



COMPUTER SCIENCE

NOVEL-RESULT

Fall detection for elderly-people monitoring using learned features and recurrent neural networks

Daniele Berardini^{*} , Sara Moccia, Lucia Migliorelli, Iacopo Pacifici, Paolo di Massimo, Marina Paolanti  and Emanuele Frontoni

Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy

*Corresponding author. E-mail: dani.berard.db@gmail.com

(Received 02 January 2020; Accepted 19 January 2020).

Abstract

Elderly care is becoming a relevant issue with the increase of population ageing. Fall injuries, with their impact on social and healthcare cost, represent one of the biggest concerns over the years. Researchers are focusing their attention on several fall-detection algorithms. In this paper, we present a deep-learning solution for automatic fall detection from RGB videos. The proposed approach achieved a mean recall of 0.916, prompting the possibility of translating this approach in the actual monitoring practice. Moreover to enable the scientific community making research on the topic the dataset used for our experiments will be released. This could enhance elderly people safety and quality of life, attenuating risks during elderly activities of daily living with reduced healthcare costs as a final result.

Keywords: Bidirectional LSTM; convolutional neural networks; fall detection; fine tuning; RGB videos

Introduction

Fall-related injury represents a major social issue. Fall frequency drastically increases among elderly: the 28–35% of people over 65 fall each year (World Health Organization, Ageing and Life Course Unit, 2008). This percentage reaches the 32–42% for people over 70. As a consequence, automatic fall detection is becoming a crucial task to increase safety of elderly people, especially when they live alone.

According to (Mubashir, Shao, & Seed, 2013), fall detection approaches are based on: wearable, ambience and vision devices (Mehmood, Nadeem, Ashraf, Alghamdi, & Siddiqui, 2019; Liciotti, Bernardini, Romeo, & Frontoni, *in press*; Shojaei-Hashemi, Nasiopolous, Little, & Pourazad, 2018; Wang, Chen, Zhou, Sun, & Dong, 2016). Regarding wearable devices, elderly often forget or refuse to wear them (Mubashir et al., 2013) and approaches that use ambient devices suffer from high sensitivity to noise (Delahoz & Labrador, 2014). A possible solution would be to monitor, through RGB camera, home environment and develop a fall-detection algorithm that could alert caregivers once falls occur.

Objective

The goal of this work is to develop an automatic solution to perform fall detection from RGB video sequences (the code is available on request on GitHub¹). The proposed approach exploits VGG16 (Simonyan & Zisserman, 2014), a convolutional neural network (CNN), as feature extractor and a

