

AHA-3D: A Labelled Dataset for Senior Fitness Exercise Recognition and Segmentation from 3D Skeletal Data

João Antunes²

joaoa@andrew.cmu.edu

Alexandre Bernardino¹

<http://www.isr.ist.utl.pt/~alex>

Asim Smailagic²

<http://www.cs.cmu.edu/~asim>

Daniel Siewiorek²

<https://www.cs.cmu.edu/~dps/>

¹ Instituto Superior Técnico

Lisboa

Portugal

² Carnegie Mellon University

Pittsburgh

USA

Abstract

Automated assessment of fitness exercises has important applications in computer and robot-based exercise coaches to deploy at home, gymnasiums or care centers. In this work, we introduce AHA-3D, a labeled dataset of sequences of 3D skeletal data depicting standard fitness tests on young and elderly subjects, for the purpose of automatic fitness exercises assessment. To the best of our knowledge, AHA-3D is the first publicly available dataset featuring multi-generational, male and female subjects, with frame-level labels, allowing for action segmentation as well as the estimation of metrics like risk of fall, and autonomy to perform daily tasks.

We present two baseline methods for recognition and one for segmentation. For recognition, we trained models on the positions of the joints achieving $88.2\% \pm 0.077$ accuracy, and on joint positions and velocities, achieving $91\% \pm 0.082$ accuracy. Using the Kolmogorov-Smirnov test we determined the model trained on velocities was superior. The segmentation baseline achieved an accuracy of 88.29% in detecting actions at frame level.

Our results show promising recognition and detection performance suggesting AHA-3D's potential use in practical applications like exercise performance and correction, elderly fitness level estimation and risk of falling for elders.

1 Introduction

Exercising and physical activity has been identified as having a fundamental role in increasing quality of life in seniors, increasing not only an individual's lifespan but also allowing a more independent one [20][9]. The ability to perform daily activities without requiring help is a determinant factor not only for physical health, but also a promoter of social interactions and inclusion within society [9].

Automated coaching systems using gamification and motivational systems can play a crucial role in promoting fitness in elders, by providing new ways to engage the senior users in the exercise program. It can also reduce costs of labour in institutions, since a care professional can serve a bigger number of users with the help of such systems. The convenience of having these systems deployed in accessible locations like nursing homes, gymnasiums or even in the user's own home makes it possible to have care being administered to people who would otherwise not be able to afford it, or attend to it [13].

An important component to automated coaching systems is the ability to assess and monitor the progress of the fitness level of the users. In this paper, the AHA-3D, a dataset comprised of 3D skeletal data depicting both young and elderly subjects performing standard fitness tests, is presented. The fitness tests performed by the subjects are the 30-Second chair stand, 8-feet up and go, 2-minute step and unipedal stance (described in Section 2) [14]. These tests are designed to assess musculoskeletal, motor and cardiovascular/aerobic fitness, with an emphasis on lower body strength due to the role it takes in common activities for daily tasks like walking, sitting and standing on a chair, or entering a car/bus [14].

The AHA-3D dataset is comprised of a manually labelled and curated selection of subjects found in the AHA dataset [4]. In the AHA dataset the acquisition of data is multimodal including RGB cameras, 3D-cameras, and physiological signal acquisition. The AHA-3D dataset presented in this paper is built from the Skeletal data acquired from the 3D cameras (Kinect v2) in conjunction with the data from the RGB cameras to aid in the labelling process. In order to allow work involving not only action classification but also detection/segmentation the labelling process was done at a frame level.

While there are some datasets that depict humans doing exercise like WorkoutSU-10 [15] or the Weight Lifting Exercises Dataset [16] the AHA-3D is the only one depicting exercises designed for and performed by elderly people.

