

Detection of Falls from Non-Invasive Thermal Vision Sensors Using Convolutional Neural Networks [†]

Javier Medina Quero ^{1,*}, Matthew Burns ², Muhammad Asif Razzaq ³, Chris Nugent ² and Macarena Espinilla ¹

¹ Department of Computer Science, University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain; mestevez@ujaen.es

² School of Computing, Ulster University, Newtownabbey, Co. Antrim, Northern Ireland BT15 1ED, UK; Burns-M19@ulster.ac.uk (M.B.); cd.nugent@ulster.ac.uk (C.N.)

³ Ubiquitous Computing Lab in Kyung Hee University, Seoul 446-701, Korea; asif.razzaq@oslab.khu.ac.kr

* Correspondence: jmquero@ujaen.es; Tel.: +34-953-212897

[†] Presented at the 12th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI 2018), Punta Cana, Dominican Republic, 4–7 December 2018.

Published: 24 October 2018

Abstract: In this work, we detail a methodology based on Convolutional Neural Networks (CNNs) to detect falls from non-invasive thermal vision sensors. First, we include an agile data collection to label images in order to create a dataset that describes several cases of single and multiple occupancy. These cases include standing inhabitants and target situations with a fallen inhabitant. Second, we provide data augmentation techniques to increase the learning capabilities of the classification and reduce the configuration time. Third, we have defined 3 types of CNN to evaluate the impact that the number of layers and kernel size have on the performance of the methodology. The results show an encouraging performance in single-occupancy contexts, with up to 92% of accuracy, but a 10% of reduction in accuracy in multiple-occupancy. The learning capabilities of CNNs have been highlighted due to the complex images obtained from the low-cost device. These images have strong noise as well as uncertain and blurred areas. The results highlight that the CNN based on 3-layers maintains a stable performance, as well as quick learning.

Keywords: thermal vision sensor; fall detection; convolutional neural networks

1. Introduction

The global population of those 60 years old or above exceeded 962 million in the year 2017, which is more than twice as large as it was in 1980. This number of elderly persons is expected to double again by 2050 worldwide. As the average age of the population continues to rise [1], elderly people are continuing to suffer from certain chronic diseases like dementia, hypertension, diabetes, gait issues, etc. Therefore, it has become important to address the needs and interests of older persons, specifically needs related to health-care and other forms of daily life. In an indoor environment, impact related accidents such as falling and collisions have been identified and studied in an attempt to avoid falls or reduce the aid response time [2]. In 2012–2013, 55% of all unintentional injury deaths among adults aged 65 and over were due to falls [3]. This opens up the research avenue for tracking elderly people's daily routine related activities, specifically those that guarantee the safety. Over the last decade, interest in wearable and mobile ubiquitous computing technologies has provided researchers with enough opportunities to design monitoring and intervention platforms, which could provide continuous 24/7 real-time monitoring in a *smart-home* environment.

To date, significant advances have been made towards reliable recognition of various activities in highly instrumented *Smart-home* environments [4], as people spend most of their time indoors.

These aforesaid environments mostly include elaborate and augmented object/location schemes, with a broad range of appliances and wearable sensors such as acceleration sensors, RFID reader etc. In contrary, human activity recognition in *smart-environment* with limited augmentation is possible with minimal or no use of wearable sensors, but it is still an open research issue.

These methods of activity recognition most often involve the use of a camera. A camera's usage, however, invades the user's privacy, as it is deployed 24/7 to monitor the inhabitant in the *smart-home* environment. In order to increase comfort for the inhabitant, minimize the operation cost and simplify the use of technologies, non-wearable sensors are introduced. Keeping these issues in mind, different types of non-wearable and non-intrusive human identification sensors are deployed in the *smart-home* environment. The maturity of non-wearable activity recognition, however, is not high and most of these technologies are verified only in a *smart-lab* environment. It is evident that falling is a big problem for elderly people in nursing care, as a fall or accident could happen in any living environment such as a room, corridor or wash-room. Most importantly around 58% of falls occur in the living room for the elderly [2]. Therefore, even with the mentioned privacy concerns, 24/7 monitoring is required for inhabitant tracking.