¹git clone <https://github.com/daniebera/deep-learning-fall-detection.git>

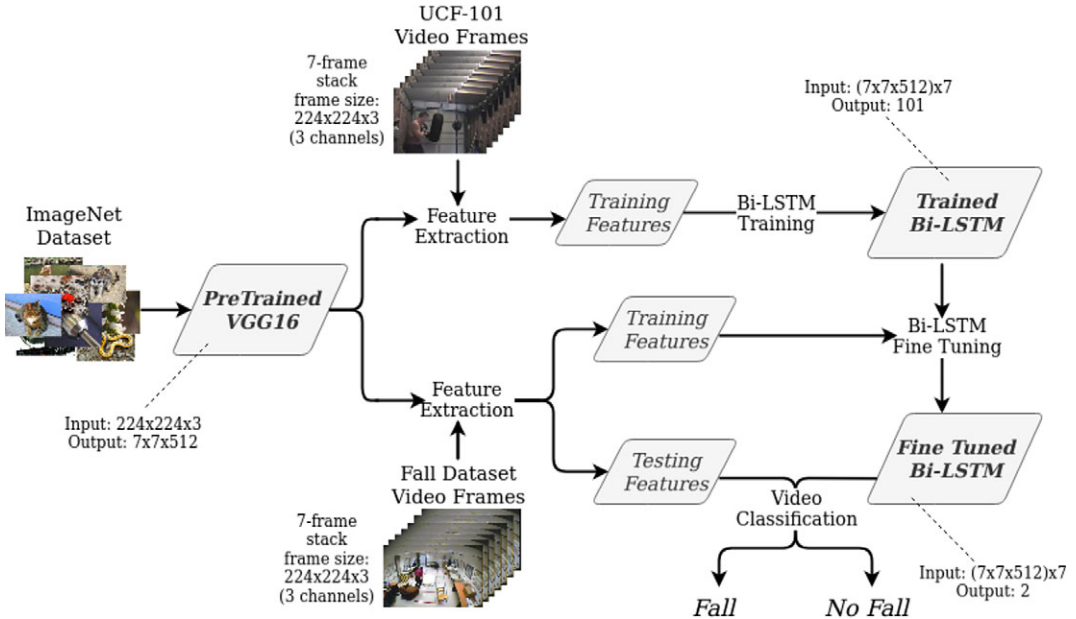


Figure 1. Workflow of the proposed approach for fall detection from RGB video sequences.

Bidirectional Long Short Term Memory (Graves & Schmidhuber, 2005) (Bi-LSTM), a recurrent neural network (RNN), as feature classifier. Due to the lack of large and annotated publicly available datasets in the field, we used VGG16 pretrained on the ImageNet dataset and Bi-LSTM pretrained on the UCF-101 action recognition dataset.² We then fine-tuned the Bi-LSTM for the fall-detection task on a custom-built dataset (the fall dataset) publicly available for the scientific community.³ Figure 1 shows an overview of the workflow of the proposed method.

Methods

To exploit temporal information, we processed stacks of 7 consecutive frames (inter-frame distance was 1 s). To extract features from UCF-101 videos, we used two pretrained VGG16 configurations: with (*Top*) and without (*No Top*) fully connected layers, resulting in a vector of 4096×7 and 25088×7 features, respectively. These features were used to train two Bi-LSTMs (one per VGG16 configuration). VGG16 was used as feature extractor since, unlike other deeper networks (*e.g.*, ResNet50), with its few layers it can extract more general features from images improving the generalization power of the Bi-LSTM classifiers.

We selected the configuration giving the highest macro recall (*Rec*) on UCF-101, and fine tuned it on the fall dataset (without freezing any layers (*No Freeze*) and freezing the first layer (*Freeze*)) to perform fall detection. To fine tune the selected configuration we added two dense layers (32 and 2 neurons, respectively) on top of the original Bi-LSTM architecture. We initialized these last two layers with the standard *Glorot* initialization, while the other layers with the weights resulting from the training on UCF-101. The fall dataset is composed of 216 videos (108 falls and 108 activities of daily living) collected from different sources. From each video, the first 7 frames were extracted (inter-frame distance = 225 ms). The Bi-LSTM was fine tuned for 150 epochs using Adam to minimize the categorical cross entropy. The best Bi-LSTM model was retrieved as the one that maximised *Rec* on the validation set.

²<https://www.crcv.ucf.edu/data/UCF101.php>

³<http://192.168.2.30/owncloud/index.php/s/1FmsIiKWvO9APVw>

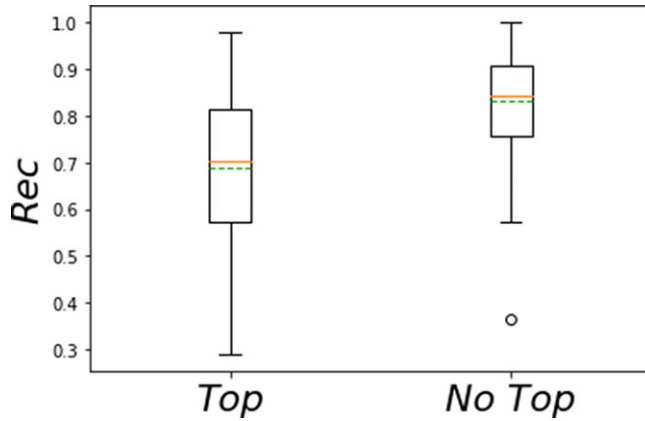


Figure 2. Boxplots of the *Rec* for classification on UCF-101 achieved with the *Top* and the *No Top* configurations. *Mean* (green) and *Median* (orange) of the *Rec* are showed too.

