Comparative Study Of Computer Vision Algorithms For Fall Detection

Estudio Comparativo De Algoritmos De Visión Computacional Orientados Al Reconocimiento De Caídas

José Camilo Eraso Guerrero [iD]
Universidad CESMAG, Colombia

Elena Muñoz España [iD] **, Mariela Muñoz Añasco** [iD]
Universidad del Cauca, Colombia

Miller Manuel Rúales Luna [iD]
Universidad CESMAG, Colombia

Abstract

Objective: To compare the performance of three human fall recognition algorithms, focusing on computer vision. The comparison will be carried out by evaluating their performance on various databases commonly used by the scientific community, as well as on a new database called CAUCAFall. **Methodology:** The study compares three algorithms, selected through a systematic review that considered articles working with computer vision, RGB cameras and public databases. The selected algorithms focus on feature extraction and convolutional neural networks using YOLO and OPENPOSE. **Results:** The study found that all three algorithms perform well on databases commonly used by the scientific community. However, inferior performance was observed when evaluating the algorithms on CAUCAFall, which contains loosely controlled environments closer to reality. **Conclusions:** The research highlights the importance of evaluating human fall recognition algorithms in more realistic scenarios. It raises the importance of future research that focuses on creating and evaluating algorithms on databases that contain scenarios closer to reality, which would be a significant advance in the area of human fall recognition.

Keywords: Human fall recognition, computer vision, feature extraction, convolutional neural networks, YOLO, OPENPOSE, uncontrolled environments.

Resumen

Objetivo: Comparar el rendimiento de tres algoritmos de reconocimiento de caídas humanas, centrados en visión computacional. La comparación se llevará a cabo evaluando su desempeño en diversas bases de datos comúnmente utilizadas por la comunidad científica, así como en una nueva base de datos denominada CAUCAFall. **Metodología:** En el estudio se comparan tres algoritmos, seleccionados mediante una revisión sistemática que consideró artículos que trabajan con visión artificial, cámaras RGB y bases de datos públicas. Los algoritmos seleccionados se centran en extracción de características y redes neuronales convolucionales usando YOLO Y OPENPOSE. **Resultados:** El estudio encontró que los tres algoritmos tienen buen rendimiento en bases de datos comúnmente utilizadas por la comunidad científica. Sin embargo, se observó un desempeño inferior al evaluar los algoritmos en CAUCAFall, que contiene entornos poco controlados y más cercano a la realidad. **Conclusiones:** La investigación destaca la importancia de evaluar algoritmos de reconocimiento de caídas humanas en escenarios más realistas. Se plantea la importancia de futuras investigaciones que se centren en crear y evaluar algoritmos en bases de datos que contengan escenarios más cercanos a la realidad, lo que sería un avance significativo en el área de reconocimiento de caídas humanas.

Palabras claves: Reconocimiento de caídas humanas, visión computacional, extracción de características, redes neuronales convolucionales, YOLO, OPENPOSE, entornos no controlados.

Introduction

The World Health Organization (WHO) [1] predicts a substantial increase in the world population over 60 years of age, from 11% to 22% between 2000 and 2050. This is equivalent to an increase in the number of older adults from 605 million to 2 billion people. This increase in the older adult population requires special attention, because as people age, the incidence of falls increases [2]. A fall, as defined in [3], is an unexpected incident in which a person falls to the ground due to a push, environmental obstruction, loss of consciousness, or any similar problem related to limitations or health disorders. This involves an unintentional change in the person's posture, resulting in the person ending up on the ground.

Falls are the second most common cause of accidental deaths worldwide and a leading cause of injury or disability. In the United States, every 11 seconds, an elderly person who has fallen is taken to an emergency room and every 19 minutes, one such person dies. It is predicted that by 2030 in the United States, one elderly person could die every hour due to a fall. In 2015, the medical cost of falls in that country exceeded $50 billion [3, 4].

In addition, a large percentage of elderly persons live or remain alone in their homes. Therefore, if they suffer a fall and do not receive immediate attention, they can suffer serious injuries and even death. Automatic detection of falls or movements that may affect health can significantly reduce post-incident consequences and can also track abnormal daily behavior patterns in adults at high risk for falls, alerting them to abnormalities [5].

Therefore, in the last decade, there has been an increase in research focused on human fall detection, using a variety of devices and technologies. These include wearable devices, such as triaxial accelerometers or gyroscopes, as well as environmental sensors, such as acoustic sensors used to detect and measure the sounds of falls, and pressure sensors to measure changes in weight on the ground. However, these technologies can often be expensive and impractical, as many factors can alter the signals from the sensors [4].

In contrast, the application of computer vision, which employs cameras, videos, and images, is another method for detecting human falls. This article focuses on this approach, as it presents several advantages over other detection systems, according to relevant literature [6, 7]. Not only is it non-intrusive, but it is also more resilient to the effects of environmental noises. Furthermore, the vision-based method enables the retrieval of significant amounts of data, including speed, angles, rotations, as well as time and distances of objects or people in front of the camera. Moreover, the vision-based method has limitations, such as compromising privacy by constantly running a camera. Additionally, the method has high computational costs, making it difficult to run in real-time with resource-limited devices, and the performance of this technique is highly dependent on the camera position, as highlighted in [8].

While some applications of computer vision fall recognition have demonstrated promising results, these evaluations are often conducted in highly controlled environments. Research studies by [9] and [10] have focused on the performance of human activity detection algorithms in less controlled environments. They have revealed that most research focusing on human activity recognition utilize short video segments taken under optimal conditions without obstructions, lighting fluctuations, changes in room texture or fall angles, or clothing that blends with the environment. Debard & Mertens, et al. [9] selected algorithms that exhibit a good percentage of activity recognition when used with databases created in controlled, actuated scenarios. Then, they implemented them in real environments with real

falls of elderly people, concluding that they did not present the same efficiency, since their development had not considered image quality, overexposure problems, obstructions, and changes in lighting conditions.