This work proposes detection of falls using Convolutional Neural Networks (CNNs) by employing a non-invasive thermal vision sensor (TVS) in a real-environment. In the proposed methodology, we collect occupancy frames using a single TVS attached to the ceiling. The fall situation and the non-fall situation is detected with a CNN by using augmented posture transitions.

The paper is organized as follows: Section 1 gives an overview of the state-of-the-art non-intrusive sensor categories. Section 2 describes the architecture of the components, and proposed methodology to detect falls. Section 3 discusses the experimental set-up in a *smart-room*. Section 4 shows the results from the 3 types of CNN, as well as the performance evaluation and a detailed discussion. Section 5 provides the concluding remarks.

Related Works

The use of smart homes for different applications such as health and activity recognition is rapidly increasing worldwide [5] for the elderly. Different sensing technologies are used around the home to monitor the activities of the inhabitants. Based on these activities and behaviours, personalized services are provided to them. In all cases, the combinations of events needed to recognize such activities require the implementation of several, often complex sensors around the inhabitant. An example is a camera, which can, in principle, provide all the information from a single source. While there has been a lot of work on video-based activity recognition in household environments [6], it infers privacy concerns. In the last decade, significant work has been done in this area to identify different sets of activities using other non-wearables. Some approaches consider using environment placed sensors, such as cameras, PIR (pyroelectric) [7], thermal sensors [8], and others. Some techniques also considered smart floor and pressure sensors. The research community, however, has suggested that non-wearable technologies can generate alerts of elapsed sedentary time [9], gait deformities, falls and collisions, which could all be detected and avoided in our daily routines without privacy issues. Non-wearable human context identification technologies are classified into four main categories, object-based, footstep-based, body-shape, and gait-based identification technologies.

In the context of fall detection, in [10] authors present a wearable based solution, which was located on the waist. It collects data from both a triaxial accelerometer and gyroscope to classify several human actions, including three types of falls: falling forward, falling to the left and falling to the right. In [11] authors describe a fall detection system using a wearable sensor, which was optimized to determine the optimal transmission and sampling rates.

In [12] authors integrate a thermal infrared array sensors to detect falls computing the maximal thermal difference between the background and foreground pixel. The configuration was sensible to room temperature and brightness. In [13] authors proposed a system used to recognize human activities, which include falls, by means of a single thermal infrared sensor. Several features based

on temperature thresholds are proposed to be evaluated by a Support Vector Machine (SVM) in the classification. It achieved a good performance but including some confusion in distinguishing between the falling and sitting.

Based on previous work, the proposed methodology aims to:

- Integrate a low resolution thermal vision *Heimann HTPA 32 × 31* [8].
- Offers a degree of privacy not offered by traditional camera methods as occupants cannot be identified, providing a more positive acceptance than standard vision cameras [14].
- Avoid that the inhabitant carries a wearable sensor or external device. The wearable based approaches involve a maintenance and battery charge of the external device.
- Avoid the evaluation by on human defined features based on thresholds and parameters using a quick configuration with data augmentation and Convolutional Neural Networks.

2. Methodology

In this section, we detail the methodology proposed to detect falls from non-invasive TVSSs. First, we detail an agile data collection to label images in order to compose a dataset that describes several cases of normal occupancy with standing inhabitants and target situations with a fallen inhabitant. Second, we detail the data augmentation techniques included in the approach, to increase the learning capabilities of the classification. Third, we have defined 3 types of CNN to evaluate the impact that the number of layers and kernel size has on the performance of the methodology.

In Figure 1 we show the architecture of components described in the next Sections.

