Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-08812-x

Elderly fall detection based on multi-stream deep convolutional networks

Chadia Khraief [1] · Faouzi Benzarti [1] · Hamid Amiri [1]

Received: 28 February 2019 / Revised: 21 December 2019 / Accepted: 28 February 2020
Published online: 25 March 2020

Abstract
Fall is the biggest threat to seniors, with significant emotional, physical and financial implications. It is the major cause of serious injuries, disabilities, hospitalizations and even death especially for elderly people living alone. Timely detection could provide immediate medical service to the injured and avoid its harmful consequences. Great number of vision-based techniques has been proposed by installing cameras in several everyday environments. Recently, deep learning has revolutionized these techniques, mostly using convolutional neural networks (CNNs). In this paper, we propose weighted multi-stream deep convolutional neural networks that exploit the rich multimodal data provided by RGB-D cameras. Our method detects automatically fall events and sends a help request to the caregivers. Our contribution is three-fold. We build a new architecture composed of four separate CNN streams, one for each modality. The first modality is based on a single combined RGB and depth image to encode static appearance information. RGB image is used to capture color and texture and depth image deals with illumination variations. In contrast of the first feature that lacks the contextual information about previous and next frames, the second modality characterizes the human shape variations. After background subtraction and person recognition, human silhouette is extracted and stacked to define history of binary motion HBMI. The last two modalities are used to more discriminate the motion information. Stacked amplitude and oriented flow are used in addition to stacked optical flow field to describe respectively the velocity, the direction and the

✉ Chadia Khraief
chadiaKhraief@gmail.com

Faouzi Benzarti
benzartif@yahoo.fr

Hamid Amiri
hamidlamiri@gmail.com

[1] SITI Laboratory, National Engineering School of Tunis (ENIT), University of Tunis El Manar, Tunis, Tunisia

motion displacements. The main motivation behind the use of these multimodal data is to combine complementary information such as motion, shape, RGB and depth appearance to achieve more accurate detection than using only one modality. Our second contribution is the combination of the four streams to generate the final decision for fall detection. We evaluate early and late fusion strategies and we have defined the weight of each modality based on its overall system performance. Weighted score fusion is finally adopted based on our experiments. In the third contribution, transfer learning and data augmentation are applied to increase the amount of training data, avoid over fitting and improve the accuracy. Experiments have been conducted on publicly available standard datasets and demonstrate the effectiveness of the proposed method compared to existing methods.

Keywords Fall detection · Elderly people · Smart home · Video surveillance · Deep learning · Multi stream CNN · RGB-depth cameras · Kinect cameras · Optical flow · Motion of binary motion images · Data augmentation · Transfer learning

1 Introduction

The elderly population is increasing dramatically over the world. In China, 10.1% of the total population are people over 65 and it will be raised to about 35% by 2050 [46]. In Tunisia, the proportion represented 9.8% in 2009 and it will be 17.7% in 2029 and around one from five will be over 60 years of age in 2039 [21]. With this fast growth of elder population, more and more seniors want to live alone [40]. Indeed, 90% of the seniors still live in their homes in Canada and approximately one third in Europe live alone. Unfortunately, unexpected situations like falls might happen suddenly and influence their health, security and well-being. Nearly one-third of adults aged over 65 reports a fall every year in United States and the annual cost of falls is about $31 billion [7]. They can't alert anyone for help particularly if they were unconscious and had serious injuries sustained [24]. It was reported that 50% of the elderly who lay on the floor for more than one hour after falls died within six months after the accident [45]. Moreover, fear of falling and not receiving instant medical service can limit the activities of elderly, resulting in social isolation, depression and helplessness [5]. In other hand, the long-term nursing care at home is very expensive [7]. Therefore, fall is the biggest threat to elderly people, with significant emotional, physical and financial implications. It is considered a major health concern especially for those living alone. For the above reasons, it is essential to develop intelligent system that can automatically detect falls, assist elderly people and help them to live safely and independently in their home. In other hand, it will reduce the cost and the time required for caregivers to intervene avoiding its harmful consequences.

In recent years, related research of fall detection systems has increased remarkably with the rapid development of new smart sensors and technologies that can extract different kinds of information from the environment. According to the kind of sensors used, fall detection methods can be divided into three main categories [33], namely, (i) wearable sensor-based, (ii) ambient sensor-based, and (iii) vision-based. Wearable sensor-based methods use specialized devices attached to human body for detecting any variation of

his activities. There are many wearable sensors used for fall detection such as accelerometer [41], gyroscope, smart watch [31] or fusion of them [8]. Unfortunately, elderly people frequently forget to wear them particularly those with cognitive impairments. There are around 35.6 million dementia patients worldwide and it will be the double by 2030 and more than triple by 2050 [52]. Ambient sensor-based systems exploit vibration or pressure devices installed under the bed or in floor to analyze the sound and the vibrations. These sensors are not expensive and do not disturb the elderly person. Their main shortcomings are low fall detection accuracy [13] and tendency to generate more false alarms. Finally, vision-based methods use one or more cameras to detect falls. These cameras are installed in several everyday environments and provide very rich information about persons and their activities. Fall detection is very challenging in the real-world environment due to light changing, illumination, occlusions, shadows, pose changes, etc. To overcome these uncontrolled conditions, several fall detection methods are proposed. Zang et al. [59] reviews recent works based on single RGB cameras, multiple RGB cameras, and depth cameras. The emergence of the depth camera, overcomes some weakness of RGB cameras. In fact, the depth camera is invariant to color and texture changes and insensitive to illumination variations. It is reliable for estimating body silhouette and skeleton, and reconstructs 3D structural information of the object. Thus, depth information provides a significant contribution to handle several challenges related to fall detection's environment. But, depth images suffer also from distance limitation. Recently, Kinect sensor has been proposed. In addition to its low-cost device, Kinect sensor combines an RGB camera to a depth sensor. Based on a comparative study of different RGB and Depth cameras, Xu et al. [53] has indicated that Kinect has replaced the RGB camera and became the most promising type of sensor used in fall detection systems after 2014.

Among the above three methods, the vision-based method using Kinect sensor is considered in this study. Inspired by the fast sensors progress, our proposed method use both RGB and depth image to detect falls. The main characteristic of RGB data is its shape, color and motion information. These modalities are used in addition to depth information to get more discriminative description of the fall event and perform an accurate detection. Cappitelli et al. [11] survey recent works based on RGB-Depth sensors and classify different methods into two categories based on the data exploited for fall detection, namely methods using only depth data, and methods based on multiple data fusion. All these methods follow the same approach that is the extraction of the best discriminating features and then the choice of the best classifier such as SVM or KNN. Despite the effectiveness of these methods specially with fusion of multiple data, it is very difficult to take into account when designing such as features all the complex fall models, the similar or confounding daily life activities such as lying or sitting down on a sofa. More recently, deep learning has shown very high classification accuracies compared to hand-crafted features on various computer vision tasks such as object detection, action recognition and specially fall detection. A deep learning schema based on depth information is proposed by [15] to distinguish humans from the background and to deal with dynamic variations of environment such as shadows, illumination. PCAnet is used by [49] to extract features from color images and then applied a SVM to classify activities. Adrian et al. [1] implemented a CNN for fall detection based on optical flow displacements and has obtained satisfactory results.

Adhikari et al. [2] fed convolutional neural networks by a single combined subtracted RGB and depth image. These methods are based only in one modality color, shape or motion. But, the fusion of these modalities has also been proven successful at carrying out several tasks on multi-modal data. For example, Simonyan et al. [43] combined RGB and motion features extracted by the CNN for activity recognition and demonstrated that better performance was obtained compared to other networks with single modality. This fusion performed better than individual features. Many tasks that benefit from multiple modalities can be found in the recent deep learning literature such as video classification [54], emotion recognition [47], etc.

Inspired by the promising results of these methods, we propose a multi-stream CNN architecture which exploits the full multi-modal information provided by RGB-D cameras such as color, depth, shape and motion. The main motivation behind the use of multi-modal data is to benefit from the complementary information of different modalities and improve the overall system performance better than using only one. But, these modalities didn't have the same contribution. For example, as proven by [43] motion modality is more effective than RGB appearance. Thus, we have used different weights to fuse these modalities based on their prediction accuracy.