This paper proposes the comparison of three computer vision-based algorithms ([11, 12, 13]) for human fall recognition. Their performance is evaluated on public databases widely used by the human fall research community, as well as on the CAUCAFall database [14]. The latter database simulates realistic environments with different obstructions, lighting changes (natural, artificial, and nighttime), varying fall angles, different floor and room textures, various clothing items, individuals of different ages, weights, and heights, and even different dominant limbs (see Figure 1). In addition, this database is the only one that provides the intensity of illumination in the environments and the distances from the fall location to the camera position. The objective is to analyze the performance of the selected algorithms in an environment as close as possible to reality.

Figura 1. CAUCAFall with different environments, participants, lighting, occlusion, fall angles, distance to the camera and distracting elements.
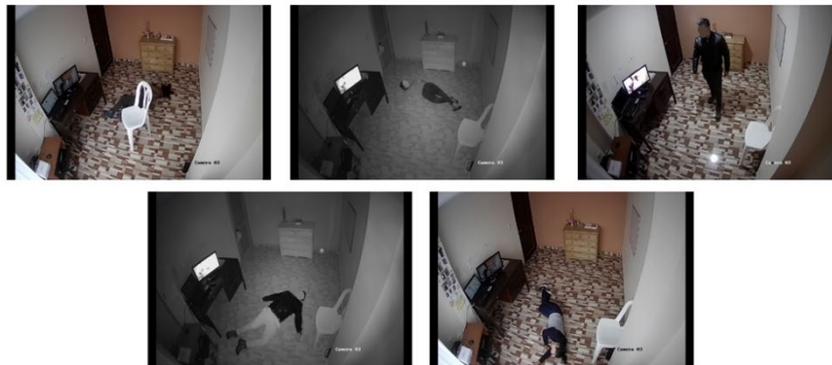
Source: [14]

The paper is organized as follows: Section 2 provides a comprehensive review of recent studies on fall detection, highlighting the advantages and disadvantages of the proposed approaches. Section 3 outlines the methodology employed to select the algorithms for comparison. Section 4 describes the experiments performed. Section 5 presents the results and subsequent discussion. Finally, Section 6 the paper concludes with contributions and future work.

Related Works

Successful feature extraction from images or videos is essential for accurate fall recognition [7, 15]. This review focuses on detection methods that employ both local and global feature extraction, depth-based representations, and novel convolutional neural network (CNN) based approaches.

Several authors [7, 16, 17, 18] use global feature extraction to locate a subject in images or videos, isolate it from the background and acquire its silhouette. From here, they perform various feature extraction, such as center of mass calculation, identification of human body orientations [7, 17], ellipse delineation [19], volume calculation [20] and optical flow analysis [21], in order to detect a fall. However, according to [22], most of these methods are sensitivity to noise, obstructions, and changes in viewpoint, resulting in significant challenges.

Other authors [6, 23, 24] prefer to perform local feature extraction, which, according to [25], are fundamental techniques for extracting robust points of interest from a video or image, such as corners, lines, curves, and isolated points where maximum

or minimum intensity is reached. This contributes to the recognition of various visual features. [23] extracts points of interest and analyzes the image histogram to identify actions performed by the individual, including falls. On the other hand, [26] focuses on action recognition by encoding local 3D spatio-temporal gradient features within a sparse coding framework. In this method, each local spatio-temporal feature is transformed into a linear combination of a few "atoms" in a dictionary trained to detect local motion and appearance features. Although this approach provides higher scale invariance and can recognize some basic activities, [27] finds difficulties in the method due to its sensitivity to changes in camera view, background motion, and camera motion.

In recent years, the use of CNNs in human fall recognition has gained increasing importance and impact. These networks are supplied with images containing information such as optical flow [28, 29, 30], bone maps of the human silhouette [13] and depth maps [31, 32]. Three-dimensional networks [33, 34] and multi-stream CNNs [35] are also used. However, [36] notes that despite the success of CNN applications in fall recognition, their effectiveness is limited to highly controlled environments and none of these networks demonstrate flexibility to perform adequately outside their domain.

It is important to note that while previous research has achieved high fall recognition rates, the database environments used are often tightly controlled [37]. The authors' purpose is to evaluate the performance of these approaches using a database that reflects characteristics of real-world live environments.

Methodology

This research performs the comparison of three computer vision algorithms oriented to fall recognition. This section describes the process that was carried out for their selection and implementation using different databases.

Selection of implemented algorithms

For the present work, the research of authors Gutiérrez, Rodríguez and Martín, published in [38], is taken as a starting point. These authors conducted a comprehensive review of computer vision-based fall detection systems in public databases, including resources such as ScienceDirect, IEEE Explorer and the Sensors database. In addition, they supplemented their search with academic literature focused on healthcare, drawing on sources such as PubMed and MedLine.

The literature review identified 929 papers utilizing the terms "fall detection" and "vision". Of these, 499 were discarded, as the titles of those investigations did not align with the content needed for the study in question. The remaining 430 papers were then further filtered to exclude investigations that focused on fall prevention or the recognition of human activities other than falls. Research that utilized mixed technologies, including sensors other than cameras, was excluded. This process resulted in a selection of 81 investigations that met the criteria for fall detection using computer vision.

To further refine the search, this research applied an additional filter to the 81 investigations. Those that employed and implemented RGB images were specifically

José Camilo Eraso Guerrero, Elena Muñoz España, Mariela Muñoz Añasco, Miller Manuel Rúales Luna

selected, as this work is based on RGB images. This second filter reduced the number of investigations to 50, which were used to identify the databases most commonly used by the scientific community in fall detection with RGB images. The results of this process are presented in Table 1.

Table 1. Databases used by the scientific community.