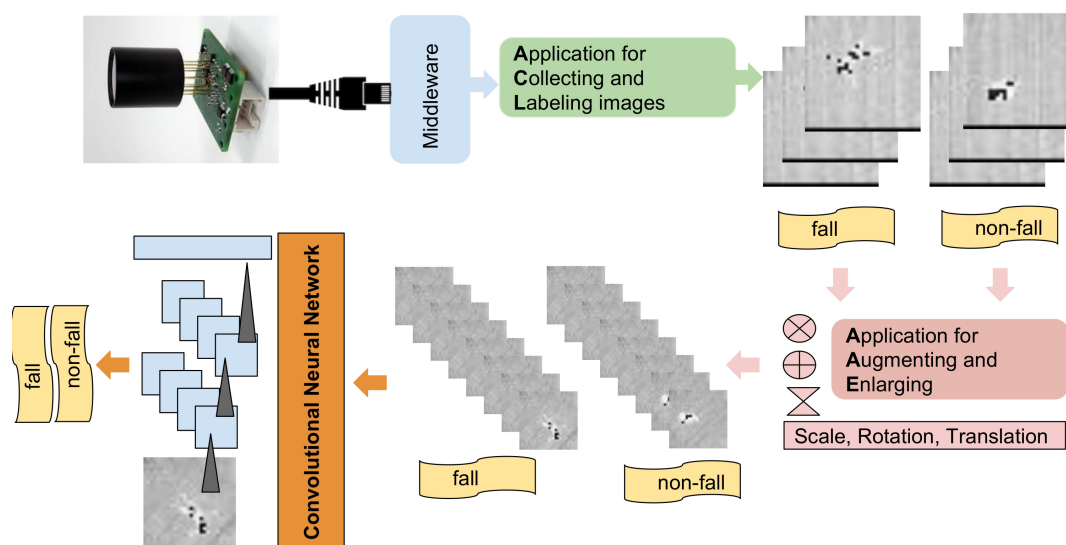


Figure 1. Architecture of components in methodology.

2.1. Agile Data Collection from Thermal Vision Sensor

In this section we describe a agile process of data collection and the processing of images in order to optimize the learning of CNN by means of data augmentation.

First, we note the TVS is located in the room’s ceiling, collecting an zenithal view of the occupants, which offers the advantage of (i) providing a clear view of fall situations and (ii) reducing occlusion in multi-occupancy. For each image, the data collected by the TVS (*Heimann HTPA 32 × 31*) is defined by a 32×31 matrix, where each pixel defines a heat point whose value is defined within the range of 0 and 255. The image data is collected from the TVS by means of a twisted ethernet cable which is connected to the local area network. The middleware *SensorCentral* [15] integrates the TVS as a sensor source, providing the data within a Web Service in JSON format.

Second, we have developed an application for collecting and labeling images (ACL), which (i) connects to the Web Service from the TVS, in order to collect data in real-time, and (ii) stores the data as an image in PNG format within a labeled local folder. We note that due to the TVS proposed being a low-cost sensor, the data collection rate is limited to a range of 1–2 images per second. The use of ACL enables developers to straightforwardly generate a dataset, from a case scene, by means of human-supervised labeling. In Section 3 we describe several case studies based on single or multi-occupancy.

2.2. Data Augmentation

Learning from a CNN needs a large amount of data [16]. Collecting images from different inhabitants, orientations and cases by the ACL could take a great deal of effort, which could make the customization and configuration in different contexts hugely difficult. Fortunately, the data augmentation provides a sort solution to enlarge the number of learning cases from a limited set [17] and therefore, reducing the over-fitting [18].

We have developed an application for augmenting and enlarging (AAE) the image data from the original dataset, collected by ACL by means of translation, rotation and scale transformations:

- *Translation.* The original image is relocated within a maximal window size $[t_x, t_y]^+$ using a random process which generates a random translation transformation $[t_x, t_y], t_x \in [0, t_x^+], t_y \in [0, t_y^+]$.
- *Rotation-Scale.* The rotations are provided by two methods. First, the translated image is flipped horizontally and vertically, using a random process that applies the transformation to a percentage of cases, defined by wH, wR respectively. Second, a rotation and scale transformation is defined by a maximal rotation angle α^+ and a scale factor s^+ , which generates a random rotation with an angle $\alpha \in [0, \alpha^+]$ and a random scale $s \in [1 - s^+, 1 + s^+]$ which are then applied in the center of the image. We note that this rotation overcomes the original image size, also providing a random scale of the image.
- *Crop-scale.* A final image of window size $[s_x, s_y]$ is cropped from the center of the image, which contains the most relevant visual information.

