	2.41
	Epochs	100	100	100	4.45	9.52	5.33
	Filter 1	43	100	63	1.82	1.62	2.27
	Filter 2	74	91	100	2.41	1.54	1.65
	Filter 3	100	48	86	1.69	1.45	2.09
	Learning rate	0.072	0.7 m	0.061	9.76	31.91	14.10
	Regulariz. rate	0.1 m	0.1 m	0.1 m	21.67	21.56	16.22
	Optimizer	*SGD*	*Adam*	*SGD*	52.92	27.29	50.13

TABLE II. OPTIMIZED HYPER-PARAMETERS USING SIGOPT FOR PAMAP2 AND UCI HAR DATASETS INDIVIDUALLY AND ALSO AFTER TRANSFER TO FRAILSAFE (*SGD*: STOCHASTIC GRADIENT DESCENT; *RMS*: RMSPROP ALGORITHM; m= 10^{-3})

	Hyper-parameters	**Values**			**Contribution (%) in the model**		
		CNN1	***CNN2***	***CNN3***	***CNN1***	***CNN2***	***CNN3***
PAMAP2 dataset (N=3)	Batch	44	32	100	2.21	1.29	1.29
	Dense Layer Size	10	10	962	1.57	1.41	1.41
	Dropout prob.	0.6	0.6	0.33	1.84	1.87	1.87
	Epochs	100	100	100	5.46	4.23	4.23
	Filter 1	100	100	52	2.63	2.06	2.06
	Filter 2	100	100	100	2.09	1.98	1.98
	Filter 3	100	100	43	1.35	1.66	1.66
	Learning rate	0.1 m	0.1 m	0.037	23.92	46.24	26.61
	Regulariz. rate	3.7 m	0.1 m	0.1 m	21.00	9.32	21.00
	Optimizer	*RMS*	*RMS*	*SGD*	8.85	18.26	18.26
	Transfer from PAMAP2 to FrailSafe				**Trained parameters (%)**		
	Optimizer	*SGD*	*RMS*	*SGD*	9.56	9.87	79.93
	Learning Rate	0.03	0.1 m	0.03			
UCI HAR dataset (N=2)	Batch	32	84	32	1.32	0.87	4.38
	Dense Layer Size	10	10	51	1.36	0.90	1.31
	Dropout prob.	0.6	0	4.94	3.64	1.26	1.51
	Epochs	100	100	87	2.02	1.32	2.14
	Filter 1	100	10	100	1.30	1.31	1.34
	Filter 2	10	100	46	1.31	2.36	1.84
	Filter 3	10	94	35	1.29	1.78	1.12
	Learning rate	0.1 m	0.1 m	0.1 m	68.10	69.01	67.89
	Regulariz. rate	0.1 m	0.098	0.1 m	6.68	15.22	10.34
	Optimizer	*RMS*	*RMS*	*RMS*	12.98	5.97	8.12
	Transfer from UCI HAR to FrailSafe				**Trained parameters (%)**		
	Optimizer	*RMS*	*RMS*	*RMS*	13.35	98.21	94.68
	Learning Rate	0.003	0.3 m	0.003			

TABLE III. AVERAGE AND STANDARD DEVIATION OF TEST ACCURACY (%) OF THE FRAILSAFE (FS) DATASET OVER THE 4-FOLDS WITH OR WITHOUT MAGNETOMETER (N = 3 OR 2, RESPECTIVELY)

Technique	Accuracy (average ± standard deviation)		
	CNN1	*CNN2*	*CNN3*
Random Init. FS (N = 3)	81.91(±2.45)	78.49(±3.66)	**82.47**(±4.24)
Augmentation of FS (N = 3)	83.04(±2.08)	**82.71**(±2.34)	82.43(±1.80)
Transfer from PAMAP2 (N = 3)	**84.89**(±0.96)	81.97(±1.19)	81.34(±1.43)
Transfer from UCI-HAR (N = 2)	83.16(±1.52)	81.44(±4.15)	80.98(±5.40)