The contributions resulting from this work are as follows: a novel dataset is provided, depicting both elderly and young subjects of both genders performing exercises designed to assess the fitness level of elderly people. These exercises are chosen by their correlation to the main fitness dimensions of interest: strength, endurance, gait speed, agility and balance [8][14]. Furthermore, the annotations provided in the AHA-3D are more finely acquired than other fitness datasets since one label per frame is provided for every video in the dataset. Two different baseline methods are provided: one for action classification and one for action detection. The results obtained indicate that the AHA-3D is suitable for feature extraction by automatic methods like joint angle estimation, counting the number of repetitions performed, or time spent performing an exercise (since frame-level labels are available), among others. Furthermore all code pertaining to these models as well as the dataset itself will be made available for free, online.

The rest of this document is organized as follows: in Section 2 the AHA-3D is described, as well as the data modalities made available, the different types of subjects that performed the exercises and the exercises themselves. In Section 3 the methods used to define the baseline performances for the recognition and detection tasks are described in detail. In Section 4 the results obtained are presented and explained, with the conclusions and future work considerations finishing the paper in Section 5.

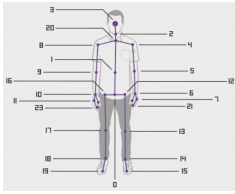


Figure 1: Joint identification from 3D Skeletal data obtained with Kinect v2

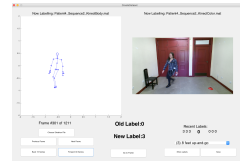


Figure 2: GUI developed to aid in the labelling process. Skeletal data is matched with the corresponding RGB frame.

2 AHA-3D Description

The AHA-3D dataset depicts exercises being performed by young and elderly subjects, in the form of 3D Skeleton data(see Fig. 1). There are 79 Skeleton videos each containing between 1 to 3 runs of the same exercise. Of 21 subjects performing the exercises, 11 were young and 10 were elderly. 5 of the subjects were male, and 16 were female. Table 1 shows how many frames were acquired from each subject during each exercise.

The Skeletal data is captured at 30 FPS. To aid in the labelling process a GUI was developed that shows the corresponding RGB image at the same timestep (see Fig. 2).

The data and all code pertaining to this dataset (as well as other datasets) is available at <http://vislab.isr.ist.utl.pt/datasets/>.

The exercises depicted in the dataset are now described.

2.1 30-second chair stand

In this exercise, the subjects are instructed to to stand up and sit back down as many times as possible, in 30 seconds, from a seated position. The goal is to measure how many repetitions the subjects can perform in 30 seconds. This assesses the lower body strength of the subject and has been shown to be correlated to laboratory measures of lower body strength like one repetition maximum load leg press, knee extensor and knee flexor strength [14]. An example of the data obtained from this exercise is shown in Fig. 3.

2.2 8-feet up and go

This exercise measures how long it takes for the subject to get up from a chair, walk 2.4 meters maneuver around a plastic cone placed on the floor, walk back and return to the seated position. It has been shown to be correlated with the risk of fall. Subjects that take longer than 8.5 are shown to have high fall risk [14]. An example of the data obtained from this exercise is shown in Fig. 4.

2.3 2-minute step test

This exercise determines the maximum amount of steps the subject can do in 2 minutes. From a standing position, the subject is instructed to raise each knee (alternating) to the height of the midway point between the knee-cap and hip. It has been shown to be a good indicator of the overall functional fitness needed for daily tasks [14]. An example of the data obtained from this exercise is shown in Fig. 5.

Exercise(#Frames)/ Subject ID	30-second chair stand	8-feet up and go	2-minute step test	Unipedal Stance	Total:
1, Young Female	1457	1720	3876	3379	10432
2, Young Female	1353	1562	3498	1670	8083
3, Young Female	1259	1411	3777	1926	8373
4, Young Female	1215	1211	3598	1420	7444
5, Young Male	1254	1188	4026	1235	7703
6, Elderly Female	0	1443	4028	1432	6903
7, Elderly Female	0	2280	3817	1393	7490
8, Elderly Female	1266	1474	3465	1359	7564
10, Elderly Male	0	2023	3985	3301	9309
11, Elderly Female	0	1894	3706	1394	6994
13, Elderly Female	1428	1584	3948	1551	8511
14, Elderly Female	1643	1625	3860	1661	8789
15, Elderly Female	0	2451	4307	1730	8488
16, Elderly Female	0	1642	4466	0	6108
17, Elderly Female	0	2109	4132	1441	7682
18, Young Female	2557	1626	4515	1692	10390
19, Young Female	1750	1362	3660	1729	8501
20, Young Male	1199	1742	3541	2008	8490
21, Young Female	1930	1655	4057	0	7642
22, Young Female	1287	1274	3806	1944	8311
23, Young Female	1515	1281	4076	1674	8546
Total:	21113	34557	82144	33939	171753