Database	Number of investigations	%
UR Fall Detection [39]	21	42%
Multicam Fall Dataset [40]	9	18%
LE2I [41]	8	16%
Fall Detection Dataset [42]	7	14%
UP-Fall [43]	1	2%
MOT Dataset [44]	1	2%
COCO Dataset [45]	1	2%
Center For Digital Home [46]	1	2%
ntu rgb+d [47]	1	2%

Source: Own elaboration

However, not all the databases mentioned in Table 1 are publicly accessible, which prevents a performance comparison with CAUCAFall. For this reason, a filter was applied to select only journal articles that were the result of research conducted with at least one publicly accessible database. This resulted in the identification of 12 research studies, which are detailed in Table 2, along with a brief description of their methodology and the databases they used.

Subsequently, criteria such as the innovativeness of the algorithms proposed in the investigations and the amount of information provided by the authors to enable their implementation were evaluated. Following these selection criteria, 3 investigations were identified ([11, 12, 13]) to be used in the comparison presented in this paper.

Table 2. Research pre-selected for performance analysis

Work	Data	UR Fall Detection	Multicam Fall Dataset	LE2I	UPFALL
F. Harrou et al. [48]	Background subtraction / depth characterization	✓	-	-	-
Syed F. Ali et al. [49]	Background subtraction (GMM) / global Characterization	✓	✓	-	-
D. Kumar et al. [50]	Silhouette segmentation / global characterization / silhouette center /angular velocity determined by long short-term memory (LSTM)	✓	-	-	-
F. Harrou et al. [51]	Background subtraction / global characterization	✓	-	-	-
Swe N. Htun et al. [11]	Background subtraction / global characterization	-	-	✓	-

Yaxiang Fan et al. [52]	CNN / local characterization	-	✓	✓	-
W. Min et al. [53]	Object recognition through CNN / local characterization	✓	-	-	-
Chao Ma et al. [54]	Feature maps obtained through CNN / local characterization	✓	✓	-	-
Ricardo Espinosa et al. [55]	Global characterization / feature maps obtained through CNN / local characterization	-	-	-	✓
B. Wang et al. [56]	Human keypoints identified by OpenPose / DeepSORT / local characterization	✓	-	✓	-
Qi Feng et al. [12]	Feature maps obtained through of CNN and LSTM / local characterization	✓	✓	-	-
Qingzhen Xu et al. [13]	Human keypoints identified by OpenPose and CNN / local characterization	✓	✓	-	-

Source: Own elaboration

Feature Extraction

The first algorithm implemented in this research is based on the work done by Htun, Zin and Tin in [11]. In their study, they performed feature extraction from videos and y used the Hidden Markov Model to distinguish crashes from normal activities. Moreover, the same authors in [57] proposed the use of SVM (Support Vector Machine) as a classifier. In our implementation, we selected the best features of both researches as the basis of the algorithm to be used and compared.

The proposed monitoring system is based on the selection of a specific mixture of Gaussians (MoG) distribution to model the foreground, which is updated frame by frame. A specific low-rank subspace learning is also used to model the background in each successive frame of the video sequences. Subsequently, an Expectation and Maximization (EM) algorithm is used to update the foreground and background parameters in each new frame. In this way, we proceed to feature extraction, such as the distance between a virtual grounding point and the centroid of the human silhouette, the area of the human silhouette, and its aspect ratio (which is the ratio between the width and length of the silhouette).

Then, we conducted an analysis of abnormal events and their classification using SVM. It is important to note that the extracted features have specific thresholds that determine when a fall is considered to have occurred. The mathematical and statistical calculations are detailed in the original research ([11, 57]).

Figure 2 shows the algorithm at work using some of the most popular public databases, which were obtained from Table 1.

Figura 1. (a) fall in CAUCAFall, (b) no fall in CAUCAFall, (c) fall in URFD, (d) no fall in URFD, (e) no fall in le2i, (f) fall in MCF.
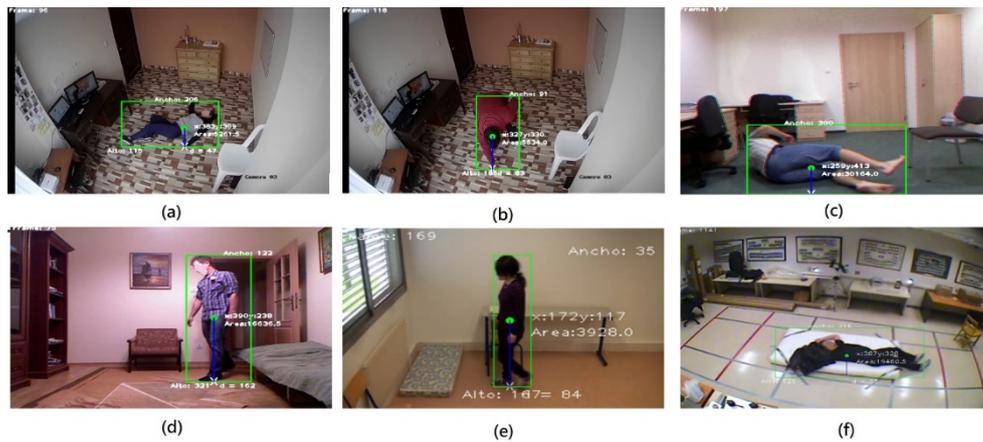
Source: Own elaboration

YOLO / CNN / LSTM

The second fall recognition algorithm implemented in this paper is based on the research proposed by Feng and Gao, et al [12]. The authors also mention that there is little research on fall detection in complex scenarios and propose their research as a robust solution for poorly controlled situations.

[12] It uses YOLO (You Only Look Once) V3 [58] to detect people in the videos and is monitored by a Deep Sort tracking module. Subsequently, features are extracted from the frames through a convolutional neural network (CNN) pre-trained with a VGG model. Finally, drop events are detected by attention-guided LSTM. For specific implementation parameters, we invite to review the authors' publication [12]. It is also important to mention that the authors' proposal detects pedestrians only when there are fall events. In our work, we detect fall events and non-fall events. Figure 3 shows the visualisation of the detection results of the proposed algorithm in different databases. In the figure, green marked falls and pink marked non-falls are recognised.