V. DISCUSSION

Overall, augmentation and transfer learning seem to boost performance of CNN1 and CNN2, while the performance of CNN3 deteriorates slightly in respect to class averages, but becomes more stable (smaller standard deviation) when the instances are augmented using the FrailSafe or PAMAP2 datasets. On the contrary, the UCI HAR dataset seems to add additional variation on CNN3, such that the network does not benefit in respect to neither performance nor robustness. To elaborate more on this, convolution across *height* and *width* dimensions allows the sharing of neurons in order to preserve statistical invariance, while along the *depth* dimension parameter sharing is not desired because data arranged in *depth* are expected to carry unique information. In CNN3 the *depth* dimension has vanished because we have arranged all channels in *height* dimension enforcing the sharing of weights across the different sensors (accelerometer etc). While preserving statistical invariance across sensors seems beneficial when handling purely the FrailSafe data (1st row of TABLE III), it is not the case when introducing data from other studies in which the measured physical quantities might have some deviations. Finally, the decreased contribution of the UCI HAR dataset along with the high standard deviation in predictions for CNN2 and CNN3 might be attributed to the fact that it does not contain recordings of magnetometer.

In comparison to previous work, the CNN1 architecture, optimized with transfer learning from PAMAP2, outperforms (with an accuracy of 84.89%) our previous method evaluated on the FrailSafe dataset [42], which achieved 81.7% accuracy using time-domain and spectral features and a support vector machines classifier. Furthermore, the accuracy is higher compared to all three baseline convolutional networks [32] (obtaining 81.91%, 78.49% and 82.47%, respectively), which were trained without incorporating data augmentation or transfer learning.

In respect to other studies, the classification accuracy reported in the original works presenting the PAMAP2 [6] and UCI HAR [7] datasets are higher, i.e. 82.4-89.2% for leave-one-subject-out 9-fold cross-validation in [6] and 96% in [7], evaluated however most possibly on data-split across time windows and not across subjects. For different type of activities, many works exist, such as the comparative study in [43] using four benchmark datasets (Ambient Kitchen 1.0, Darmstadt Daily Routines, Skoda Mini Checkpoint, Opportunity) with accelerometer recordings. Graphs illustrate an accuracy between

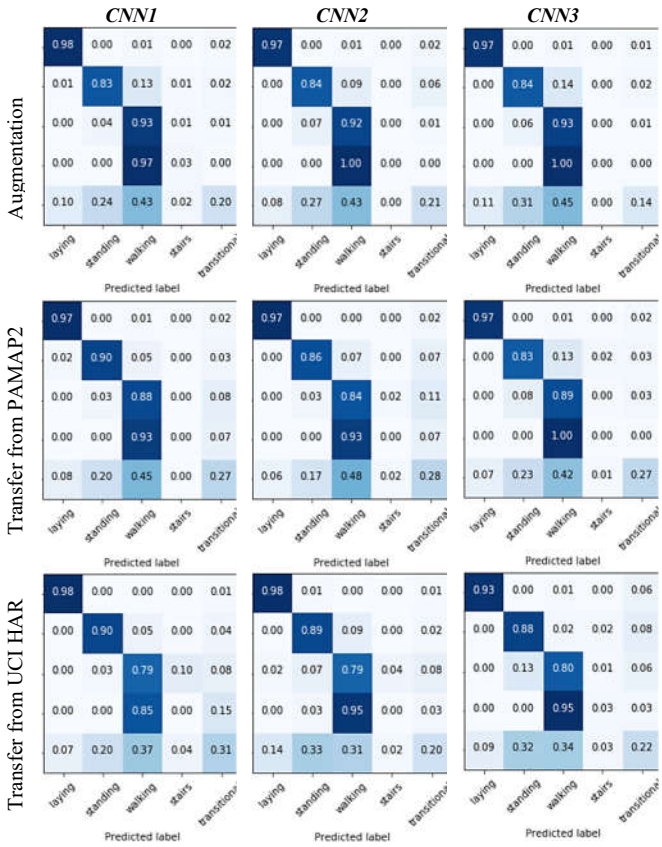


Fig. 4. Average confusion matrices for the test set using the 3 investigated CNNs when only data augmentation (1st row), transfer learning using PAMAP2 (2nd row), or transfer learning using UCI HAR (3rd row) is applied. For each matrix the vertical axis corresponds to true labels and the horizontal axis to predicted labels. The classes are arranged horizontally and vertically in the same order (from left to right: laying, standing, walking, ascending/descending stairs, transition).