2.3. Design of the Convolutional Neural Network

In this section we describe 3 types of CNNs (CNN) for classifying the fall situations of inhabitants. There are two target classes for the CNN to learn: fall situation and non-fall situation. As we discussed in the previous Section 2.2, the augmented dataset provides a wide range of images of size 28×28 for the two cases.

Based on the visual information of the images, the CNN learn kernels are related to key visual patterns, whose similarity with source visual regions is defined by a convolutional operation. They enable summarizing and reducing of the spatial information defined by a bi-dimensional kernel size $[K_x, K_y]^l$. Moreover, the CNN is configured within layers that are hierarchically ordered $L = \{l_1, \dots, l_L\}$. In addition, some methods/sub-layers are usually included to configure the network, reduce the number of parameters and avoid over-fitting, such as, *dropout*, *pooling* or *activation functions*.

Next, we briefly discuss the functionality of such methods/sub-layers within the layers of the CNN:

- *Convolutional kernels.* They represent the main contribution in the CNN. The kernels (also called filters) define a pattern which must be learned within a local visual region. The convolution operation of a local visual region and a kernel generate an output which represents its similarity. The convolutional sub-layer is defined by the size of the kernels $K_i^l = [K_x, K_y]^l$ and the number of them $|K^l|$ in a layer l .
- *Activation function.* The outputs of kernels within layers can be filtered by an activation function A_l to train the neural network several times faster [18]. Rectified linear unit (Re-LU) is a simple and encouraging function widely applied.

- Pooling. This sub-layer provides a spatial reduction in the size of the outputs within the layers, to control over-fitting [17]. It is defined by a window size $[P_x, P_y]$ and aggregation function P_1^+ , such as maximal or average.
- Dropout. In the last layer, reducing the network by randomly removing nodes [19] reduces over-fitting. The random punning is controlled by a probability threshold d_α .
- Dense Layer. Finally, the output from last CNN layer configures a dense layer of features whose weights describe the input image. A dense layer is introduced as a classifier or regressor that learns the relation between all of the weights from the last layer, and the labeled target class. The classical classifier is a multilayer perceptron, with a final activation function A^+ , such as *soft-max* or *sigmoid* functions. This relates the weights of the last layer to a probability distribution for each possible target class.
- Cost function and back-propagation. The lost/cost function C^* defines the performance of the learned model while training. Several cost functions have have been proposed, such as, *quadratic cost* or more recently, *cross-entropy* [20] in vision recognition. The cost function is computed by the optimization of the weights of neurons within the back-propagation process. In this approach, the gradient descent optimization algorithm [21] is commonly used in back-propagation, to optimize the network by calculating the gradient of the cost function.

In this work, three types of configurations of CNN are evaluated to classify the fall detection from vision thermal images: (i) CNN_2^0 a 2-layer CNN with optimized configuration for MINIST dataset [22], (ii) CNN_2^+ 2-layer CNN with new kernels configuration and (iii) CNN_3 a 3-layer CNN. Based on the previous definitions we have described them in Table 1. The performance of the CNNs is shown in next Section 4.

Table 1. Configurations of Convolutional Neural Networks.

Parameter	CNN_2^0	CNN_2^+	CNN_3
$ L $	2	2	3
$[K_x, K_y]^1 \times K^1 $	$[5,5] \times 16$	$[3,3] \times 16$	$[3,3] \times 16$
A^1	Re-Lu	Re-Lu	Re-Lu
$[P_x, P_y], P^1$	$[2,2], \max$	$[2,2], \max$	$[2,2], \max$
$[K_x, K_y]^2 \times K^2 $	$[5,5] \times 16$	$[5,5] \times 64$	$[5,5] \times 32$
A^2	Re-Lu	Re-Lu	Re-Lu
$[P_x, P_y], P^2$	$[2,2], \max$	$[2,2], \max$	$[2,2], \max$
$[K_x, K_y]^3 \times K^3 $			$[7,7] \times 16$
A^3			Re-Lu
d_α	0.5	0.5	0.5
A^+	soft max	soft max	soft max
C^*	cross-entropy	cross-entropy	cross-entropy

3. Experimental Setup

In this section, we detail the experimental setup of the case study that has been developed to evaluate the methodology for fall detection in a smart environment.