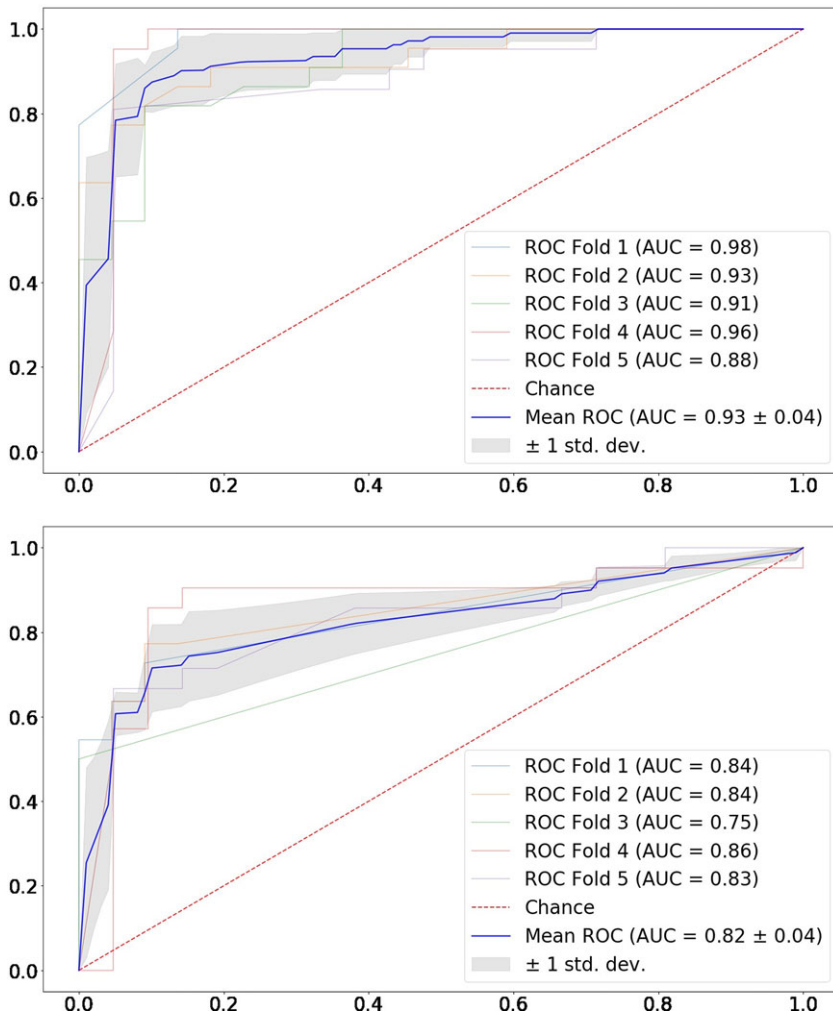


Figure 3. ROC curves for the five folds of the *No Freeze* (up) and the *Freeze* (down) approaches. Mean AUC (\pm standard deviation) is reported, too.