55.1% and 90% depending on the context and method applied, while the accuracy is increased in a more recent work implementing deep convolutional and LSTM recurrent neural networks [17].

In general, a direct comparison with other studies is not feasible due to differences in the experimental setup including the type and number of activities, the incorporated sensors, and the use of different classification performance metrics. We focus particularly on monitoring systems for older adults for which functional status and age variability are serious confounding factors. By contrast, most works reporting higher classification accuracy are evaluated on data from more homogeneous groups, such as younger participants in good physical condition. Furthermore, we make use of only one sensor located at the sternum underneath the clothing (for discreetness in appearance), whereas the combination of more sensors located in different parts of the body, such as legs, could facilitate the recognition of activity.

Additionally, the data of this study were acquired in each participants' house using a prototype research platform (FrailSafe system) with possible measurement errors, whereas most of the studies report accuracies in controlled, simulated or laboratory settings with consistent data. Appropriate processing

steps, such as orientation recognition and classifier adaptation, were proposed in [42] to mitigate the effects of misplacement, and mis-orientation in the data. Such corrective post-acquisition actions are effective for reducing the measurement bias and for automating the analysis, but do not remove other types of noise in the data. Similarly, another work [31] has applied semi-supervised clustering to reduce data variations. Although highly appreciated, this technique is appropriate to correct for potential shifts in the subject's behavior, whereas in our case the intra-subject data variations are considered less critical than the inter-subject variations, caused mainly by large differences in the physical status across participants, and acquisition-related confounding factors, such as the use of a prototype device instead of commercial products.

Finally, and more importantly, the CNN architectures were assessed by data-split on the subjects, i.e. the model was trained using measurements from subjects not used during testing. This gives us an estimate on the method's accuracy when applied on recordings of new subjects. Other studies [2] [44] [45] [46] do not describe in details the validation procedure or assign the individual instances (time windows) to the training or test set without considering the subject exclusiveness in each set. The latter can increase the reported accuracy significantly, since different instances in recordings of the same subject can have very similar patterns. Similarly, in our deep learning approach, the training accuracy can reach the value of 97.9% for CNN2 with transfer learning, but we don't consider this value as indicative of model performance, since it does not generalize to new coming persons. A relevant type of analysis involves the construction of subject-specific models [3], where a unique model is created for each subject using part of the data for training and the remaining for testing. The obtained accuracies are expected to be higher and not directly comparable to the ones of subject-independent models, such as the models developed in this work.

VI. Conclusions

With the raise in the ageing population the use of technology promoting active and healthy ageing has gained increased attention in the last years. Assisted living systems largely rely on wearable devices equipped with activity classification capabilities. In this paper, we introduced data augmentation and transfer learning into 3 different CNN architectures in order to improve performance of physical activity recognition for senior adults. Bayesian optimization was used for hyper-parameter tuning in the original as well as modified architectures. The results show that each technique contributes differently in every CNN architecture. Overall the highest accuracy (84.89%) was observed with transfer learning from PAMAP2 on the CNN1 architecture, in which 1D convolution is performed on each of the 9 channels along the temporal dimension.

In the future, we plan to combine data augmentation and transfer learning, in order to further increase the classification accuracy. Finally, we aim to address possible differences in walking speed, especially when data from younger population groups are combined with recordings of older people, by investigating the use of dynamic time warping to align the weights of the convolutional filters in our CNNs [47], and therefore to reduce time-scale variations.