Table 1: Number of frames per exercise found in AHA-3D. When the exercise was discarded, because of problems on the acquisition or deemed insufficient in quality to be included in this curated dataset the number of frames was set to 0.

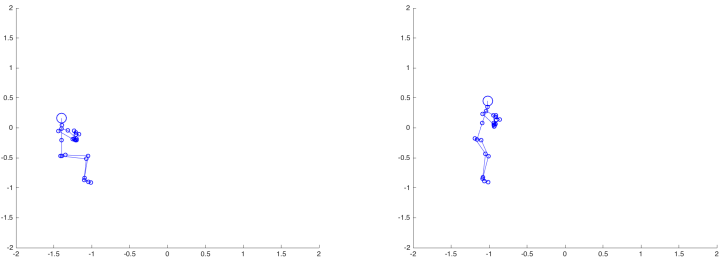


Figure 3: Two skeletons from the same 30-second chair stand exercise. In this exercise, the subjects are instructed to start off in a seated position and then rise to a standing position and returning to the initial seating position as many times as possible in 30 seconds. This test assesses the lower body strength of the subject and is related to lower body strength indicators like one repetition maximum load leg press, knee extensor and knee flexor strength

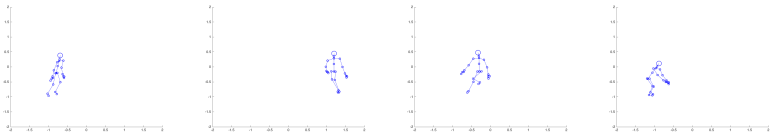


Figure 4: Four skeletons from the same 8-feet up and go exercise. In this exercise the subjects are instructed to start off in a seated position, stand up, walk 2.4 meters (8 feet) maneuver around a plastic cone, walk back and take a seat again. It has been shown that taking longer than 8,5 seconds to perform this task is an indicator of a high risk of falling. [14]

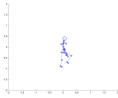


Figure 5: Example of the data obtained from the 2-minute step test exercise. It has been shown to be a good indicator of the overall functional fitness level required for daily tasks. [14]

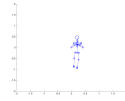


Figure 6: Example of the data obtained from the Unipedal Stance exercise. This test assesses static balance. [14]

2.4 Unipedal Stance

This exercise determines how long the subject can balance him/herself on only one limb. Starting from a standing position, with his/her arms crossed on his/her chest, time starts counting when one of the feet leaves the ground. Time is interrupted anytime the subject either used his arms, moved the weight-bearing foot to maintain balance, moved the raised foot for balance (either touching the floor or the other limb). The maximum score is 45 seconds (uninterrupted). This test assesses static balance [14]. An example of the data obtained from this exercise is shown in Fig. 6.

3 Baseline methods

Along with the AHA-3D dataset, two baseline methods are also presented: a classification method that predicts which exercise is being executed, and a detection method that outputs one label per frame describing whether or not an exercise is being executed. Note that, in the detection baseline, the specific exercise being executed is irrelevant, and only whether the subject is exercising or resting is determined.

The AHA-3D is comprised of 79 different 3D Skeletal videos, each containing 1-3 exercises being performed, the data was split as 39 videos for training, 20 for validation and 20 for testing. These splits were decided at random and fixed fair comparison of the various networks employed in designing a baseline.