The case study has been carried out in the Smart Lab of the University of Ulster (<https://www.ulster.ac.uk/research/institutes/computer-science/groups/smart-environments>). Three participants (one woman, two men) between 25 and 35 years old were involved to collect data in the hall room, using a TVS installed in the ceiling. The height of the participants were 1.72, 1.68 and 1.83 m.

The data to detect falls in *single occupancy* composed of three subcategories: (i) empty room, (ii) one person standing/walking, and (iii) one person fallen. In *multi occupancy*, we added two new subcategories: (iv) 2–3 people standing/walking and, (v) one person fallen with another person standing/walking. The image from the three participants was collected in a static way due to the purpose of the work is based on evaluating an image. The vision of the TVS within the hall room is defined by a square bounding box of 3.5 m. When collecting data, each person simulated several

natural positions in case of fall, as well as, took a walk around the vision area of the TSV in case of walking.

As we have introduced, one of the main motivations of the work is providing a quick configuration using data augmentation. So, the duration for each subcategory was 4 min and limited to 260 frames. This allows for the methodology to provide a quick configuration and agile deployment in several contexts.

Data Augmentation

Once data from the original dataset is collected, we include a data augmentation step for enlarging the visual information of images, which increases the learning performance of the CNN.

- *Translation.* The original image is translated within a maximal window size $[t_x, t_y]^+ = [3, 3]$.
- *Rotation-Scale.* Each image is flipped horizontally and vertically by a random probability $wH = 0.5, wR = 0.5$ respectively, i.e., horizontally in half of the cases, and vertically in the other half. Second, a rotation and scale transformation is defined by a maximal rotation angle $\alpha^+ = \pi$ and scale $s^+ = 0.1$. We note both rotation methods generate a random rotation from the complete range of possible orientations.
- *crop-scale.* In this work, we are working on a final centered image with a window size of 28 pixels $[s_x, s_y] = [28, 28]$.

In Figure 2, we show an example of augmented images generated from the same image collected by the TVS.

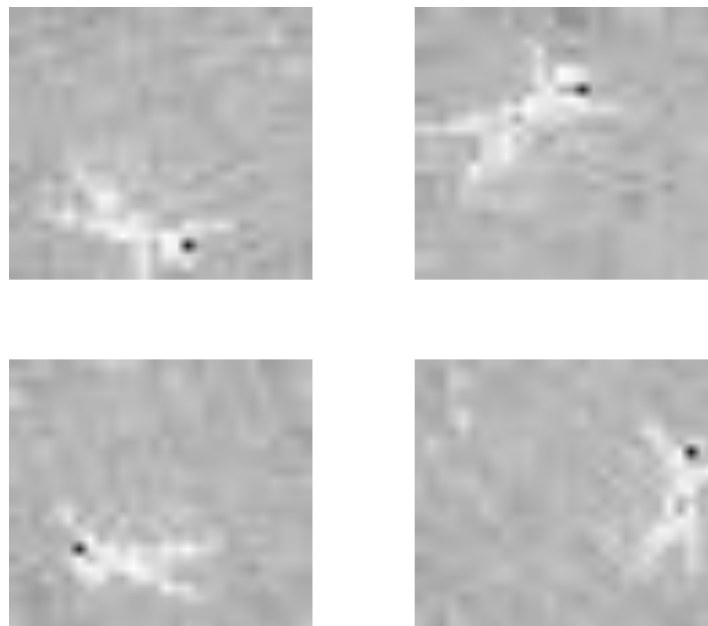


Figure 2. Example of data augmentation applied to the image from the thermal vision sensor, where a fall is presented.

4. Results

In this section, we detail the results from the 3 types of CNNs, to detect falls from thermal vision images. In order to evaluate the dataset, it has been divided into 10% for testing and 90% for training using cross-validation. So, 10 training and testing datasets have been further evaluated. The accuracy and time thrown by the 3 types of CNNs have been collected for over 2000 learning steps.

In Table 2, we include the data for the *single occupancy* dataset. In Table 3, we include the data for the *multi occupancy* dataset. The results from 5 datasets for both single and multi-occupancy, provide

(i) the average and standard deviation of accuracy from the last 20 learning steps, and (ii) the average of the time wasted in the 2000 steps. We include the average of accuracy from the last 20 learning steps due to the variance and evolution of the performance in Deep Learning while the training. In Figure 3 we describe the evolution of accuracy in the learning process of while the 2000 steps from the 3 types of CNNs.