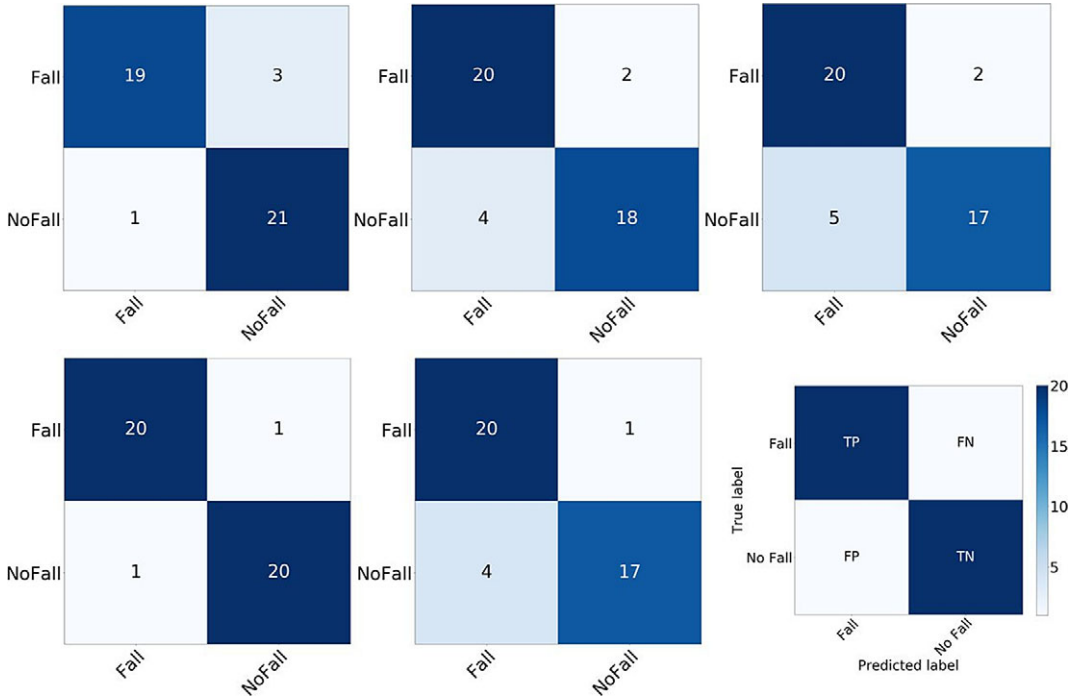


Figure 4. Confusion matrices for 5-fold cross-validation with the *No Freeze* configuration. The colorbar indicates the number of test samples.

Results

Figure 2 shows the descriptive statistics for *Rec* for the two Bi-LSTM tested on UCF-101 using the (left) *Top* and (right) *No Top* configuration. The best result was achieved by the *No Top* configuration (mean *Rec* = 0.836).

For robust result evaluation, performance on the fall dataset was assessed with 5-fold cross validation. For each fold, the 30% of the training set was used as validation. Results were evaluated with the area under the ROC curves (*AUC*) and the confusion matrices (one per fold). Figure 3 shows the ROC curves for the (left) *No Freeze* and (right) *Freeze* configuration. The best result was achieved with *No Freeze* ($AUC = 0.93 \pm 0.04$).

Figure 4 shows the confusion matrices for the five folds, achieved on fine tuned Bi-LSTM without freezing any layers. The mean *Rec* for all the folds was 0.916 and 0.860 for the fall and no fall class, respectively.

Discussions

The proposed approach implemented fine tuning to overcome challenges related to the small size of the fall dataset. The *No Top* approach achieved higher macro *Rec* (Fig. 2.). Indeed, the Bi-LSTM is fed with features from the last VGG16 convolutional layer, which are less specific than those from the fully connected one (used in the *Top* configuration).

Figure 3 shows that the *No Freeze* approach achieved the highest performance, which may be attributed to the considerable difference between videos in UCF-101 and in fall dataset. Confusion matrices in Fig. 4 show comparable values among the two classes, pointing out the stability on the proposed approach.

Conclusions

In this paper, we proposed a deep-learning method for automatic fall detection from RGB videos. The proposed method showed promising results (mean *Rec* = 0.916 for the fall class), which could be eventually

enhanced by enlarging the training dataset and testing more advanced architectures based on spatio-temporal features (Colleoni, Moccia, Du, De Momi, & Stoyanov, 2019). As future work we will investigate longer frame sequences to exploit the full potential of Bi-LSTM. Moreover, we plan to implement a CNN with 3D convolutions (able to process features both in space and time) to accomplish fall-detection task. This model could be used for comparison against the approach proposed in this contribution.