Acknowledgment

The authors want to thank all ICT partners from the FrailSafe Project for software or hardware support, and especially their colleagues K. Deltouzos and S. Kalogiannis from University of Patras for help with the data management. They also want to thank the medical team for the data acquisition.

References

[1] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive and mobile computing,* vol. 10, pp. 138-154, 2014.

[2] B. Andò, S. Baglio, C. O. Lombardo, V. Marletta, E. A. Pergolizzi, A. Pistorio and A. Valastro, "ADL Detection for the Active Ageing of Elderly People," in *Ambient Assisted Living*, Springer, 2015, pp. 287-294.

[3] E. Pippa, I. Mporas and V. Megalooikonomou, "Feature Selection Evaluation for Light Human Motion Identification in Frailty Monitoring System.," in *ICT4AgeingWell*, 2016.

[4] D. C. Ranasinghe, R. L. S. Torres and A. Wickramasinghe, "Automated activity recognition and monitoring of elderly using wireless sensors: Research challenges," in *Advances in Sensors and Interfaces (IWASI), 2013 5th IEEE International Workshop on*, 2013.

[5] G. Sebestyen, I. Stoica and A. Hangan, "Human activity recognition and monitoring for elderly people," in *Intelligent Computer Communication and Processing (ICCP), 2016 IEEE 12th International Conference on*, 2016.

[6] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *16th IEEE International Symposium on Wearable Computers (ISWC)*, 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," in *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)* , 2013.

[8] E. Pippa, E. I. Zacharaki, I. Mporas, V. Tsirka, M. P. Richardson, M. Koutroumanidis and V. Megalooikonomou, "Improving classification of epileptic and non-epileptic EEG events by feature selection," *Neurocomputing,* vol. 171, pp. 576-585, 2016.

[9] I. Mporas, V. Tsirka, E. Zacharaki, M. Koutroumanidis and V. Megalooikonomou, "Evaluation of time and frequency domain features for seizure detection from combined EEG and ECG signals," in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, 2014.

[10] E. I. Zacharaki, I. Mporas, K. Garganis and V. Megalooikonomou, "Spike pattern recognition by supervised classification in low dimensional embedding space," *Brain informatics,* vol. 3, pp. 73-83, 2016.

[11] T. Papastergiou, E. I. Zacharaki and V. Megalooikonomou, "Tensor decomposition for multiple instance classification of high-order medical data," *Complexity,* 2018.

[12] S. Sempena, N. U. Maulidevi and P. R. Aryan, "Human action recognition using dynamic time warping," in *IEEE International Conference on Electrical Engineering and Informatics (ICEEI)*, 2011.

[13] D. McGlynn and M. G. Madden, "An ensemble dynamic time warping classifier with application to activity recognition," *Research and Development in Intelligent Systems XXVII,* pp. 339-352, 2011.

[14] I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, Deep learning, Cambridge: MIT press, 2016.

[15] J. Yang, M. N. Nguyen, P. P. San, X. Li and S. Krishnaswamy, "Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition.," in *IJCAI*, 2015.

[16] 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 *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*, 2014.

[17] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors,* vol. 16, no. 1, p. 115, 2016.

[18] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *23rd ACM international conference on Multimedia*, 2015.

[19] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha and others, "Collecting complex activity datasets in highly rich networked sensor environments," in *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*, 2010.

[20] H. Sagha, S. T. Digumarti, J. d. R. Millán, R. Chavarriaga, A. Calatroni, D. Roggen and G. Tröster, "Benchmarking classification techniques using the Opportunity human activity dataset," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011.

[21] M. Fekri and M. O. Shafiq, "Deep Convolutional Neural Network Learning for Activity Recognition using real-life sensor's data in smart devices," in *IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2018.

[22] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," in *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, 2011.

[23] R. Saeedi, S. Norgaard and A. H. Gebremedhin, "A closed-loop deep learning architecture for robust activity recognition using wearable sensors," in *IEEE International Conference on Big Data*, 2017 .

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

[25] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Systems with Applications,* vol. 105, pp. 233-261, 2018.