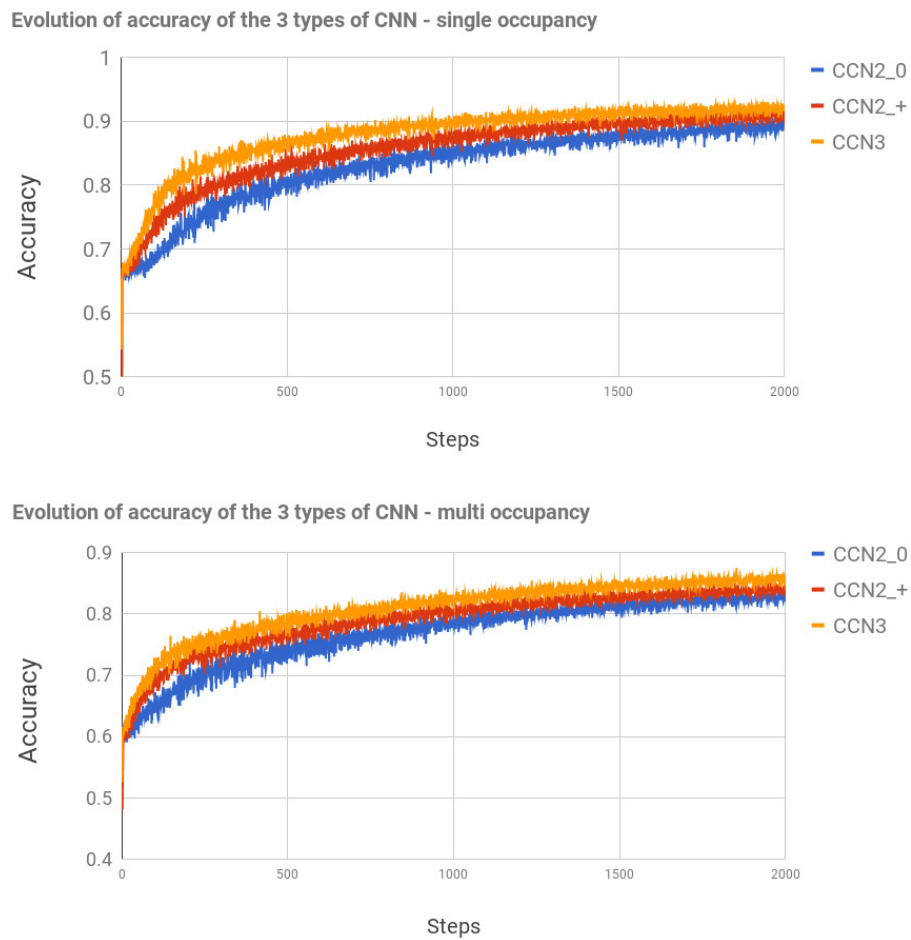


Figure 3. Evolution of accuracy in the learning process for single and multi occupancy from the 3 types of CNNs.

Table 2. Table summarizing the results of the single occupancy data.

	$\Sigma(\text{Time})$	$\Sigma(\text{Accuracy})\%$
CNN_2^0	22.48 m	89.31
CNN_2^+	30.36 m	90.82
CNN_3	27.86 m	91.93

Table 3. Table summarizing the results of the multiple occupancy data.

	$\Sigma(\text{Time})$	$\Sigma(\text{Accuracy})\%$
CNN_2^0	22.53m	83.87
CNN_2^+	30.36m	84.01
CNN_3	28.00m	85.70

Discussion

Before discussing the results, we note the complexity of recognizing falls from the low-cost device studied here. The most important issue is the precision of the thermal points, which configure the image of the TVS. It includes strong amounts of noise and uncertain and blurring areas between hot points. This can generate problematic images in detecting falls. An example is included in Figure 4, which summarizes the complexity of the problem that the current approach has been faced with.