To account for privacy issues, future work could also deal with depth-video processing, which has already shown promising results for movement analysis (Moccia, Migliorelli, Pietrini, & Frontoni, 2019). This work, in integration with an alert system, could have a positive impact on elderly people safety and quality of life by ensuring prompt implementation of first aid.

Acknowledgements. We thank our colleagues from Department of Information Engineering, Università Politecnica delle Marche, who provided insight and expertise that greatly assisted the research.

Author Contributions. Daniele Berardini, Sara Moccia and Emanuele Frontoni conceived and designed the study and the methodology. Iacopo Pacifici and Paolo di Massimo conducted data gathering. Daniele Berardini and Sara Moccia performed statistical analyses. Daniele Berardini, Sara Moccia and Lucia Migliorelli wrote the article. Marina Paolanti reviewed and corrected the article.

Funding Information. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of Interest. Author Daniele Berardini, author Sara Moccia, author Lucia Migliorelli, author Iacopo Pacifici, author Paolo di Massimo, author Marina Paolanti and author Emanuele Frontoni declare none.

Data availability. The dataset that support the findings of this study is available from the author D. Berardini at <http://192.168.2.30/owncloud/index.php/s/1FmsLiKWvO9APVw> upon reasonable request.

The code necessary to reproduce the work is available on GitHub using the following bash command: “git clone <https://github.com/daniebera/deep-learning-fall-detection.git>”, from the corresponding author, D. Berardini, upon reasonable request.

References

- Colleoni, E., Moccia, S., Du, X., De Momi, E., & Stoyanov, D. (2019). Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers. *IEEE Robotics and Automation Letters*, 4, 2714–2721.
- Delahoz, Y. S., & Labrador, M. A. (2014). Survey on fall detection and fall prevention using wearable and external sensors. *Sensors*, 14, 19806–19842.
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18, 602–610.
- Liciotti, D., Bernardini, M., Romeo, L., & Frontoni, E. (in press). A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing*.
- Mehmood, A., Nadeem, A., Ashraf, M., Alghamdi, T., & Siddiqui, M. S. (2019). A novel fall detection algorithm for elderly using SHIMMER wearable sensors. *Health and Technology*, 9, 1–16.
- Moccia, S., Migliorelli, L., Pietrini, R., & Frontoni, E. (2019). *Preterm infants' limb-pose estimation from depth images using convolutional neural networks*. IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, Siena – Tuscany, Italy.
- Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. *Neurocomputing*, 100, 144–152.
- Shojaei-Hashemi, A., Nasiopolous, P., Little, J. J., & Pourazad, M. T. (2018). *Video-based human fall detection in smart homes using deep learning*. IEEE International Symposium on Circuits and Systems, Florence, Italy.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Wang, S., Chen, L., Zhou, Z., Sun, X., & Dong, J. (2016). Human fall detection in surveillance video based on PCANet. *Multimedia Tools and Applications*, 75, 11603–11613.
- World Health Organization, Ageing and Life Course Unit. (2008). *WHO global report on falls prevention in older age*. World Health Organization.

Peer Reviews


Reviewing editor: Dr. Adín Ramírez Rivera

UNICAMP, Institute of Computing, Av. Albert Einstein 1251, Campinas, São Paulo, Brazil, 13083-872

This article has been accepted because it is deemed to be scientifically sound, has the correct controls, has appropriate methodology and is statistically valid, and met required revisions.

doi:10.1017/exp.2020.3.pr1

Review 1: Fall detection for elderly-people monitoring using learned features and recurrent neural networks

Reviewer: Dr. Utku Kose 

Suleyman Demirel Universitesi, Computer Engineering, Suleyman Demirel University, Dept. of Computer Engineering, Faculty of Engineering, West Campus, E9 Block, Z-23, 32260, Isparta, Turkey, 32260

Date of review: 18 January 2020

Published online:

© The Author(s) 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence <http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conflict of interest statement. Reviewer declares none.