All the code for both baselines was made in Python using the Keras library [9] with a Tensorflow backend [10].

#Layers	2
#Hidden Units	512/256
Joints Used	1-21 (See Fig.1)
Learning Rate	10^{-3}
Dropout Rate	50%
Activation Function	ReLU
#Epochs	300

Table 2: Hyperparameters and design choices used on the best performing network in the action recognition task.

#Layers	1
#Hidden Units	64
Joints Used	1-21 (See Fig.1)
Learning Rate	10^{-3}
Lookback	30 frames
Overlap	25 frames
#Epochs	300

Table 3: Hyperparameters and design choices used on the best performing network in the action detection task.

3.1 Recognition

The current state-of-the-art for action recognition is populated mainly with CNN’s with deep architectures, with the most prevalent approaches being some variant of a two-stream Convolutional Neural Network (CNN) [10]. One stream is used to process individual raw video frames performing action recognition on still images, called the Spatial Stream. The other stream the input data is some representation of movement to complement the still image classification, the temporal stream. The most common way to codify the input for this network is using optical flow [10]. Both streams are usually pre-trained on a large dataset like the ImageNet [8]. For a more thorough state-of-the-art on action recognition we point to João Carreira *et al.*’s paper [9] where he proposes a new approach to Two-Stream CNN’s that seems to outperform every other CNN so far by inflating the kernels used by both streams to 3-D and taking tensors as input on both streams: a collection of raw frames on the spatial stream, and a collection of optical flow frames on the temporal stream.

For the action recognition baseline a feed-forward neural network was employed. The hyperparameters and design choices that yielded the best performing network, as well as the pre-processing of the data that was necessary to run this network are discussed next. The network that yielded the best results on the validation set was a 2-layer feed-forward neural network with the hyperparameters show in Table 2.

The joints used were 1-21. The remaining joints were discarded as the position of the thumbs/hands were not relevant for the exercises present in AHA-3D.

In order to classify each video, we subdivide each video into 60 frame snippets that are fed to the network. Then, each snippet gets classified independently, by the network. The final video label is then decided by majority voting of the labels predicted for each snippet.

3.2 Action Detection

The current state-of-the-art for action detection revolves around two key concepts: Actionness and Completeness. Actionness is a probability measure of a given action proposal window depicting an action [8][10][14]. Completeness is a concept that defines whether or not a given proposal window depict an action from start to finish, as opposed to only a fraction of an action [14]. The best performing methods on action detection employ inference tools (usually neural networks) to estimate the quality of action proposal windows with regards to both its actionness and completeness.

For the action detection baseline Long Short Term Memory (LSTM) networks are used. This recurrent neural network architecture was designed to be able to store information

across long periods of time, dealing with vanishing gradient limitations of common recurrent neural network architectures [9].

The hyperparameters and design choices that yielded the best performing network, as well as the pre-processing of the data that was necessary to run this network, are now explained.

The hyperparameters of the best performing LSTM network are shown in Table 3.

The lookback hyperparameter refers to how many frames are input to the LSTM at each time step. The overlap hyperparameter describes how much overlap exists between consecutive time slices input to the network. As such, with Lookback = 30 the LSTM will, at each iteration, receive a tensor of 30 frames of skeletal data (1 second time span) and use those frames along with the internal representation already calculated on previous iterations to classify the label that corresponds to the last frame in that tensor. With Lookback = 30 and Overlap = 25 the granularity of the labels produced by this model is 5 frames (*i.e.* given a video with 100 frames, this model will output labels corresponding to frames 30, 35, 40, ..., 100). The goal of this baseline is to determine the time windows in which the subject is performing an exercise. For performance speedup, the number of frames found on the longest video is calculated and all other videos are zero padded until all the videos have the same number of frames.

4 Experiments and Results

4.1 Experiments

To assess performance of the baseline models proposed in Sec.3 each model was tested on 100 runs. On each model run the 79 videos present in AHA-3D are randomly sorted into 39 for training, 20 for validation and 20 for testing. All performance metrics shown were calculated by averaging over 100 runs.