This paper presents the following main contributions:

- Multi-stream deep architecture to extract four modalities that are RGB, depth, motion and shape modalities. We demonstrate that the multi-stream network is able to digest complementary information and improve significantly the overall system performance.
- Depth images are used as input to the network to grant the detection during night time and under insufficient illumination.
- Different motion images are employed in order to extract efficiently the movement's information. It is based not only on optical flow displacement as done in many works [1, 42] but also based on amplitude and orientation of optical flow. More precisely, our method captures the velocity and the direction of the movement.
- The shape variations are defined based on binary motions images which are a combination of human silhouettes after background subtraction and person recognition. This stream gives on important idea for posture changes over time.
- We investigate and evaluate early and late fusion strategies to combine these modalities and generate the final decision for fall detection. We have defined the weight of each modality based on its performance contribution. Weighted score fusion is finally adopted based in our experiments.
- Transfer learning and data augmentation are used in order to surmount the insufficiency of amount of training data. In fact, deep convolutional networks have achieved excellent success for object and action recognition. However, for fall detection, the enhancement of deep convolutional networks is not so evident. First, the training dataset of fall detection is extremely small compared with the ImageNet dataset, and as a result over-fiting on the training dataset will be produced without data augmentation. In addition, experimental results have shown that the accuracy is improved after increasing the amount of data.

The rest of the paper is organized as follows. In Section 2, the related works are discussed. Section 3 details our proposed method. Section 4 presents the experimental results compared to state-of-the-art methods. Section 5 concludes our paper and future work is suggested.

2 Related works

This section gives an overview of the current effective methods on fall detection research based on RGB-D cameras and an introduction of multi stream deep convolutional neural networks.

2.1 Existing fall detection methods based on RGB-D cameras

Many vision-based methods are proposed to fall detection. They can be divided into two categories based on the kind of features used: hand-crafted or deep learned based methods:

Hand-crafted features based methods Extracting the best discriminating features is not a trivial task in fall detection. Many researchers are searching the more efficient features to get more accurate detection. Pathak and Bhosale [37] use only 3D information from the RGB-D camera located under the ceiling and directed vertically making the method invariant to ambient light. The fall is determined by the threshold of 0.4 m from the floor of the person's coordinates. But, this method includes a limited field of view at this location of the camera leading to incorrect detection if the distance is not respected and in the presence of confounding everyday activities such as lying. Albawendi et al. [3] proposed three features to detect a fall which are motion information based on tMNI, human shape variation based on angle of fitting ellipse and projection histogram. Rougier et al. [39] recognized falls based on the motion history image (MHI) and human shape changes. Sehairi et al. [42] exploited a set of features to detect falls such as vertical velocity of the head, area, height/width ratio and orientation. Gasparrini et al. [18] put Kinect on-ceiling and let it face downwards to detect fall by analyzing the raw depth data, which provided an automatic, privacy-preserving fall detection method for the indoor environment. Similar to using a ceiling-mounted 3D depth camera, Kepski et al. [20] analyzed the potential fall action by depth images. Fall detection method based on body parts movement and shape analysis is proposed by Khraief et al. [23]. Charfi et al. [9] extracted spatio-temporal features after background subtraction and classified using a SVM. Human head position and center of mass velocity are extracted by [32] from depth images. Human fall detection is proposed by [34] based on position and velocity of subject. Mastorakis and Makris [30] extracted features from the bounding rectangle for fall detection using RGB-D camera mounted on a tripod. Fall is identified based on shape variations by measuring the rate of decrease or increase in width, height and depth of the bounding rectangle. Alhimale et al. [4] used neural network to fall detection based on human silhouette after background subtraction.

Deep learned features based methods Deep learning searches for the relevant features in images, avoiding the feature engineering tasks and providing flexible automatic feature extraction. Wang et al. [49] extract features from still images after background subtraction using PCAnet and then applied a SVM classifier to distinguish fall from other daily life activities. Another method is proposed by Wang et al. [50] combining hand-crafted features such as Local Binary Pattern (LBP) and Histograms of Oriented Gradients (HOG) to features learned from a Caffe neural network. A convolutional neural networks based on optical flow is proposed by Adrian et al. [1] for elderly fall detection showing the importance of the motion information in the fall detection system. Lu et al. [29] applied also three-dimensional convolutional neural network (3D CNN) after extracting the human region using LSTM (Long

Short-Term Memory) with spatial visual attention schema. An adaptative deep learning schema is adopted by Doulamis et al. [15] to distinguish humans from the background and to deal with uncontrolled conditions of environment. Adhikari et al. [2] combined background subtracted RGB and depth image and fed it to convolutional neural networks to recognize different poses starting from the idea that fall event is a sudden change of pose. But, this method doesn't automatically detect falls and distinguish exactly fall like events but recognize only 5 activities. To detect falls, further feature engineering are needed such as speed of change of pose, aspect ratio and inclination angle. Recurrent neural network and long short-term memory network are used by [26] based on skeletons generated with 14 joints to detect a fall. Lin et al. [27] used the Kinect device to generate binary images designing the silhouette of the person. These images are analyzed by a network combining convolution layers followed by LSTM layers. Finally, Fan et al. [17] has adopted also convolutional neural network to detect falls using a generated dataset from YouTube videos.

In contrast of single-stream architecture, multi-stream deep architecture extracts multiple modalities. The related works and applications of the multi-stream CNN architecture are briefly presented in next subsection.

2.2 Multi-stream convolutional neural network

Conventional neural network architecture consists of extracting a single feature representation from the given training data. From the widely used feature representations for fall detection, we can cite optical flow displacement [1], RGB-D subtracted image [2] and single RGB frame [50]. A convolutional neural network is trained on this feature representation. The same features are extracted from the given test data during testing, and passed through the classifier to get the final score prediction. This architecture is referred to as single-stream architecture due to single feature representation extracted from the data, and a single classifier. By contrast, multi-stream architectures consist of extracting multiple feature representations from the given training data and testing data. Several convolutional neural networks are trained, one for each feature representation. Unlike single-stream architecture that gives only one feature vector, multi-stream system produce several feature vectors. These features can be merged and then passed forward one classifier (softmax) to get the final score. It is the early fusion strategy. In contract, in the late fusion strategy, each feature is passed throw its classifier producing its own score prediction. These score should be fused to get the final score. Thus, the multi-stream architecture is based on multiple features and multiple classifiers. Each feature is affected by noise differently. This is the main motivation behind using multiple classifiers in multi-stream architecture.

The architecture of multi-stream CNN has been recently proposed in various computer vision tasks. Wang et al. [51] used the RGB images and gradient images as input of two-stream CNN for face detection. Simonyan et al. [43] combined RGB and optical flow displacements as motion features extracted by the CNN for activity recognition. Zhang et al. [58] used three-stream CNN architecture for dense crowd counting. Liu et al. [28] combined deep and shallow networks for person re-identification task. Rahman et al. [38] proposed a three-stream multi-modal CNNs based deep network architecture for RGB-D object recognition. The three streams are based on surface normal, color, and RGB channel. Zhou et al. [60] proposed a fall detection system based on the combination of optical camera and radar. Three convolutional neural network (CNN) are used for action recognition. Two CNNs is trained

to classify the TF images extracted from radar images, and one CNN is used to predict the shape variations after bounding box extraction of the human silhouette. Then, the result is given by fusing the decision of these three CNNs. Wang et al. [48] proved that multi-stream convolutional networks based on temporal information where optical flow fields are fed into CNNs outperformed RNNs in action recognition task [48]. Espinosa et al. [16] present a fall detection system based on a 2D CNN inference method and multiple cameras. Their results showed that multi-vision-based approach detects human falls and achieves an accuracy of 95.64% compared to state-of-the-art methods with a simple CNN network architecture.