Figura 2. (a) fall in CAUCAFall, (b) no fall in CAUCAFall, (c) fall in URFD, (d) no fall in le2i, (e) fall in MCF, (f) no fall in MCF.
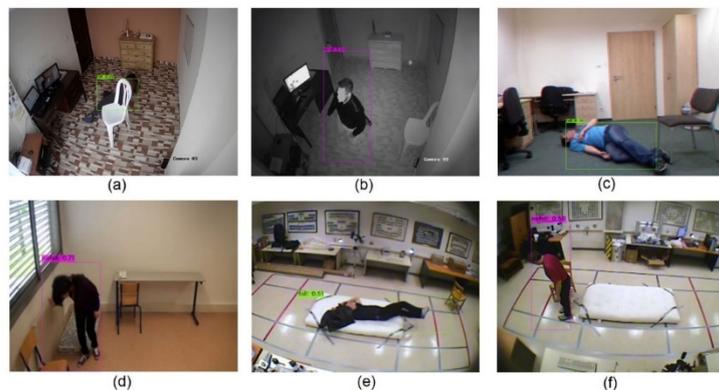
Source: Own elaboration

OpenPose/CNN

Finally, as a third implementation algorithm, an innovative research presented in [13] has been selected. Its fall prediction is based on the extraction of human skeleton maps from 2D images using OPENPOSE [59]. This allows the creation of a dataset to develop a new model using convolutional neural networks and transfer learning, specifically the Inception_ResNet_V2 model. In the end, it is determined whether or not it is a human fall. Figure 4 shows the visualisation of the OPENPOSE results in various databases.

Figura 3. (a) y (b) fall in CAUCAFall, (c) y (d) no fall in URFD, (e) y (f) no fall in le2i, (g) y (h) no fall in MCF.

(a) (b) (c) (d)

(e) (f) (g) (h)
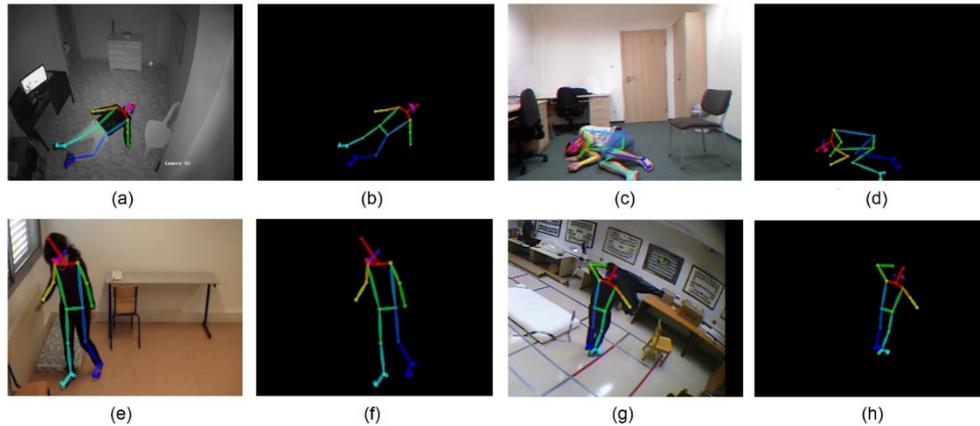
Source: Own elaboration

Proposed Experiments

In order to compare the performance of the three algorithms ([11, 12, 13]) in human fall detection, the most popular public databases in the scientific community specialised in fall detection, such as UR Fall Detection [39], Multicam Fall Dataset [40], LE2I [41], UP-Fall [43], as well as the inclusion of CAUCAFall [14] have been used to evaluate their performance in an uncontrolled environment [60]. The following experiments have been designed:

(a) Experiment 1: Each of the algorithms will be evaluated with each of the aforementioned databases individually. In this process, 50% of the data will be allocated for training, 25% for validation and 25% for testing.

(b) Experiment 2: Since the objective is to analyse whether research using databases of human falls in controlled environments is able to generalise predictions of falls in realistic situations, the model will be trained using a combination of different databases ([39, 40, 41, 43]), and then the model will be tested in each of the conditions proposed by CAUCAFall, separately.

(c) Experiment 3: The same model trained in experiment 2 will be evaluated in the full CAUCAFall.

(d) Experiment 4: The model will be trained using the full CAUCAFall and its performance will be evaluated using the various proposed databases ([39, 40, 41, 43]).

The architectures were implemented in the Python language, using libraries such as OpenCV, TensorFlow, Keras, Scikit-learn, for the different stages of the

José Camilo Eraso Guerrero, Elena Muñoz España, Mariela Muñoz Añasco, Miller Manuel Rúales Luna

implementation, such as background subtraction, feature extraction, model training, SVM classifier, among others. For the implementation of [11], which incorporates feature extraction techniques, and to test and evaluate the models based on convolutional neural networks of [12] and [13], an ACER laptop with Intel(R) Core i7 CPU at 2.80 GHz and 8 GB of RAM, which also has a GeForce® MX330 graphics card, was used. To train the different models implemented in this research, the Google Colaboratory tool was used.

The performance criteria to evaluate the efficiency of the proposed methods using the different databases is based on True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), which allow the calculation of the indicators of Accuracy, Sensitivity (Recall) and F-Score (F-Score), as shown in the following equations:

$$Accuracy = \frac{TP}{TP + FP} \quad , \quad Recall = \frac{TP}{TP + FN} \quad ,$$
$$FScore = \frac{2 * Acc * Recall}{Acc + Recall} \tag{1}$$

Results
Results Experiment 1
As can be seen in Figure 2, Figure 3 and Figure 4, the algorithms were implemented on the different databases selected for this research. The results obtained are presented in Table 3.