For the Action Recognition experiments an extra pre-processing step was applied. We calculated the first and second derivatives for joint coordinates. All models were trained on joint positions, joint positions and velocities, and joint positions, velocities and accelerations. The results presented refer to the accuracy per video, *i.e.*: a score of 50% means that 50% of the videos were labelled correctly.

For the Action Detection experiments the results presented refer to the accuracy per frame *i.e.*: a score of 50% means that 50% of the frames got labelled correctly.

4.2 Results

4.2.1 Action Recognition Results

For the action recognition baseline a 2 layer feed-forward neural network with dropout and majority voting (see Sec. 3.1). When trained only on joint positions, our we obtained a mean testing accuracy of $88.2\% \pm 0.077$. When trained on joint positions and velocities, our best performing network had the same hyper-parameters but exhibited a mean testing accuracy of $91\% \pm 0.082$. We used the Kolmogorov-Smirnov test to determine whether the difference in performance was relevant and found that, with over 95% confidence, the difference was relevant. Training on positions, velocities and accelerations of the joints provided no improvement vs. only positions and velocities.

Predicted Label\Ground Truth	Unipedal Stance	8-FeeT Up and Go	30-Second Chair Stand	2-Minute Step	Total
Unipedal Stance	450	0	0	35	485
8-FeeT Up and Go	0	607	69	0	676
30-Second Chair Stand	0	10	246	0	256
2-Minute Step	45	0	21	517	583
Total	495	617	336	552	2000

Table 4: Confusion Matrix for our best performing network trained on joint positions and derivatives. See Sec. 3.1 for network details.) The first line of the matrix denotes that out of 485 Unipedal Stance videos 450 were correctly labelled, 0 were labelled as 8-FeeT Up and Go, 0 were labelled as 30-Second Chair Stand, and 35 were labelled as 2-Minute Step Test. The network correctly classifies the Unipedal Stance exercise 450 out of 485 times. The same reasoning can be applied to all lines in the matrix. Its apparent that 2 groups are established in this matrix. One group with the Unipedal Stance and 2-Minute Step exercises and another one with the 8-FeeT Up and Go and 30-Second Chair Stand exercises. The most obvious differences between these 2 groups are the starting positions (*i.e.* where in the room) and the starting poses. More details in Sec. 4.2.1

In Table 4 we present the Confusion Matrix resulting from our best performing network on joint positions and velocities. The first line of the matrix denotes that out of 485 videos depicting a Unipedal Stance exercise 450 videos were correctly labelled, and 35 were wrongfully labelled as 2-Minute Step Test. This shows that the network correctly classifies the Unipedal Stance exercise correctly 450 out of 485 times. The same reasoning can be applied to all lines in the matrix. It’s interesting to note that the network has established a separation between two groups of exercises, one containing the Unipedal Stance and 2-Minute Step tests and the other the 8-FeeT Up and go and 30-Second Chair Stand (see Table 4. The Unipedal Stance and 2-Minute Step Tests both start from a standing position, while 8-FeeT up and Go and 30-Second Chair stand both start from a seated position. The pose of the skeleton in these frames might be the trigger for this separation. Another possibility is that the trigger for the separation might be the starting position (text*i.e.* where in the room). Since a chair in the left-most wall of the room was where every subject sat, when instructed.

4.2.2 Action Detection Results

For the Action Detection baseline the mean accuracy of our best performing network was **88.29%**, *i.e.*: 88.29% of all the frames were labelled correctly. In Fig. 7, the results obtained on the best performing LSTM network on one run of 20 test videos are shown. There are 20 graphs, each corresponding to one video. Each graph has the ground truth label in blue and the predicted label in black. A label of 0 corresponds to the sections of the video in which the subject is not performing an exercise, and a label of 1 when he is. If the ground truth and the predicted label match a green bar will show. If a false negative occurs a yellow bar will show. If a false positive occurs a red bar will show. The results obtained follow the ground truth data except for one recurring artifact: in videos where the majority of the frames are depicting an exercise, with a small amount of samples of rest (see videos #1, #8, #15, #19, in Fig. 7) the model states that the entire video is one exercise. This was found on all models, and only happened on videos depicting the 2-minute step test exercise. This exercise is the longest one, and that skewed the results in the matter aforementioned.