Based on the aforementioned researches, we can conclude that the motion variations, shape changes, RGB and depth information are used for fall detection but these four modalities were never combined as deep learned features. Based on the rich multi-modal data given by RGB-D cameras and starting from the basic definition of fall event defined as a sudden change of shape and a rapid downward movement of position of the body from an upright, sitting or lying position to a lower inclining position [36], we propose multi-stream convolutional neural networks that fuse all these complementary features. Our proposed method will be explained in the next section.

3 Proposed method

Motivated by the success of deep learning methods and the multi-modal data provided by RGB-D cameras, we propose multi-stream convolutional neural networks that combine complementary information of motion, shape, RGB and depth information for elderly fall detection.

- The first stream is for analyzing the shape variations and models the pose deformations. The main idea behind our method is that the fall event is characterized by a defined sequence of poses that ends in lying. Adhikari et al. [2] method fed the CNN by only one single frame resulted from background subtracted RGB and depth image and then a post processing is done in order to distinguish the fall event from other daily activities by analyzing different poses predicted. In contract of Adhikari et al. [2] method, our method detect automatically fall event without any post processing by alimenting the CNN not only by one frame but by stacked frames representing a sequence of pose changes and modeling the various human silhouette deformations.
- The second stream is for extracting appearance information from depth and RGB single image in order to profit from the advantages of both RGB and Depth. Thus, our method use RGB image especially with the distance limitation of depth image. Similarly, our method is based on depth information if RGB image is influenced by illumination changes.
- The third stream is based on multi frame vertical and horizontal displacements of optical flow to extract the motion information. The person's displacement is a good indicator if a sudden movement is occurred. Thus, this stream is proposed by many tasks for fall detection [1] and action recognition by [43]. But, knowing only that a sudden movement is happened doesn't mean that this event corresponds to fall action but it can be "jumping suddenly".
- The fourth stream is added to enhance the motion information provided by the third stream by measuring the motion velocity and direction of human movement. This stream

is based on multi frame of amplitude and orientation optical flow. These features are usually used as hand-crafted feature for fall detection but not as learned deep features despite its significant contribution in action recognition. Colque et al. [12] proposed the Histograms of Optical Flow Orientation and Magnitude (HOFM) to capture the orientation and the magnitude of flow vectors improving significantly the accuracy of event recognition. Inspired by the effectiveness of these features, we propose this temporal stream based on amplitude and orientation optical flow to deal with the fast velocity and the downward direction of the movement.

Given the prediction scores of these multiple network streams, we are able to extract complementary characteristics of fall event. It is important to efficiently fuse the multi-stream to generate the final binary prediction result which is "Fall" and "Not Fall". We investigate different fusion strategies to improve the accuracy of fall detection. We have adopted weighted score fusion as a late fusion strategy. Transfer learning and data augmentation are used to deal with small amount of training data and to train each conventional neural network.

As shown in Fig. 1, our proposed method involves four main modules that are feature extraction, multi-stream convolutional neural network architecture, fusion strategy and finally transfer learning with data augmentation techniques. Each step will be explained in detail in next subsections.

3.1 Feature extraction

In this section, we describe the different information used as input for the proposed multi-steam CNN architecture. In particular, we use RGB images, depth images, optical flow displacement, history of binary motion images, amplitude and orientation flow. This set of features will cover motion, shape and appearance information of elderly person activities.

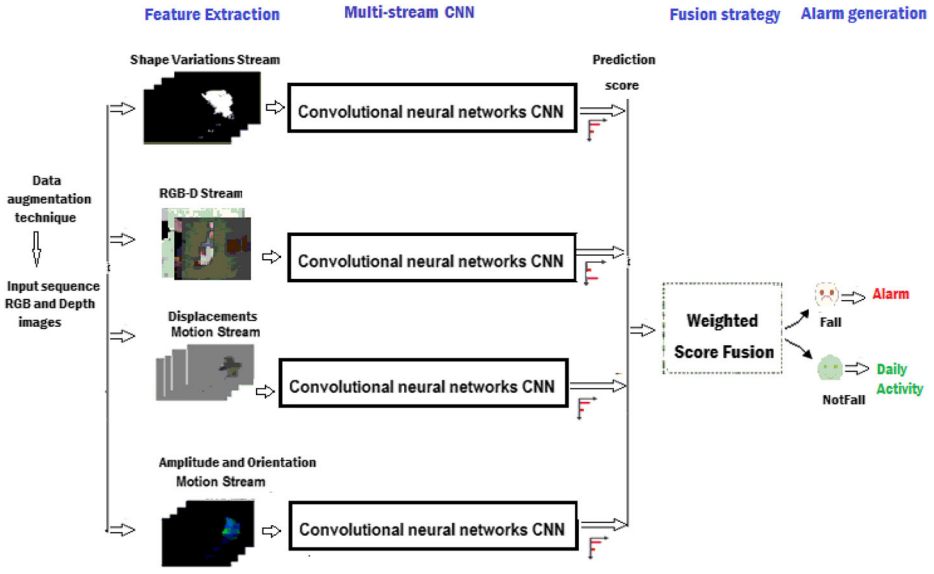


Fig. 1 Proposed method flow

A. History of binary motion image HBMI

Extraction of the human body is the first step in our method in order to classify pixels of each frame into background or foreground and recognizing human from other moving objects. We used the method of khraief et al. [22] for moving person detection and tracking. First, improved Visual Background Extractor (ViBe) [6] is used to extract foreground regions. Then dollar et al. [14] method is applied to recognizing person from others moving objects. Second, level-Set based region method is applied to track persons in each frame. Finally, to deal with partial occlusion, Level-Set based edge is used to obtain more accurate objects contours.

After human silhouette extraction, we combine many consecutive binary silhouettes into one image in order to model human action as illustrated by Fig. 2. We used the following to generate the history of binary motion image HBMI.

$$\mathrm{HBMI}(x, y) = \sum_{t=0}^{n} f(t) M_{xy}(t) \tag{1}$$

where

- M(t) is the binary image containing the human silhouette only without background.
- f(t) is the weight function that gives more higher weight to recent frames

B. The optical flow displacement

Optical Flow measures the motion of objects between two consecutive frames. Several algorithms exist in the literature for optical flow computation. They are based on the brightness constancy assumption. If the pixels I (x, y, t) in an image moves by distance (dx, dy) in another image taken after dt time. Since the intensity of the same pixels will not change considerably within a small-time interval. Thus, the following brightness formula can be given by eq2.

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{2}$$

Thus, the optical flow constraint equation (OFCE) is obtained by using Taylor expansion in (Eq. 2) and dropping its nonlinear terms. Therefore, the OFCE can be expressed in the form as flowing,

$$I_x V_x + I_y V_y + I_t = 0 \tag{3}$$

where (V_x, V_y) represent the optical flow vectors (δx, δy) and (I_x, I_y, I_t) represent the derivatives of image intensities at coordinate (x, y, t).

Optical flow algorithm can be classified into sparse or dense methods. Sparse algorithm computes the flow only for certain specified pixels, while dense algorithm computes the flow

Fig. 2 Examples of MCF Dataset (a) and their history of binary morion image (b)

for all the pixels. Despite, sparse method is often faster but dense method gives more flow vectors and can lead to better motion estimation.

We used the dense optical flow method of Gunnar Farneback method [19]. This method uses quadratic polynomials that gives us the local signal model expressed in a local coordinate system such that,

$$\mathcal{F}(x) \sim x^T A x + b^T A x \tag{4}$$

Where A is a symmetric matrix, b a vector and c a scalar.

Optical flow is calculated from images generating vertical and horizontal components (V_x, V_y) for each image. We will use only 10 optical flow images from each video sequence this resulting in 20 channels as an input to CNN model for feature extraction and training. These deep features were used as input of a classifier that gives the event output as "fall" or "no fall."

C. Optical flow amplitude and orientation

We introduce a new motion stream to further improve the motion features extraction by CNN. It is based on magnitude and orientation extracted from the optical flow.

Our method extracts not only the displacement, but also magnitude and orientation of the fall event. In fact, this information is frequently used to characterize motion information in various hand-crafted features such as Histograms of Optical Flow Orientation and Magnitude (HOFM) [12], Histogram of Oriented Flow (HOF) [39], etc. Various fall detection methods are based on the magnitude and the orientation as hand-crafted features rather than optical flow displacement. We propose to use the magnitude and orientation of optical flow to extract deep motion features.