Table 3. Results Experiment 1

Algorithm	Database	Accuracy	Recall	F-Score
Feature Extraction	CAUCAFall	87.02%	86.26%	85.15%
	UR Fall Detection	93.61%	93.70%	93.61%
	Multicam Fall Dataset	96.82%	96.70%	96.61%
	LE2I	97.43%	96.95%	97.07%
	UP-Fall	99.23%	99.23%	99.22%
YOLO/CNN	CAUCAFall	91.93%	91.76%	91.78%
	UR Fall Detection	96.20%	96%	96.04%
	Multicam Fall Dataset	92.59%	92.25%	92.20%
	LE2I	96.75%	96.75%	96.74%
	UP-Fall	100%	100%	100%
OpenPose/CNN	CAUCAFall	97.25%	97.25%	97.25%
	UR Fall Detection	87.88%	87.75%	85.40%
	Multicam Fall Dataset	90.05%	88.5%	88.46%
	LE2I	96.52%	96.25%	96.25%
	UP-Fall	100%	100%	100%

Source: Own elaboration

Table 3 presents the results of Experiment 1. It can be seen that the feature extraction algorithm achieves higher accuracy with the UP-Fall database [43]. This database is characterised by a highly controlled environment, with the use of 2 video cameras, no scene changes, no occlusions, no lighting variations, and only 2 fall angles that are recorded at a constant distance from the camera. In addition, some of the participants simulate falls by wearing reflective waistcoats, which contributes to a high performance in fall detection.

On the other hand, the lowest performance was obtained with CAUCAFall, the database proposed by the authors. This is because CAUCAFall has occlusions, movement in the background, different distances between the falls and the camera, which generates diversity in features such as area, centroid-to-ground distance and aspect ratio, making it difficult to detect falls.

The algorithm based on YOLO detectors and convolutional neural networks again shows outstanding performance using UP-Fall [43] (see Table 3), due to the recognition facilities it offers. In contrast, CAUCAFall [14] still has the lowest performance. The difference lies in the fact that CAUCAFall considers partial occlusions (see Figure 3 (a)), variations in illumination (see Figure 5), different fall angles (see Figure 6) and various distances from the camera (see Figure 7).

Figura 4. (a) Natural lighting, (b) Artificial lighting, (c) No lighting.

Source: [14]

Figura 5. (a) fall with 47.8° angle, (b) fall with 182.3° angle, (c) fall with 277.5° angle.

Source: [14]

Figura 6. (a) fall distance: 2.7m, (b) fall distance: 3.12m, (c) fall distance: 3.4m.

Source: [14]

In addition, the results of the evaluation of the databases in the OpenPose-based algorithm are presented in Table 3. The performance of this algorithm depends on the high quality of the videos and frames, which must be of high definition. In addition, OpenPose faces difficulties when dealing with occlusions, as can be seen in Figure 8. It also requires good lighting or adequate segmentation of the human silhouette. In case the image is dark, the algorithm will not be able to identify the bone map of the subject. In contrast, CAUCAFall is a database created with high-resolution cameras, which facilitates the detection of the human skeleton more effectively. In addition, it works with a camera that incorporates night vision, which ensures the availability of images in which the human body can be recognised, even in environments with little or no illumination.

Figura 7. OpenPose with occlusions.
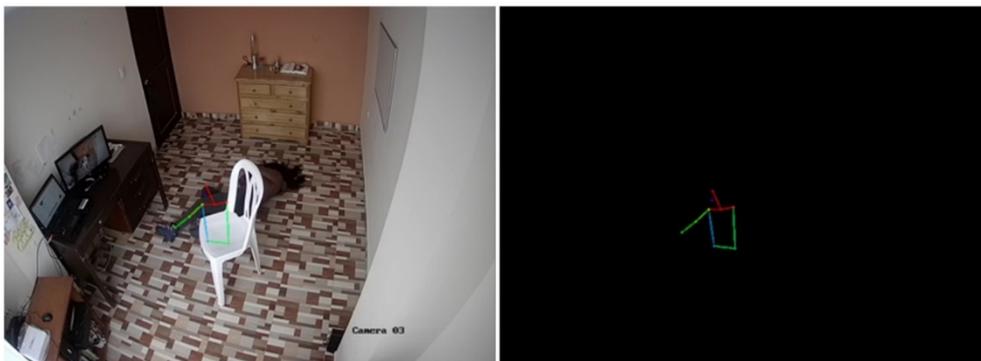
Source: Own elaboration

However, the most outstanding results are observed when using UP-Fall [43], as this database is also created with a high-resolution camera. Moreover, it is characterised by considering only two fall angles, which simplifies its recognition. Although neither of these two angles is optimal for OpenPose, leading to the loss of bone points in the human silhouette, excellent accuracy is achieved because the algorithm is trained, evaluated and tested on the same images, as shown in Figure 9.

Figura 8. Example of fall recognition in UP-Fall.
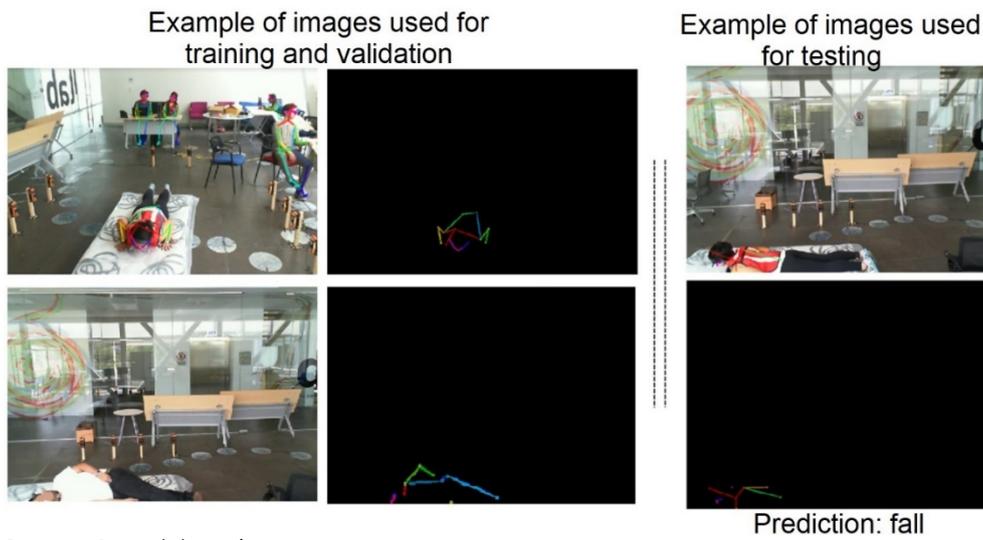
Results Experiment 2