The results obtained in both baselines indicate that AHA-3D would be well suited for

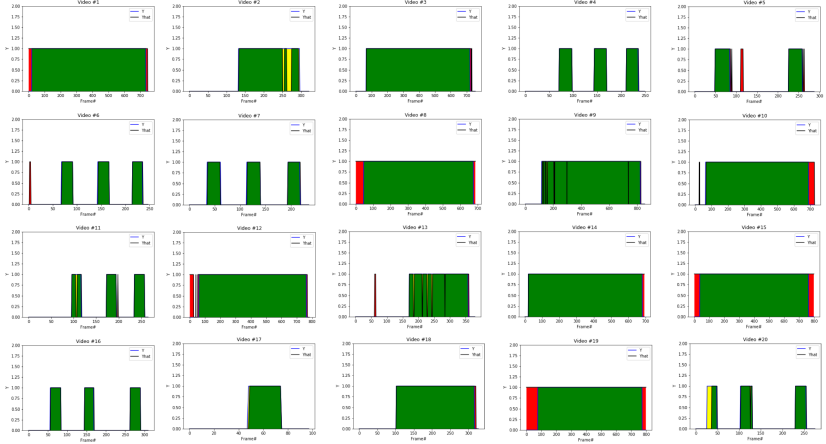


Figure 7: Results obtained on the best performing LSTM network. Each graph corresponds to one video. Each graph has the ground truth label in blue and the predicted label in black. A label of 0 corresponds to no action and refers to the sections of the video in which the subject is not performing an exercise. A label of 1 corresponds to the frames in which the subject is performing an exercise. If the ground truth and the predicted label match a green bar will be underneath the plot. If a false negative occurs a yellow bar will be underneath the plot. If a false positive occurs a red bar will be underneath the plot.

models that attempt to learn metrics that can be extracted from skeletal data (*p.e.* measuring joint angles, calculating exercise duration/number of repetitions). Furthermore, the exercises depicted in AHA-3D are standard fitness tests used to determine metrics describing autonomy and quality of life in elders.

5 Conclusions

A dataset is presented, AHA-3D, composed of 3D skeletal data of young and elder subjects performing four standard fitness tests. 11 young and 10 senior subjects were recorded and made their data available for research. To the best of our knowledge, this is the only dataset made available with senior and young subjects of both genders performing standard fitness tests. Furthermore, our dataset is manually labelled with one label per frame, instead of the more common one label per video approach. This type of labelling permits more delicate measures to be obtained with confidence, like determining the start/end of an exercise. The AHA-3D, as well as all code pertaining to the baseline methods described in this paper are available for free. Baseline methods trained on this dataset achieve high accuracy rates in both recognition and detection tasks, 91% and 88.29% respectively, which justifies the notion that the AHA-3D is appropriate to extract and evaluate metrics on exercise quality like repetition counting, exercise segmentation in time or accurate joint angle calculations. These metrics are standard for determining factors like the fall risk, or the autonomy that and elderly subject can be expected to have. It also points to the theory that this dataset would also be adequate for inferring exercise performance quality, which can have real world applications on automated fitness coaches for use at home, gymnasiums or care centers.

6 Acknowledgements

This work was partially supported by the FCT projects [UID/EEA/50009/2013], AHA CMUP-ERI/HCI/0046/2013, doctoral grant [SFRH/BD/106068/2015].