To incorporate such information on the temporal stream, we compute the optical flow and then we calculate the magnitude OF_{mag} and orientation OF_{phase} information as flows,

$$OF_{mag} = \sqrt{V_x^2 + V_y^2} \tag{5}$$

$$OF_{phase} = \tan^{-1}\left(\frac{V_y}{V_x}\right) \tag{6}$$

A HSV (Hue-Saturation-Value) model based color-coding converts the optical flow vector to an RGB image. At each pixel, the flow direction OF_{phase} is coded as hue while the flow magnitude OF_{mag} is coded as saturation. The input is composed by 10 stacked images. The magnitude and orientation flow are illustrated by Fig. 3.

3.2 Network architecture

Several deep convolutional neural network architectures have been proposed in computer vision in the recent years. We adopted the VGG-16 net [42] for our task, motivated by its high accuracy obtained in other related domains. The VGG-16 network architecture was initially

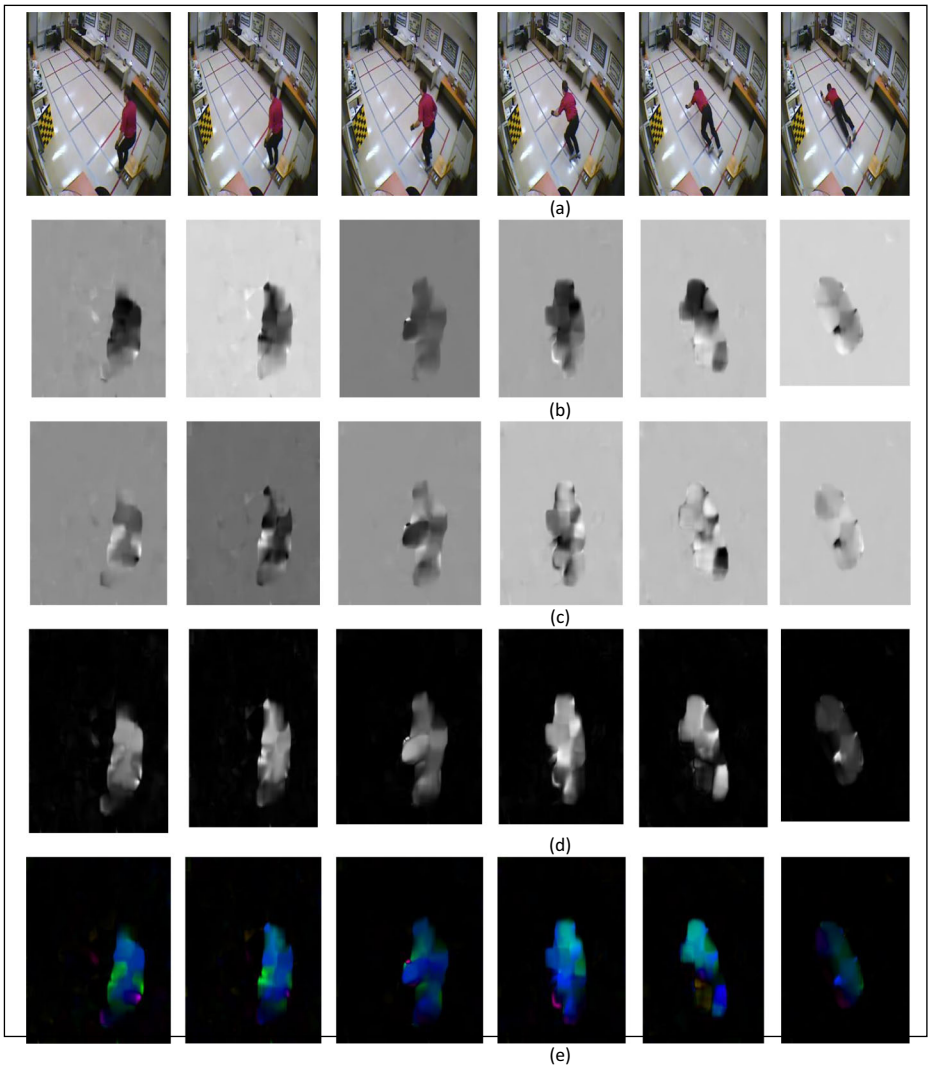


Fig. 3 Frames Multiple Cameras Fall Dataset (a) and their optical flow horizontal displacement (b), vertical displacement (c), amplitude(d) fusion of amplitude and orientation flow (e)

proposed by Simonyan and Zisserman [42]. Its architecture contains 13 layers to form five blocks of convolutional layers. Each convolutional layer use filters with size 3×3 followed by max Pooling with filter size 2×2, a stride of 2 and rectified linear unit activation. Two fully-connected layers with 4096 and ReLU activated units are then used before the softmax layer. The final prediction gives 1000 classes.

Given that we are addressing the fall detection, there are only two possible output values of the CNN architecture, namely "Fall" and "Not Fall". Therefore, the last fully connected layer produces only two classes as illustrated by the Fig. 4. We replaced the input layer of VGG-16 net in order to generate multiple streams that treat effectively dynamic motion, depth and RGB shape variations.

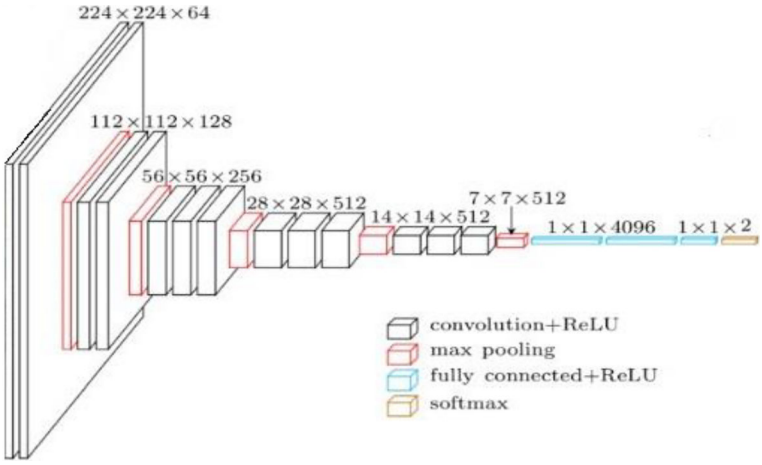


Fig. 4 Our VGG-16 architucture producing two classes

3.3 Fusion strategy

There are three principal strategies to fuse information in deep learning models: early fusion, late fusion and hybrid fusion. The last one is essentially a combination of early and late fusion variants. We have evaluated early and late fusion.

A. Early fusion

In early fusion strategy also referred to feature level fusion, the data is fused before it is feed into the model as illustrated in Fig. 5. All features from the last convolutional layer

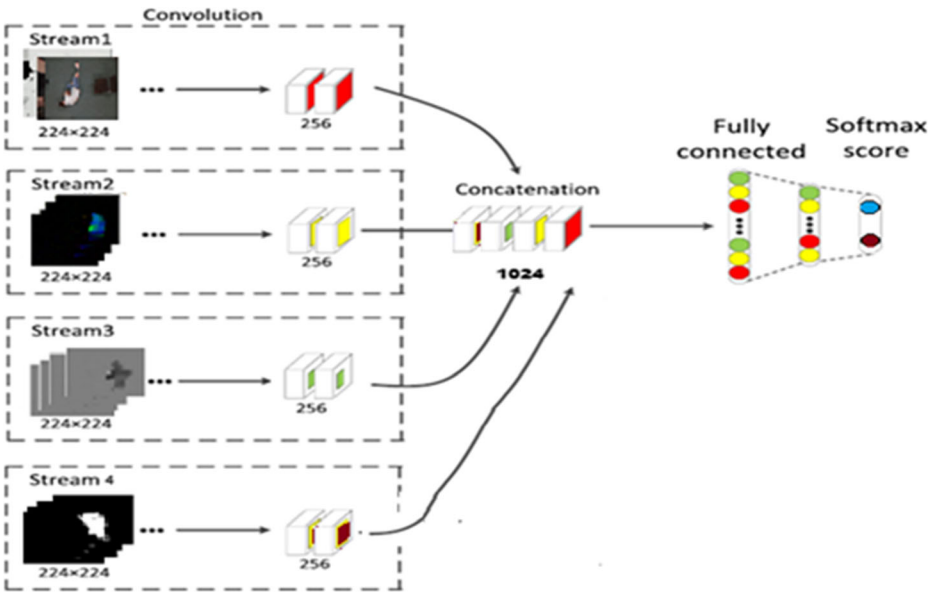


Fig. 5 Early fusion strategy

are integrated into a single stacked multimodal vector and trained to produce the final prediction.