First, the CNNs have provided an encouraging performance in detecting falls in single occupancy, which achieves 90% accuracy. The use of different kernel sizes in 2-layers of CNN in CNN_2^0 and CNN_2^+ provides a slight improvement in accuracy, and the second version CNN_2^+ wastes a significant amount of time on learning, >35% compared to CNN_2^0 . The use of a CNN based on 3-layers has improved the accuracy in detecting falls up to 91.93%, without requiring an increase in time spent on learning >24%. Finally, CNN_2^+ and CNN_3 have shown to learn faster than CNN_2^0 , which could reduce the learning time in real-time deployments and configurations.

Second, the multi-occupancy scenes have shown a more complex visual configuration, which has reduced the accuracy to 85%. As we detail in Figure 4, it is due to the integration of more people, which produces similar shapes as a fallen person. The use of different kernel sizes in 2-layers of CNN in CNN_2^0 and CNN_2^+ has provided a slight increase in accuracy. We note that the use of a CNN based on 3-layers has again improved the accuracy in detecting falls +2%, producing the same increase in time spent on learning, >24%, as single occupancy. Similar to single occupancy, CNN_2^+ and CNN_3 have shown quicker learning than CNN_2^0 .

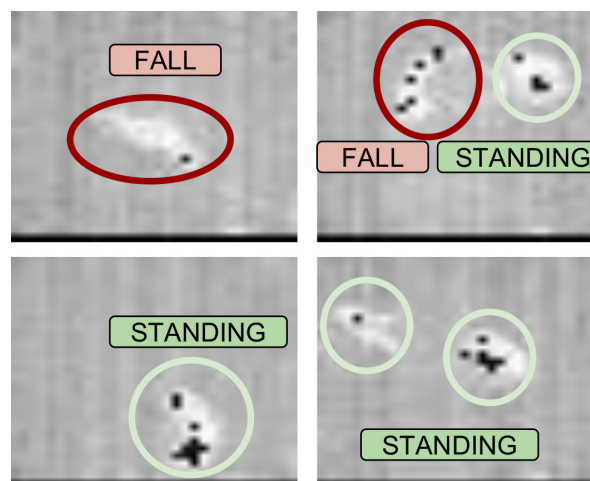


Figure 4. Example of problematic images to detect falls from thermal vision sensor. In Top, two falls are presented. Bottom, standing/walking people is collected.

In summary, the use of a CNN, specially based on 3-layers, has provided a robust performance in detecting falls from thermal vision images.

5. Conclusions and Ongoing Works

The CNNs have been evaluated as suitable detectors of inhabitant falls from the images collected by a TVS. In order to solve the problem of collecting a huge amount of data, we have included data augmentation to provide a quick configuration.

The results have shown an encouraging performance in single-occupancy contexts with up to 91% accuracy, and a 6% reduction in accuracy for multi-occupancy. The learning capabilities of CNNs have been highlighted due to the complex images obtained with strong amounts of noise and uncertain and blurring areas from the low-cost device used here. The results highlight that the CNN based on 3-layers maintains a stable performance, as well as quick learning.

In future works we will include a robust system where the temporal evaluation of the stream frames from a TVS are analyzed. A great number of inhabitants will be involved to increase the soundness of the recognition. Therefore, the probability of fall detection from TVS, in the methodology described here, will be increased as richer information will be involved. Such information includes movement, speed and tracking information from a sequence of frames, to provide a competitive detection of falls based on the context of the environment and inhabitants.

Moreover, for solving the problems of resolution and precision of *Heimann HTPA 32 × 31*, other recent mobile TVSS, such as FLIR (<https://www.flir.com/products/flir-one-gen-3/>) will be evaluated to obtain a higher quality in the images collected, which could improve the performance in the detection of falls. It will also include a visible spectrum camera to provide a comparative performance between thermal and non-thermal vision sensors.

Author Contributions: Conceptualization, C.N. and J.M.; Methodology, M.E. and J.M.; Investigation: J.M., M.B., M.A.R., C.N. and M.E.; Funding Acquisition, C.N. and M.E.