References

- [1] Alexandre Bernardino, Christian Vismara, Sergi Bermudez i Badia, Ãl'vio Gouveia, FÃtima Baptista, Filomena Carnide, SimÃo Oom, and Hugo Gamboa. A dataset for the automatic assessment of functional senior fitness tests using kinect and physiological sensors. In *Technology and Innovation in Sports, Health and Wellbeing (TISHW), International Conference on*, pages 1–6. IEEE, 2016.
- [2] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4733. IEEE, 2017.
- [3] Eliza F. Chakravarty, Helen B. Hubert, Vijaya B. Lingala, and James F. Fries. Reduced disability and mortality among aging runners: a 21-year longitudinal study. *Archives of internal medicine*, 168(15):1638–1646, 2008.
- [4] FranÃois Chollet and others. Keras. GitHub, 2015. URL <https://github.com/fchollet/keras>.
- [5] Xiyang Dai, Bharat Singh, Guyue Zhang, Larry S. Davis, and Yan Qiu Chen. Temporal Context Network for Activity Localization in Videos. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5727–5736. IEEE, 2017.
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [7] Felix A. Gers, JÃirgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with LSTM. 1999.
- [8] C Jessie Jones and Debra J Rose. *Physical activity instruction of older adults*. Human Kinetics, 2005.
- [9] Abby C. King and Diane K. King. Physical activity for an aging population. *Public Health Reviews*, 32(2):401, 2010.
- [10] Tianwei Lin, Xu Zhao, and Zheng Shou. Temporal Convolution Based Action Proposal: Submission to ActivityNet 2017. *arXiv preprint arXiv:1707.06750*, 2017.
- [11] MartÃn Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion ManÃl, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda ViÃgas, Oriol Vinyals, Pete Warden,

- Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. URL <https://www.tensorflow.org/>.
- [12] Farhood Negin, F  rat   zdemir, Ceyhun Burak Akg  jl, Kamer Ali Y  ijkseel, and Ayt  jl Er  gil. A decision forest based feature selection framework for action recognition from rgb-depth cameras. In *International Conference Image Analysis and Recognition*, pages 648–657. Springer, 2013.
- [13] Ferda Ofli, Gregorij Kurillo,   t  p  n Obdr    lek, Ruzena Bajcsy, Holly Brugge Jimison, and Misha Pavel. Design and evaluation of an interactive exercise coaching system for older adults: lessons learned. *IEEE journal of biomedical and health informatics*, 20(1):201–212, 2016.
- [14] Roberta E. Rikli and C. Jessie Jones. *Senior fitness test manual*. Human Kinetics, 2013.
- [15] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.
- [16] Barbara A. Springer, Raul Marin, Tamara Cyhan, Holly Roberts, and Norman W. Gill. Normative values for the unipedal stance test with eyes open and closed. *Journal of geriatric physical therapy*, 30(1):8–15, 2007.
- [17] Javier S    nchez P    rez, Enric Meinhardt-Llopis, and Gabriele Facciolo. TV-L1 Optical Flow Estimation. *Image Processing On Line*, 3:137–150, 2013. doi: 10.5201/ipol.2013.26.
- [18] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. Qualitative activity recognition of weight lifting exercises. In *Proceedings of the 4th Augmented Human International Conference*, pages 116–123. ACM, 2013.
- [19] Yuanjun Xiong, Limin Wang, Zhe Wang, Bowen Zhang, Hang Song, Wei Li, Dahua Lin, Yu Qiao, Luc Van Gool, and Xiaoou Tang. CUHK & ETHZ & SIAT submission to ActivityNet challenge 2016. *arXiv preprint arXiv:1608.00797*, 2016.
- [20] Laurel B. Yates, Luc Djouss    , Tobias Kurth, Julie E. Buring, and J. Michael Gaziano. Exceptional longevity in men: modifiable factors associated with survival and function to age 90 years. *Archives of Internal Medicine*, 168(3):284–290, 2008.
- [21] Yue Zhao, Yuanjun Xiong, Limin Wang, Zhirong Wu, Xiaoou Tang, and Dahua Lin. Temporal action detection with structured segment networks. In *The IEEE International Conference on Computer Vision (ICCV)*, volume 8, 2017.