B. Late fusion

Late fusion is the combination of the output probabilities of different streams. This fusion method is also named decision level fusion or score fusion as shown in Fig. 6. We have used the late schema for fusing the four streams. It consists in weighting the score of each modality and sums them up. In our network, different modalities should have different weights in the final information fusion in order to use the advantages of each modal and control its contribution. We have defined the weight of each modal and adjust it in the validation process. The late fusion is more flexible than early one because each model has its own representation space and its own hyper-parameters.

3.4 Transfer learning and data augmentation

Deep convolutional neural networks achieved state-of-the-art results in many classification tasks. However, they still have challenges to overcome. They require a large number of labeled data in training to avoid over-fitting as well as ensure generalization abilities. Over-fitting occurs when a network learns a function that perfectly model the training data and generalization ability is verified by measuring the performance of the generated model when evaluated on training data as on previously seen data versus data testing as data never seen

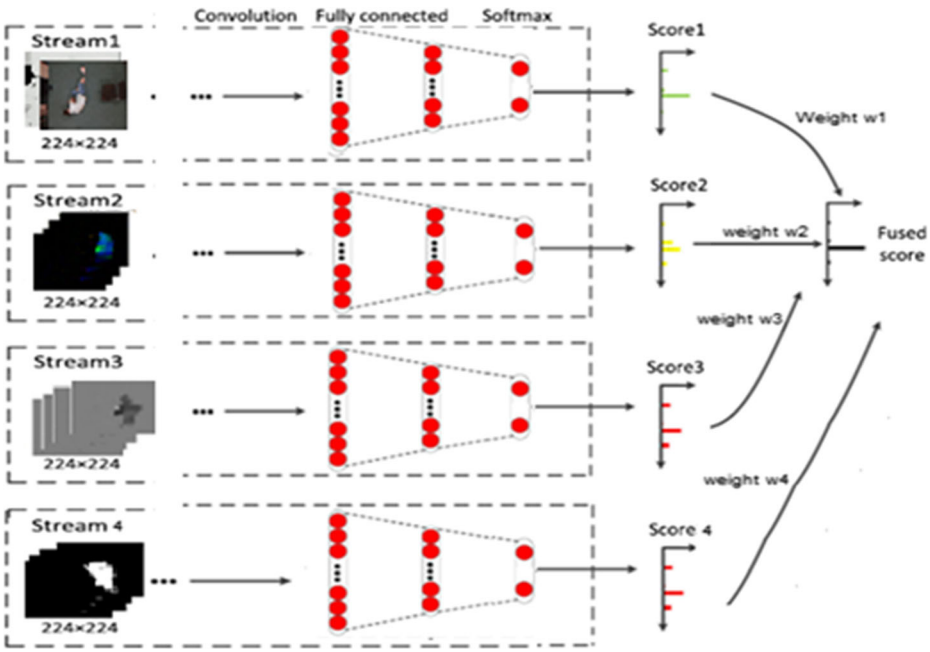


Fig. 6 Late fusion strategy

before. Unfortunately, our training data suffers from the small amount of images. Thus, we have used transfer learning and data augmentation.

A. Transfer learning

Transfer learning consists in reassign the parameters of a neural network trained with one dataset and task to another problem with a different dataset and task. When the target dataset is significantly smaller than the base dataset, transfer learning can be a powerful tool to enable training a large target network without over-fitting. We have used VGG16 as the base model, pre-trained for object detection task on the ImageNet dataset. Learning millions of parameters during this pre-training will be done in order to get generic visual features. These features will be fed to our convolutional neural network as illustrated by Fig. 7. We retrained the network on the UCF101 dataset [44] to learn more motion features that could be later used to classify falls in the motion stream. Finally, we reuse the network weights and fine-tuning the classification layers in order to generate two classes 'fall' or 'not fall'.

B. Data augmentation

Data Augmentation is a very powerful way of overcome the small amount of data by encompassing a suite of techniques that improve the size and quality of training datasets without really collecting new data. Furthermore, to reduce over-fitting and ensure the generalization of the network, each original image in the training has undergone several variations generating more data. The position and orientation of elderly person can be changed according to camera's position and point of view. Also, image lighting is affected by camera and external environment lighting.

Five augmentation techniques are used and some image's changes are illustrated by Fig. 8:

– Rotation: To make the network more robust to rotation, we generated five additional images for each image by its 45 degree rotation increments between 0 and 360 degrees

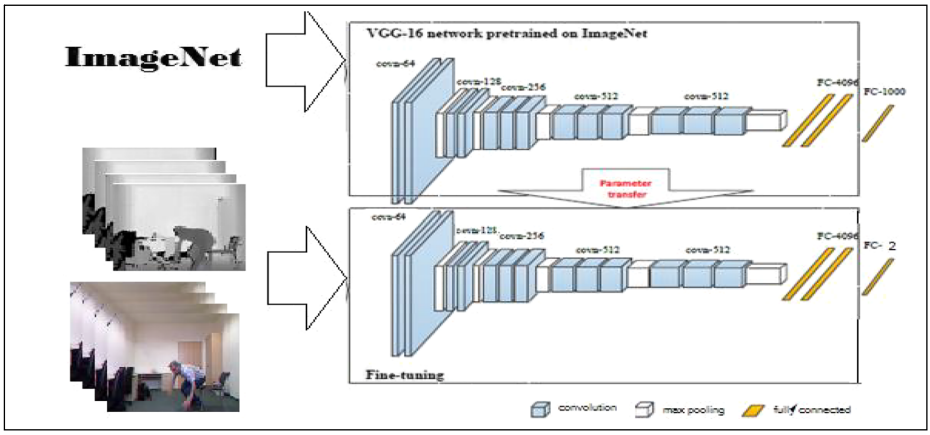
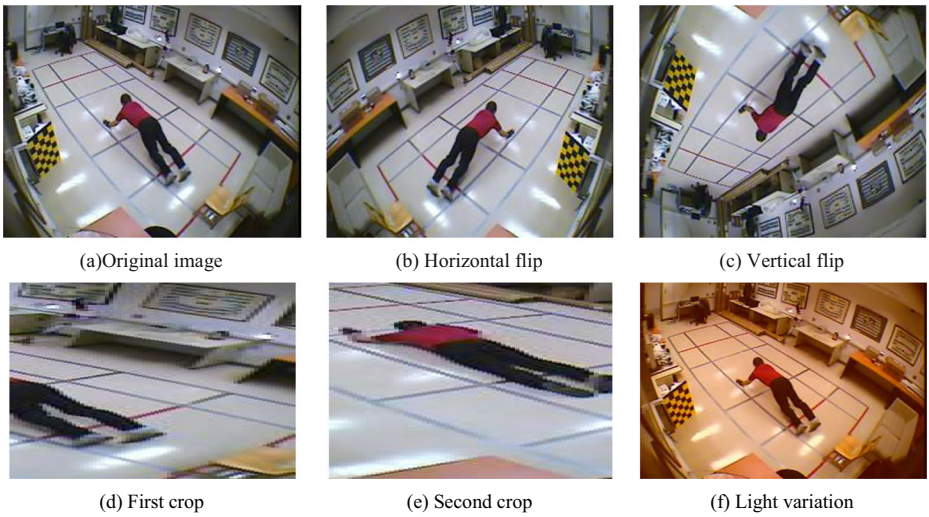


Fig. 7 The deep Convontional neural network is first pre-trained on the ImageNet dataset and then fine-tuned on the RGB and depth images as input producing two classes as output



(a)Original image (b) Horizontal flip (c) Vertical flip

(d) First crop (e) Second crop (f) Light variation

Fig. 8 MCF Dataset . data augmentation (a)Original image (b) Horizontal flip (c) Vertical flip (d) First crop (e) Second crop (f) Light variation

- Cropping: We randomly sample a section from the original image, then, this region is resized to the original image size.
- Flipping: Two more images were generated by flipping each original image horizontally and vertically.
- Lighting variations: To deal with illumination variations, three more images for each original image are produced by lightness changing.
- Scaling: Objects can appear in different sizes depending on the object and camera position, but they have the same characteristics.