Table 4 details the results of experiment 2, using the different algorithms (Alg1 [11], Alg2 [12], Alg3 [13]).

Table 4. Results Experiment 2

Conditions CAUCAFall	Accuracy			Recall			F-Score		
	Alg 1	Alg 2	Alg 3	Alg 1	Alg 2	Alg 3	Alg 1	Alg 2	Alg 3
Natural Light	84.46 %	92.41 %	90.39 %	79.53 %	91.75 %	90%	76.79 %	91.72 %	89.97 %
Artificial Light	91.69 %	89.9%	87.95 %	90.79 %	89.5%	87.75 %	89.97 %	89.47 %	87.73 %
No Light	88.57 %	91.68 %	94.72 %	86.46 %	92%	94.5%	84.71 %	91.64 %	94.49 %
Occlusions	86.91 %	71.12 %	74.57 %	65.92 %	69.5%	73.87 %	68.80 %	70.35 %	74.21 %
Camera-Fall distance									
2.27 m – 2.66 m	97.70 %	86.31 %	90.56 %	42.38 %	85.25 %	85.25 %	56.92 %	85.70 %	86.65 %
2.67 m – 3.04 m	96.64 %	91%	97.47 %	68.91 %	92%	97.5%	78.43 %	91.53 %	97.48 %
3.05 m – 3.43 m	89.43 %	83.80 %	91.08 %	60.19 %	79.5%	86.75 %	65.71 %	81.07 %	87.90 %
Falling Angles									
0º - 90º	96.39 %	90.52 %	96.77 %	70.12 %	91%	96.75 %	79.09 %	90.62 %	96.75 %

91º - 180º	82.41 %	67.94 %	85.86 %	64.83 %	62.25 %	80.25 %	64.15 %	63.60 %	80.96 %
181º - 270º	95.78 %	82.76 %	97.59 %	94.73 %	83%	97.5%	94.93 %	82.82 %	97.4%
271º - 360º	95.84 %	85.92 %	88.77 %	29.83 %	84.5%	78.75 %	40.64 %	85.08 %	81.27 %

Source: Own elaboration

Experiment 2 consists of training the models with various databases and evaluating their performance under different conditions provided by CAUCAFall. No other research presents fall recognition results that detail the conditions as well as this study does. The results when using the feature extraction algorithm in various conditions vary. Good fall recognition is achieved in both artificial light and unlit environments due to the night vision of the CAUCAFall camera. However, the performance is inferior in occlusive situations and when falls occur very close to or very far from the camera. In addition, inferior performance is observed when falls occur at angles between 91° and 180°, as well as between 271° and 360°. This is because the databases used to train the model do not contain enough examples of these conditions.

When we compare the feature extraction algorithm with the algorithm using YOLO and convolutional neural networks (see Table 4), we notice that fall recognition decreases in occlusion situations, but improves at distances near and far from the camera, as well as at different fall angles. In terms of lighting conditions, similar performance is obtained, thanks to the camera's night vision sensors, which allow capturing images with good definition even in the dark.

On the other hand, the performance of OpenPose in the CAUCAFall evaluation increases performance, as the model focuses on learning the shape of human skeletons, unaffected by environmental distractions. OpenPose only requires sharp, high-resolution images to obtain the bone map of the subject, which results in good performance, despite its high computational cost. However, problems caused by occlusion persist.

Results experiment 3

Table 5 details the results of experiment 3, using the different algorithms.

Table 5. Results experiment 3

Algorithm	Accuracy	Recall	F-Score
Feature Extraction [11]	88.57%	86.52%	84.96%
YOLO/CNN [12]	88.03%	87%	87.19%
OpenPose [13]	87.63%	82.75%	83.14%

Source: Own elaboration

The third experiment uses the model trained with the commonly used databases and is evaluated on the entire CAUCAFall. When comparing the performance indices in Table 5, very similar performance is observed. However, the best performance is achieved by the algorithm employing convolutional neural networks with YOLO,

followed by feature extraction and, ultimately, the OpenPose-based algorithm. This is because most of the images in the databases ([39, 40, 41, 43]) do not have a high resolution, which makes it difficult to correctly extract the bone map of the human silhouette, which in turn complicates the training of the network.

Results experiment 4

Table 6 details the results of experiment 4, using the different algorithms.

Table 6. Results experiment 4

Algorithm	Accuracy	Recall	F-Score
Feature Extraction [11]	90.21%	91.37%	90.83%
YOLO/CNN [12]	93.58%	92.87%	93.32%
OpenPose [13]	94.58%	93.5%	94.07%

Source: Own elaboration

In the last experiment, the different models are trained and validated using the whole CAUCAFall and evaluated with the proposed databases ([39, 40, 41, 43]). Table 6 clearly shows that the performance increased significantly, demonstrating that CAUCAFall meets the necessary conditions to allow, when used as a training dataset, the models to be able to generalise human fall recognition. All three algorithms are able to extract the most relevant features from the images and learn them, allowing them to recognise falls at different angles, distances and lighting conditions in a satisfactory way.