Comment

Comments to the Author: This paper is based on a research regarding use of deep learning for automatic fall detection of elderly-people. In detail, the detection is done over videos and the paper considers the obtained findings as well as the development flow in this manner. From a general perspective, the paper has an interesting, important topic and comes with a pure, technically-enough content. So, it seems acceptable. I only suggest a few minor revisions as final touches:

- 1- Please support the first paragraph of the Introduction section, with one or two more references. That will make more sense in terms of starting point of the research.
- 2- After starting to the 2nd paragraph of the Introduction section with “According to [2]...”, the sentence should be ended as “...devices [3-5].” by taking citations at the end of the sentence.
- 3- In Fig. 3, please provide the graphics in bigger sizes, (from up to down) in order to improve readability.
- 4- Is there any future work(s) planned? That should be indicated at the last section briefly. Thanks the author(s) for their valuable efforts to form this paper-research.

Score Card

Presentation



Is the article written in clear and proper English? (30%)

3/5

Is the data presented in the most useful manner? (40%)

4/5

Does the paper cite relevant and related articles appropriately? (30%)

4/5

Context



Does the title suitably represent the article? (25%)

3/5

Does the abstract correctly embody the content of the article? (25%)

4/5

Does the introduction give appropriate context? (25%)

4/5

Is the objective of the experiment clearly defined? (25%)

4/5

Analysis



Does the discussion adequately interpret the results presented? (40%)

4/5

Is the conclusion consistent with the results and discussion? (40%)

4/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)

4/5

Review 2: Fall detection for elderly-people monitoring using learned features and recurrent neural networks

Reviewer: Dr. Leonel Aguilar 

ETH Zurich Department of Computer Science, GESS, Zurich, Switzerland, 8092

Date of review: 13 January 2020

Published online:

© The Author(s) 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence <http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conflict of interest statement. Reviewer declares none

Comment

Comments to the Author: I found the experiments and results interesting but its description, motivation, and support for the novelty claim were lacking.

Description

A detailed description of the experimental setup, the data used and the challenges it presents should be incorporated. Although the relationship of the two neural networks can be guessed from the reported tensor dimensions this information should be clearly stated. Figure 1. lists the components but doesn't present their relationship or details of the experimental setup. For example, it is not clear how a Bi-LSTM trained to predict 101 categories of the pre-train-data is later fine-tuned to predict 2 categories.

Motivation

Although the motivation to use of VGG16 for feature extraction can be guessed by the reader, the author's should justify this choice. In general, LSTMs are justified over long sequences but in this experiment, only 7 frames are used. No comparison to alternative methods or configurations were presented for the LSTM.

Novelty

Stronger evidence in terms of the novelty of the experimental setup or its results should be made. In the introduction, the authors show alternative approaches to detect human falls but do not provide any point of comparison, e.g. in terms of recall scores or robustness against noise.

Misc

The data and code should be publically available, not "under reasonable request".

While there should be an explanation on how to reproduce your results/use your code, trivial information on how to clone a git repository should be avoided.

Avoid half-page information sparse figures.

Score Card

Presentation



Is the article written in clear and proper English? (30%)

2/5

Is the data presented in the most useful manner? (40%)

1/5

Does the paper cite relevant and related articles appropriately? (30%)

4/5

Context



Does the title suitably represent the article? (25%)

5/5

Does the abstract correctly embody the content of the article? (25%)

4/5

Does the introduction give appropriate context? (25%)

4/5

Is the objective of the experiment clearly defined? (25%)

5/5

Analysis



Does the discussion adequately interpret the results presented? (40%)

4/5

Is the conclusion consistent with the results and discussion? (40%)

4/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)

1/5