Acknowledgments: This research has received funding under the REMIND project Marie Skłodowska-Curie EU Framework for Research and Innovation Horizon 2020, under Grant Agreement No. 734355. The authors would also like to thank the Connected Health Innovation Centre (CHIC). Also, this contribution has been supported by the project PI-0203-2016 from the Council of Health for the Andalusian Health Service, Spain together the research project TIN2015-66524-P from the Spanish government.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CNN	Convolutional Neural Networks
TVS	Thermal Vision Sensor

References

1. World Health Organization. World Population Ageing Report. Available online: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf (accessed on 7 November 2017).
2. Khojasteh, S.B.; Villar, J.R.; Chira, C.; Gonzalez, V.M.; de la Cal, E. Improving fall detection using an on-wrist wearable accelerometer. *Sensors* **2018**, *18*, 1350.
3. Kramarow, E.A.; Chen, L.H.; Hedegaard, H.; Warner, M. *Deaths from Unintentional Injury Among Adults Aged 65 and Over, United States, 2000–2013 (No. 2015)*; US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics: Washington, DC, USA, 2015.
4. Amiribesheli, M.; Benmansour, A.; Bouchachia, A. A review of smart homes in healthcare. *J. Ambient Intell. Humanized Comput.* **2015**, *6*, 495–517.
5. Gjoreski, H.; Gams, M. Person Identification by Analyzing Door Accelerations in Time and Frequency Domain. In *European Conference on Ambient Intelligence*; Springer: Cham, Switzerland, 2015; pp. 60–76.
6. Rohrbach, M.; Amin, S.; Andriluka, M.; Schiele, B. A database for fine grained activity detection of cooking activities. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1194–1201.
7. Raykov, Y.P.; Ozer, E.; Dasika, G.; Boukouvalas, A.; Little, M.A. Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 1016–1027.
8. Heimann. 2017. Available online: http://www.heimansensor.com/products_imaging.php (accessed on 7 November 2017).
9. Synnott, J.; Rafferty, J.; Nugent, C.D. Detection of workplace sedentary behavior using thermal sensors. In Proceedings of the 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 5413–5416.

10. Nari, M.I.; Suprpto, S.S.; Kusumah, I.H.; Adiprawita, W. A simple design of wearable device for fall detection with accelerometer and gyroscope. In Proceedings of the International Symposium on Electronics and Smart Devices (ISESD), Bandung, Indonesia, 29–30 November 2016; pp. 88–91.
11. Gia, T.N.; Sarker, V.K.; Tcarekko, I.; Rahmani, A.M.; Westerlund, T.; Liljeberg, P.; Tenhunen, H. Energy efficient wearable sensor node for IoT-based fall detection systems. *Microprocess. Microsyst.* **2018**, *56*, 34–46.
12. Hayashida, A.; Moshnyaga, V.; Hashimoto, K. The use of thermal ir array sensor for indoor fall detection. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 594–599.
13. Mashiyama, S.; Hong, J.; Ohtsuki, T. Activity recognition using low resolution infrared array sensor. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 495–500.
14. Sixsmith, A.; Johnson, N. A smart sensor to detect the falls of the elderly. *IEEE Pervasive Comput.* **2004**, *3*, 42–47.
15. Rafferty, J.; Synnott, J.; Ennis, A.; Nugent, C.; McChesney, I.; Cleland, I. SensorCentral: A Research Oriented, Device Agnostic, Sensor Data Platform. In *International Conference on Ubiquitous Computing and Ambient Intelligence*; Springer: Cham, Switzerland, 2017; pp. 97–108.
16. Yamashita, T.; Watasue, T.; Yamauchi, Y.; Fujiyoshi, H. Improving Quality of Training Samples Through Exhaustless Generation and Effective Selection for Deep Convolutional Neural Networks. *VISAPP* **2015**, *2*, 228–235.
17. Ciresan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. High-Performance Neural Networks for Visual Object Classification. *arXiv* **2011**, arXiv:1102.0183.
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Stateline, NV, USA, 2012; pp. 1097–1105.
19. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
20. Robert, C. Machine learning, a probabilistic perspective. *Chance* **2014**, doi:10.1080/09332480.2014.91476.
21. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533.
22. LeCun, Y. The MNIST Database of Handwritten Digits. Available online: <http://yann.lecun.com/exdb/mnist/> (accessed on 19th October, 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).