On this occasion, the algorithm combining OpenPose with CNN shows the best performance, and much of its success is due to the quality of the images in CAUCAFall. In addition, the bone maps extracted from CAUCAFall, having variations in size, angles and distances, are suitable for generalising and predicting falls in other databases.
Finally, it is important to mention that in order for OPENPOSE and YOLOV3 to work in real time for fall recognition using computer vision, higher hardware capabilities are required. According to the author [13], OpenPose needs the acceleration of four GPUs, and if a conventional laptop is used, each frame of the video will have a delay of 1 to 1.5 seconds, which affects the real-time effectiveness.

Conclusion

In this paper, a review of the current state of the art of fall detection in elderly people has been carried out, focusing on advances related to computer vision. Our intention was to evaluate and highlight the performance of some innovative approaches in the field of fall detection, particularly in a database that simulates realistic and poorly controlled environments.

To achieve this, we designed and created the CAUCAFall database, which brings together the features of the most popular public databases in the scientific community. Our intention was to create realistic and detail-rich environments. This database includes individuals of different ages, weights, heights and dominant legs. The data was collected using an RGB camera in a home environment that mimicked realistic, uncontrolled conditions. It included elements such as occlusions, lighting changes (natural, artificial and night-time), varying clothing among participants, movement in the background, varying textures on the floor and in the room, and a variety of fall angles and distances from the camera. CAUCAFall is unique in that it

provides details on the intensity of the lighting in the scenarios, the distances between the human fall and the camera, and the angles of the falls with respect to the camera. It is also the only database that labels each image, identifying frames with human falls as "fall" and everyday activities as "non-fall". These labels are especially useful for algorithms employing YOLO detectors. The high resolution of CAUCAFall allows novel algorithms such as OpenPose to extract more information from the images.

In this study, we evaluate the performance of three fall detection proposals using different databases. These proposals are based on feature extraction, YOLOV3 together with convolutional neural networks, and OpenPose with convolutional neural networks. Our findings reveal that, in the case of feature extraction, performance is optimal in databases with highly controlled environments, i.e. those without changes in lighting, variable fall angles, occlusions, changes in scenery, movement in the background, variable distances from the camera, and changes in textures of both the environment and the participants' clothing. This suggests that these methods are highly sensitive to noise, occlusions and variations in viewpoint.

On the other hand, proposals that rely on Convolutional Neural Networks for their investigations have superior performance in uncontrolled environments compared to feature extraction. However, the computational cost is significantly higher, which makes them almost impossible to use in real-time fall detection applications. In addition, large and varied datasets containing examples with different postures, distances, angles, occlusions with various objects, various environments and illuminations are required for the models to perform well.

The high resolution of CAUCAFall allowed the use of modern algorithms to recognise human bone maps from 2D images, which is an advantage over other databases containing human falls.

This study underlines the importance of an efficient fall detection system and highlights the great potential for research in this field. Although the results are promising with current techniques, the environments simulated in CAUCAFall can still be improved, as they do not fully reflect the reality of human falls. It is considered of great impact that the scientific community presents research results in more realistic environments, such as those offered by CAUCAFall. In addition, it is essential to develop databases containing real adult falls in uncontrolled environments.

As future work, we plan to evaluate the proposed methods using at least two cameras, incorporate techniques such as object recognition in scenarios and analyse the speed of different human silhouettes. We also intend to evaluate novel detection methods, such as YOLOV4 and YOLOV5, in the context of CAUCAFall.

In addition, it is essential to advance research into algorithms that can perform the same tasks as OpenPose and YOLO, but at a lower computational cost, so that they can be used in real time for everyday life applications. It would also be an important advance in research to incorporate multimodal techniques and to use depth cameras and non-invasive wearable sensors, such as smart watches, for human fall detection.

Bibliographic References

[1]   OMS | "Datos interesantes acerca del envejecimiento," WHO. (2015). En: https://www.who.int/ageing/about/facts/es/

[2]   G. Horng y K. Chen, "The Smart Fall Detection Mechanism for Healthcare Under Free-Living Conditions," Springer, Wireless Personal

Communications. 118:715–753, 2021. https://doi.org/10.1007/s11277-020-08040-4.

[3]   N. Thakur y C. Han, "A Study of Fall Detection in Assisted Living: Identifying and Improving the Optimal Machine Learning Method," J. Sens. Actuator Netw. 10, 39, 2021. https://doi.org/10.3390/jsan10030039

[4]   F. Shu y J. Shu, "An eight-camera fall detection system using human fall pattern recognition via machine Learning by a low-cost android box," Scientific Reports, 2021. 11:2471. https://doi.org/10.1038/s41598-021-81115-9

[5]   H. Li, H. Shrestha, F. Fioranelli, K. Le y H. Heidari, "Hierarchical Classification on Multimodal Sensing for Human Activity Recogintion and Fall Detection," 2018 IEEE SENSORS, pp. 1–4. 2018. Doi: 10.1109/ICSENS.2018.8589797

[6]   S. Amiri, M. Pourazad, P. Nasiopoulos y V. Leung, "Improved human action recognition in a smart home environment setting," IRBM. Vol. (35), no. 6, pp. 321–328, 2014.

[7]   M. Yu, Y. Yu, Y. Rhuma, S. Naqvi, L. Wang y J. Chambers, "An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment," IEEE J. Biomed. Heal. Informatics. Vol. (17), no. 6, pp. 1002–1014, 2013.

[8]   F. Concone, G. Re y M. Morana, "A Fog-Based Application for Human Activity Recognition Using Personal Smart Devices," ACM Transactions on Internet Technology. Vol. (19), no. 2, 2019.

[9]   G. Debard, M. Mertens, M. Deschodt, E. Vlaeyen, E. Devriendt, E. Dejaeger, K. Milisen, J. Tournoy, T. Croonenborghs, T. Goedemé, T. Tuytelaars, B. Vanrumste," Camera-based fall detection using real-world versus simulated data: how far are we from the solution?," Journal of Ambient Intelligence and Smart Environments, pp. 1-19, 2015.