  This increased data makes the proposed method insensitive to scaling, color and transformation and effectively improves the robustness of the entire network.

4 Experimental results

In this section, we present the datasets, quantitative and qualitative results. The architecture has been implemented using the TensorFlow, Keras, the Python programming language and MATLAB R2018b. All procedures, training and testing were performed on Dell system with Intel® Core™ i7-7500u - 2.9GHz, 8 GB of RAM, and Windows 10 Home Premium 64-bit operating system.

For training VGG-16, we resized each image to the fixed size of 224 × 224 and we fixed the dropout regularization for the two fully-connected layers to 0.5. The learning rate was initially set to 10^{-3} and then decreased by a factor of 10. We fine-tuned on two folders using the "Fall" vs. "NotFall" image label. Adam Optimizer algorithm was used. The loss function is the cross-entropy.

The dataset was further divided into two subsets: Dtrain and DTest with a ratio of 0.8 and 0.2, respectively. The first dataset is used for training and the second is used for testing. 25% of the train set (Dtrain) is used as validation set (DValid) during the training

Fig. 9 Keyframes MCF Dataset . Top row: human falls. Bottom row: daily life activities

process in order to adjust the weights of different streams and test the generalization ability of the model.

4.1 Datasets

Experiments have been conducted based on three publicly available datasets: the Multiple Cameras Fall (MCF) dataset [39], the UR Fall Detection (URFD) dataset and Fall Detection Dataset FDD:

- The (MCF) dataset [39] is recorded from eight cameras mounted on the walls and contained 24 scenarios of simulated falls and normal daily activities such as sitting on a chair, walking, crouching, etc. Figure 9 shows some example frames from the MCF dataset.
- The UR Fall Detection (URFD)[1] is captured by two Kinect sensors and contains frontal and overhead video sequences. The frontal sequence includes 314 frames, in which 74 frames have falls and 240 frames have no falls. The overhead sequence encloses a total of 302 frames, in which 75 frames define falls, while 227 have no falls. Two types of falls were defined that are from standing position and from sitting on the chair. Figure 10 shows some example frames from the URFD dataset.
- Fall Detection Dataset FDD [2] contains 190 videos in different simulated environments including office, coffee room, home, and lecture room

Fig. 10 Keyframes from UR Fall Detection (URFD). Top row: RGB human activities. Bottom row: Depth daily life activities

Fig. 11 Examples of our proposed method results based on MCF Dataset . Top row: human falls with different directions. Bottom row: daily life activities

4.2 The qualitative evaluation results

Figure 11 presents some results of our method based on MCF dataset. As shown, our method does not depend on camera position neither on fall direction. In fact, in whatever direction the person falls, either forward or backward, to the left or to the right, the fall will be detected by our method.

From the previous related works, the fall is only detected when the person is standing and then falls on the background. Our method does no longer depend on fall speed or location only but on the person shape variations and so, it is able to detect all these kinds of falls as illustrated by Fig. 12.

4.3 Quantitative evaluation results

This section presents the results obtained for human fall detection. We used three criteria widely used to evaluate fall detection systems. The sensitivity (Se) determines the capacity of

Fig. 12 Our proposed method results on complicated situations

the method to classify fall activities correctly, specificity (Sp) is its capability to distinguish daily life activities properly and the accuracy A measures the ability of the method to differentiate between fall and daily life activities. They are formulated as bellow:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{8}$$

$$\mathrm{A} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{9}$$

Where

- TP (True Positives) is the number of falls correctly detected as a fall by the system,
- TN (True Negatives) is the number of the fall not occurred and correctly detected as a non-fall,
- FP (False Positives) is the number of incorrectly events detected as fall
- FN (False Negatives) is the number of fall events missed by the method and identified as non-fall

Figures 13 shows the confusion matrix obtained for the MCF dataset. As can be seen, most of the samples are correctly classified. An overall accuracy is 99.8% was achieved, although there appears a small confusion between "Fall" and "notFall" activities (0.1% and 0.2% respectively) caused by Falling-like activities.

We have conducted several comparisons with the methods of the state-of the-art to ensure the robustness of our method based on the same publically evaluation databases despite many researchers have created their own dataset to evaluate their works.

Comparisons to hand-crafted features based methods we have compared our method to state-of the-art approaches that are based on hand-crafted features such as Yun al. approach [56], MHI based approach [39] and Chua's approach based on three-point representation [10]. As shown in the Table 1, the hand-crafted low-level features can usually work well in

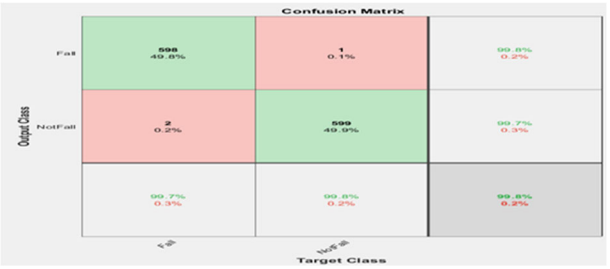


Fig. 13 Confusion matrix

Table 1 Comparison of Our Deep Features Extraction method with hand-crafted low-level features Methods in terms of sensitivity and specificity

	Proposed method	Yun et al. [56]	MHI [39]	Chua et al. [10]
Sensitivity (%)	99.70	98.55	85.70	90.50
Specificity (%)	99.80	95.84	80.00	93.30

Table 2 Comparison of proposed method with state-of-the-art approaches in terms of sensitivity and specificity

	Proposed method	Adrian et al. [1]	Wang et al. [49]	Wang et al. [50]
Sensitivity (%)	99.70	99.00	89.20	93.70
Specificity (%)	99.80	96.00	90.30	92.00

constraint environment. Yet, they are not universal for all conditions. Deep learning is treated as a better method to extract high-level features.

Comparisons to deep learning based methods *We* have compared our method to approaches based on deep learning such as Wang's approach based on PCANet [49], Wang's approach based on combination of features [50] and Adrian et al. approach [1] based on optical flow only. The results are illustrated in Table 2. Our approach outperforms their results by using multiple motion information not only optical flow as [1] and the use of the shape stream after background subtraction instead of motion stream only.

Comparisons with state-of-the-art approaches using URFD dataset and FDD dataset The evaluation of our method to other vision-based systems that have used the URFD and FDD is illustrated by Table 3 and Table 4 respectively. The results demonstrate the outperformance of the proposed method over fall detection by the state-of-the-art methods.

Comparisons of multi-stream fusion strategies We investigate two fusion strategies including early fusion and late fusion to combine the different streams. As Table 5 shown, our method achieves an accuracy rate of 99.72% for late fusion strategy and only 98.60% for early fusion strategy. Therefore, based on these experiments, late fusion schema is more suitable for combining multi-stream CNN and achieves state-of-the-art.