[26] B. Liu, X. Wang, M. Dixit, R. Kwitt and N. Vasconcelos, "Feature space transfer for data augmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[27] X. Cui, V. Goel and B. Kingsbury, "Data augmentation for deep neural network acoustic modeling," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) ,* vol. 23, no. 9, pp. 1469-1477, 2015.

[28] T. Um, F. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek and D. Kuli´c, "Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks.," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI'17)*, Glasgow, UK, 2017.

[29] S. A. Rokni and H. Ghasemzadeh, "Autonomous Training of Activity Recognition Algorithms in Mobile Sensors: A Transfer Learning Approach in Context-Invariant Views," *IEEE Transactions on Mobile Computing,* vol. 17, no. 8, pp. 1764 - 1777, 2018.

[30] W.-L. Zheng and B.-L. Lu, "Personalizing EEG-based affective models with transfer learning," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016.

[31] H. Ali, E. Messina and R. Bisiani, "Subject-dependent physical activity recognition model framework with a semi-supervised clustering approach," in *IEEE European Modelling Symposium (EMS)*, 2013.

[32] A. Papagiannaki, E. I. Zacharaki, G. Kalouris, S. Kalogiannis, K. Deltouzos, J. Ellul and V. Megalooikonomou, "Recognizing physical activity of older people from wearable sensors and inconsistent data," *Sensors,* vol. 880, pp. 1-19, 2019.

[33] "Smartex," [Online]. Available: http://www.smartex.it/en/.

[34] "FrailSafe project," [Online]. Available: https://frailsafe-project.eu/.

[35] S. Kalogiannis, K. Deltouzos, E. I. Zacharaki, A. Vasilakis, K. Moustakas, J. Ellul and V. Megalooikonomou, "Integrating an openEHR-based personalized virtual model for the ageing population within HBase," *BMC medical informatics and decision making,* pp. 19-25, 2019.

[36] S. Kalogiannis, E. I. Zacharaki, K. Deltouzos, M. Kotsani, J. Ellul, A. Benetos and V. Megalooikonomou, "Geriatric group analysis by clustering non-linearly embedded multi-sensor data," in *2018 Innovations in Intelligent Systems and Applications (INISTA)*, 2018.

[37] J. Snoek, H. Larochelle and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012.

[38] I. Dewancker, M. McCourt, S. Clark, P. Hayes, A. Johnson and G. Ke, "A Stratified Analysis of Bayesian Optimization Methods," *arXiv preprint arXiv:1603.09441,* 2016.

[39] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard and others, "Tensorflow: a system for large-scale machine learning.," in *OSDI*, 2016.

[40] F. Chollet and others, *Keras,* 2015.

[41] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in neural information processing systems*, 2014.

[42] A. Papagiannaki, E. I. Zacharaki, K. Deltouzos, R. Orselli, A. Freminet, S. Cela, E. Aristodemou, M. Polycarpou, M. Kotsani, A. Benetos and others, "Meeting challenges of activity recognition for ageing population in real life settings," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2018.

[43] T. Plötz, N. Y. Hammerla and P. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2011.

[44] J. Liu, J. Sohn and S. Kim, "Classification of Daily Activities for the Elderly Using Wearable Sensors," *Journal of Healthcare Engineering,* vol. 2017, 2017.

[45] C. Moufawad el Achkar, C. Lenoble-Hoskovec, A. Paraschiv-Ionescu, K. Major, C. Büla and K. Aminian, "Instrumented shoes for activity classification in the elderly," *Gait & posture,* vol. 44, pp. 12-17, 2016.

[46] S. Chernbumroong, S. Cang, A. Atkins and H. Yu, "Elderly activities recognition and classification for applications in assisted living," *Expert Systems with Applications,* vol. 40, pp. 1662-1674, 2013.

[47] B. K. Iwana and S. Uchida, "Dynamic Weight Alignment for Temporal Convolutional Neural Networks," *arXiv preprint ,* vol. arXiv:1712.06530, 2017.

View publication stats