Table 3 Comparison of proposed method with state-of-the-art approaches for URFD Dataset

	Proposed method	Adrian et al. [1]	Kwolek et al. [25]	Nizam et al. [35]
Sensitivity (%)	100.00	100.00	100.00	96.67
Specificity (%)	95.00	92.00	80.00	82.50

Table 4 Comparison of proposed method with state-of-the-art approaches for FDD Dataset

	Proposed method	Adrian et al. [1]	Yu et al. [55]	Charfi et al. [9]	Zerrouki and Houacine [57]
Accuracy (%)	99.72	97.00	96.09	99.61	97.02

Table 5 Comparison of proposed method with Early and late Fusion strategy for FDD Dataset

	Early fusion	Late Fusion
Accuracy (%)	98.60	99.72

Table 6 Comparison of proposed method with the same or different weights

	Same weight	Different weights
Accuracy (%)	98.31	99.72

Impact of weighted score fusion We have assumed that different modalities didn't contribute by the same manner in the detection of fall event and they should have different weights in order to exploit the advantage of each modality perfectly. Thus, we have attributed these weights 0.2, 0.2, 0.3, 0.3 respectively to RGB-D stream, shape stream, amplitude and direction motion stream and finally motion displacements stream in order to give more importance to motion information as [43]. As Table 6 shown, the same weight method means that all the modal features are treated equally and our method which gives different weights achieves the best result.

Impact of data augmentation We have compared the accuracy of our method with and without data augmentation as shown in Table 7. We note that with data augmentation, our method has achieved higher accuracy. This result proves the significant contribution of the data augmentation technique to the overall system's performance and thus reduces the potential over-fitting.

5 Conclusion

In this paper, we have proposed a multi-stream convolutional neural network for elderly fall detection using RGB-D cameras. Our method is based on deep learning architecture that analyzes both the appearance, motion and the shape variations. Human motion detection before fall recognition exclude background thus let our method to better perform on various situations, in indoor or outdoor environment. Multi-stream deep architecture are fused to extract complementary features from RGB images, depth images, optical flow displacement, history of binary motion images, amplitude and orientation optical flow. Transfer learning and data augmentation are used to overcome the problem posed by the low number of images in fall datasets and to learn generic features.

The experimental results obtained are promising compared the state-of-the-art methods but can be improved by further including the audio information as a future work. Furthermore, we

Table 7 Comparison of proposed method with and without data Augmentation

	Without data augmentation	With data augmentation
Accuracy (%)	96	99.72

will investigate the possibility of fusion of our vision-based method to wearable sensor- based method such as smart watch or smart phone to improve the results.

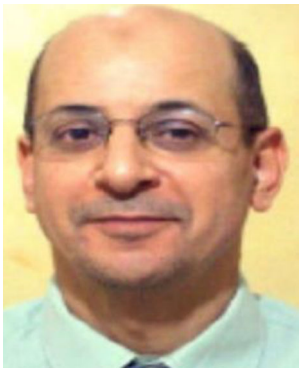
References

1. Adam MA, Azkune G, Arganda-Carreras I (2017) Vision-based fall detection with convolutional neural networks. Wirel Commun Mob Comput 1:1–16
2. Adhikari K, Bouchachia H, Nait-Charif H (2017) Activity recognition for indoor fall detection using convolutional neural network. In Proceedings of the 15th IAPR International Conference on Machine Vision Applications, Nagoya, Japan, pp. 81–84
3. Albawendi S, Lotfi A, Powell H, & Appiah K (2018). Video based fall detection using features of motion, shape and histogram. Proceedings of the 11th PErvasive technologies related to assistive environments conference on - PETRA '18. doi:10.1145/3197768.3201539
4. Alhimale L, Zedan H, Al-Bayatti A (2014) The implementation of an intelligent and video-based fall detection system using a neural network. Appl Soft Comput 18:59–69. https://doi.org/10.1016/j.asoc.2014.01.024
5. Allali G, Ayers EI, Holtzer R, Verghese J (2017) The role of postural instability/gait difficulty and fear of falling in predicting falls in non-demented older adults. Arch Gerontol Geriatr 69:15–20
6. Barnich O, Van Droogenbroeck M (2011) ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans Image Process 20(6):1709–1724
7. Burns ER, Stevens JA, Lee R (2016) (2016). The direct costs of fatal and non-fatal falls among older adults — United States. J Saf Res 58:99–103
8. Casilari E, Oviedo-Jiménez MA. (2015) Automatic fall detection system based on the combined use of a smartphone and a smartwatch. PLoS one 2015; 10(11): e0140929.
9. Charfi I, Miteran J, Dubois J, Atri M, Tourki R (2013) Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. Journal of Electronic Imaging 22(4):041106. https://doi.org/10.1117/1.jei.22.4.041106
10. Chua JL, Chang YC, Lim WK (2013) "A simple vision-based fall detection technique for indoor video surveillance". Signal, Image and Video Processing: p. 1–11
11. Cippitelli E, Fioranelli F, Gambi E, Spinsante S (2017) Radar and RGB-depth sensors for fall detection: a review. IEEE Sensors J 17(12):3585–3604
12. Colque RVHM, Junior CAC, Schwartz WR (2015) Histograms of optical flow orientation and magnitude to detect anomalous events in videos. 2015 28th SIBGRAPI
13. Delahoz Y, Labrador M (2014) Survey on fall detection and fall prevention using wearable and external sensors. Sensors 14:19806–19842
14. Dollar P, Belongie S, Perona P (2010) The fastest pedestrian detector in the west, in: British Machine Vision Conference, pp. 68.1–68.11
15. Doulamis A, Doulamis N (2018) Adaptive deep learning for a vision-based fall detection. Proceedings of the 11th *PETRA '18*
16. Espinosa R, Ponce H, Gutiérrez S, Martínez-Villaseñor L, Brieva, J., & Moya-Albor, E. (2019). A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset Computers in Biology and Medicine, 103520. doi:https://doi.org/10.1016/j.compbiomed.2019.103520
17. Fan Y, Levine MD, Wen G, Qiu S (2017) A deep neural network for real-time detection of falling humans in naturally occurring scenes. Neurocomputing 260:43–58
18. Gasparrini S, Cippitelli E, Spinsante S, Gambi E (2014) A depth-based fall detection system using a Kinect sensor. Sensors 14:2756–2775
19. Gunner Farneback (2003) "Two-Frame Motion Estimation Based on Polynomial Expansion", Image Analysis, pp363–370
20. Kepski, M.; Kwolek, B (2014) Fall detection using ceiling-mounted 3D depth camera. In Proceedings of the 2014 International conference on computer vision theory and applications (VISAPP), Lisbon, Portugal, 5–8 January 2014; Volume 2, pp. 640–647
21. Kharrat O, Mersni E, Guebsi O, Ben Salah FZ, Dziri C (2017) Qualité de vie et personnes âgées en Tunisie, NPG Neurologie - Psychiatrie - Gériatrie, Volume 17, Issue 97, Pages 5–11

22. Khraief C, Benzarti F, Amiri H (2017) Multi person detection and tracking based on hierarchical Level-Set method. In: The 10th International Conference on Machine Vision (ICMV) pp.91–98
23. Khraief C, Benzarti F, Amiri H (2018) Vision-based fall detection for elderly people using body parts movement and shape analysis. In: The 10th International Conference on Machine Vision (ICMV)
24. Kistler BM, Khubchandani J, Jakubowicz G, Wilund K, Sosnoff J. (2018). Falls and Fall-Related Injuries Among US Adults Aged 65 or Older With Chronic Kidney Disease. Prev Chronic Dis 2018;15:170518.
25. Kwolek B, Kepski M (2014) Human fall detection on embedded platform using depth maps and wireless accelerometer. Comput Methods Prog Biomed 117:489–501
26. Lie WN, Le AT, Lin GH (2018) Human fall-down event detection based on 2D skeletons and deep learning approach. In Proceedings of the International Workshop on Advanced Image Technology, Chiang Mai, Thailand, pp. 1–4
27. Lin HY, Hsueh YL, Lie WN (2016) Abnormal event detection using Microsoft kinect in a smart home. In Proceedings of the 2016 International Computer Symposium, Chiayi, Taiwan, pp. 285–289
28. Liu J, Zha ZJ, Tian Q, Liu D, Yao T, Ling Q, Mei T (2016) Multi-scale triplet cnn for person re-identification. In: Proceedings of the 2016 ACM on Multimedia Conference, MM '16, pp. 192–196. ACM, New York, NY, USA. https://doi.org/10.1145/2964284.2967209
29. Lu N, Wu Y, Feng L, Song J (2019) Deep learning for fall detection: three-dimensional CNN combined with LSTM on video kinematic data. IEEE journal of biomedical and health Informatics, 23(1), 314-323. doi: https://doi.org/10.1109/jbhi.2018.2808281
30. Mastorakis G, Makris D (2012) Fall detection system using Kinect's infrared sensor. J Real-Time Image Process 9:635–646. https://doi.org/10.1007/s11554-012-0246
31. Mauldin TR, Canby ME, Metsis V, Ngu AHH, Rivera CC (2018) SmartFall: A Smartwatch-Based Fall Detection System Using Deep Learning. *Sensors (Basel)*. 18(10):3363. Published 2018 Oct 9. doi:10.3390/s18103363
32. Merrouche F, Baha N (2017) " Fall Detection using Head Tracking and Centroid Movement Based on a Depth Camera", International Conference ICCES pp. 29–3
33. Mubashir M, Shao L, Seed L (2013) A survey on fall detection: principles and approaches. Neurocomputing 100:144–152
34. Nizam Y, Mohd MNH, Jamil MMA (2017) Human fall detection from depth images using position and velocity of subject. Procedia Computer Science 105, 131-137
35. Nizam Y, Mohd M, Jamil M (2018) Development of a user-adaptable human fall detection based on fall risk levels using depth sensor. Sensors 18(7):2260
36. Noury N, et al (2007) "Fall detection principles and methods." 2007 29th Annual International Conferenceof the IEEE Engineering in Medicine and Biology Society. IEEE
37. Pathak D, Bhosale VK (2017) Fall detection for elderly people in homes using Kinect sensor. Int. J. Innov. Res. Comput. Commun. Eng. 5, 1468–1474. Doi:https://doi.org/10.15680/IJIRCCE.2017
38. Rahman MM, Tan Y, Xue J, Lu K (2017) RGB-D object recognition with multimodal deep convolutional neural networks. In Proceedings of the IEEE International Conference on Multimedia and Expo., Hong Kong, China, 10–14, pp. 991–996
39. Rougier C, Meunier J, St-Arnaud A, and Rousseau J (2007) "Fall detection from human shape and motion history using video surveillance," in Proceedings of AINAW'07, pp. 875–880
40. Roy N, Dubé R, Després C, Freitas A, Légaré F (2018) Choosing between staying at home or moving: a systematic review of factors influencing housing decisions among frail older adults. PLoS One 13(1): e0189266. https://doi.org/10.1371/journal.pone.0189266
41. Santos GL, Endo PT, Monteiro KHC, Rocha EDS, Silva I, Lynn T (2019) Accelerometer-Based Human Fall Detection Using Convolutional Neural Networks. *Sensors (Basel)*. 19(7):1644. Published 2019 Apr 6. doi:10.3390/s19071644.
42. Sehairi K, Chouireb F, Meunier J (2018) Elderly fall detection system based on multiple shape features and motion analysis. 2018 international conference on intelligent systems and computer vision
43. Simonyan and Zisserman A (2014) Very deep convolutional networks for large-scale image recognition, arXiv 1409.1556 , 2014
44. Soomro K, Zamir AR and Shah M (2012) UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild.CRCV-TR-12-01
45. Stevens JA, Rudd RA (2015) Circumstances and contributing causes of fall deaths among persons aged 65 and older, J. Am. Geriatr. Soc., Vol. 62, No. 3, pp. 470–475, 2015.
46. Uddin M, Khaksar W, Torresen J (2018) Ambient sensors for elderly care and independent living: a survey. Sensors 18(7):2027. https://doi.org/10.3390/s18072027
47. Vielzeuf V, Pateux S, and Jurie F (2017) Temporal mul-timodal fusion for video emotion classification in the wild. InProceedingsof the 19th ACM International Conference on Multimodal Interaction,pages 569–576. ACM

48. Wang L, Xiong Y, Wang Z, Qiao Y, Lin D, Tang X, and Van Gool L (2016) Temporal segment networks: Towards goodpractices for deep action recognition. InEuropean Confer-ence on Computer Vision, pages 20–36. Springer
49. Wang S, Chen L, Zhou Z, Sun X, Dong J (2016) Human fall detection in surveillance video based on PCANet. Multimed Tools Appl 75(19):11603–11613
50. Wang K, Cao G, Meng D, Chen W, and Cao W (2016) "Automatic fall detection of human in video using combination of features," IEEE International Conference BIBM, pp. 1228–1233,China
51. Wang W, Lu X, Song J, Chen C (2016) A two-column convolutional neural network for facial point detection. In: 2016 International Conference on Progress in Informatics and Computing (PIC), pp. 169–173. https://doi.org/10.1109/PIC.2016.7949488
52. Wortmann M (2012) Dementia: a global health priority - highlights from an ADI and World Health. Organization report. Alzheimers Res Ther 4(5):40
53. Xu T, Zhou Y, Zhu J (2018) New advances and challenges of fall detection systems: a survey. Appl Sci 8: 418. https://doi.org/10.3390/app8030418
54. Yang X, Molchanov P, and Kautz J (2016) Multilayer and mul-timodal fusion of deep neural networks for video classification. InPro-ceedings of the 2016 ACM on multimedia conference, pages 978–987.ACM
55. Yu M, Rhuma A, Naqvi SM, Wang L, Chambers J (2012) A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. IEEE Trans Inf Technol Biomed 16(6): 1274–1286
56. Yun Y, Gu IY-H (2016) Human fall detection in videos via boosting and fusing statistical features of appearance, shape and motion dynamics on Riemannian manifolds with applications to assisted living. Comput Vis Image Underst 148:111–122
57. Zerrouki N, Houacine A (2017) Combined curvelets and hidden Markov models for human fall detection. Multimed Tools Appl 77(5):6405–6424
58. Zhang Y, Zhou D, Chen S, Gao S, Ma Y (2016) Single-image crowd counting via multi-column convolutional neural network. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 589–597. https://doi.org/10.1109/CVPR.2016.70
59. Zhong Z, Christopher C, Vassilis A (2015) A survey on vision-based fall detection, Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, doi: 10.1145/2769493.2769540
60. Zhou X, Qian L-C, You P-J, Ding Z-G, Han Y-Q (2018) Fall detection using convolutional neural network with multi-sensor fusion. 2018 IEEE international conference on Multimedia & Expo Workshops (ICMEW). https://doi.org/10.1109/icmew.2018.8551564

Dr. Chadia Khraief received the Engineer's degree and Master degree in Computer Science from the National Engineering School of Tunis (ENIT) respectively in 2008 and 2011. Currently, she is pursuing the Ph.D. degree in the National Engineering School of Tunis (ENIT) and she is a member of research group in the Image, Signal and Pattern Recognition SITI Laboratory. Her current research interests include human detection, activity recognition, Pattern Recogntition, Deep Learning, Image Processing and Video Analysis



Prof. Faouzi Benzarti is a professor in the High School of Techniques and Sciences of Tunis(ESSTT). He received Engineer's degree in Electrical Engineering from the National Engineering School of Monastir, and his M. S degree in Biomedical Engineering from the Polytechnic School of Montreal CANADA (Ecole Polytechnique de Montréal). He obtained his Ph.D degree in Electrical Engineering from the National Engineering School of Tunis (ENIT) in 2006. He is presently a member of research group in the Image, Signal and Pattern Recognition SITI Laboratory. His current researches include: Human activity recognition, Image De convolution, Image In painting, Image retrieval, Image segmentation, face recognition, Video Analysis and 3Dimage reconstruction.

Prof. Hamid Amiri received the Diploma of Electrotechnics, Information Technique in 1978 and the PhD degree in 1983 at the TU Braunschweig, Germany. He obtained the Doctorates Sciences in 1993. He was a Professor at the National School of Engineer of Tunis (ENIT), Tunisia, from 1987 to 2001. From 2001 to 2009 he was at the Riyadh College of Telecom and Information. Currently, he is again at ENIT. His research is focused on Image Processing, Video Analysis, Speech Processing, Document Processing and Natural language processing.