[10]   Y. Fan, M. Levine, G. Wen, S. Qiu," A deep neural network for real-time detection of falling humans in naturally occurring scenes," Neurocomputing. Vol.260, pp.43-18, oct, 2017.

[11]   S. Nwe Htun, T. Zin y P. Tin. " Image Processing Technique and Hidden Markov Model for an Elderly Care Monitoring System," J. Imaging 6, 49, 2020. doi:10.3390/jimaging6060049

[12]   Q. Feng, C.Gao, L. Wang, Y. Zhao, T. Song, Q. Li, "Spatio-temporal fall event detection in complex scenes using attention guided LSTM", Pattern Recognition Letters .2020. doi: https://doi.org/10.1016/j.patrec.2018.08.031

[13]   Q. Xu, G. Huang, M. Yu, Y. Guo" Fall prediction based on key points of human bones," Physica A. 2019. https://doi.org/10.1016/j.physa.2019.123205

[14]   J.C. Eraso, E. Muñoz, M. Muñoz, J. Pinto, "Dataset for human fall recognition in an uncontrolled environment," Data in Brief, Volume 45, December 2022. https://doi.org/10.1016/j.dib.2022.108610

[15]   G. Goudelis, G. Tsatiris, T. Karpouzis y S. Kollias, "Fall detection using history triple features," 8th ACM Int. Conf. PErvasive Technol. Relat. to Assist. Environ. PETRA 2015 - Proc., 2015.

[16]   Z. Khan y W. Sohn, "Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care," IEEE Trans. Consum. Electron. Vol. (57), no. 4, pp. 2021.

[17]   M. Yu, S. Naqvi, A. Rhuma y J. Chambers, "One class boundary method classifiers for application in a video-based fall detection system," IET Comput. Vis. Vol. (6), no. 2, pp. 90–100, 2012.

[18]   H. Foroughi, B. Aski, y H. Pourreza, "Intelligent Video Surveillance for Monitoring Fall Detection of Elderly in Home Environment," Proc. 11th Int. Conf. Comput. Inf. Technol. pp. 219–224, 2008.

[19]   C. Rougier, J. Meunier, A. St-Arnaud y J. Rousseau, "Fall detection from human shape and motion history using video surveillance," Proc. - 21st Int. Conf. Adv. Inf. Netw. Appl. Work. AINAW'07, 2007.

[20]   E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau y J. Meunier, "Fall detection with multiple cameras: An occlusion-resistant method based on 3-D silhouette vertical distribution," IEEE Trans. Inf. Technol. Biomed. Vol. (15), no. 2, pp. 290–300, 2011.

[21]   A. Efros, A. Berg, G. Mori y J. Malik, "Recognizing action at a distance," Proc. IEEE Int. Conf. Comput. Vis., pp. 726–733, 2008.

[22]   S. Zhang, Z. Wei, J. Nie, L. Huang, L. Wang y Z. Li, "A Review on Human Activity Recognition Using Vision-Based Method," Journal of Healthcare Engineering, jul, 2017.

[23]   S. Venkatesha y M. Turk, "Human activity recognition using local shape descriptors," Proc. - Int. Conf. Pattern Recognit., pp. 3704–3707, 2010.

[24]   X. Peng, L. Wang, X. Wang y Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," Comput. Vis. Image Underst., 2016.

[25]   D.Dawn y S. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector," Vis. Comput., mar, 2015.

[26]   Y. Zhu, X. Zhao, Y. Fu y Y. Liu, Y, "Sparse coding on local spatial-temporal volumes for human action recognition," Springer-Verlag Berlin Heidelberg, pp. 660–671, 2011.

[27]   Zhang et al., "A comprehensive survey of vision-based human action recognition methods," Sensors (Switzerland). Vol. (19), no. 5, pp. 1–20, 2019.

[28]   Y. Hsieh y Y. Jeng, "Development of Home Intelligent Fall Detection IoT System based   on Feedback Optical Flow Convolutional Neural Network," IEEE Access. Vol. (6), pp. 6048 - 6057, 2017.

[29]   A.Núñez, G. Azkune, I. Arganda, " Vision-Based Fall Detection with Convolutional Neural Networks," Wireless Communications and Mobile Computing. 2017.

[30]   R. Espinosa, H. Ponceb, S. Gutiérreza, L. Martínez, J. Brievab, E. Moya, "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. 2019. doi: https://doi.org/10.1016/j.compbiomed.2019.103520

[31]   M. Rahnemoonfar, H. Alkittawi, "Spatio-temporal convolutional neural network for elderly fall detection in depth video camerasl," IEEE International Conference on Big Data, 2018.

[32]   K. Adhikari, H. Bouchachia y H. Nait, "Activity Recognition for Indoor Fall Detection Using Convolutional Neural Network," International Conference on Machine Vision Applications (MVA). 2017.

[33]   N. Lu, X. Ren, J. Song, Y. Wu, "Visual Guided Deep Learning Scheme for Fall Detection, "IEEE Conference on Automation Science and Engineering (CASE). 2017.

[34]   Y. Guo, A. Shimada, H. Uchiyama, C. Ma, H. Nagahara, R. Taniguchi, " CNN for face detection with thermal image," 2018(16・18-23):2018.3.26 p.5-8, 2018. http://id.ndl.go.jp/bib/028999107

[35]   C. Khraief, F. Benzarti y H. Amiri, " Elderly fall detection based on multi-stream deep convolutional networks," Multimedia Tools and Applications volume 79, pages19537–19560, 2020. https://link.springer.com/article/10.1007%2Fs11042-020-08812-x

[36]   A. Kaid, K. Baina, J. Baina, "Reduce False Positive Alerts for Elderly Person Fall Video-Detection Algorithm by convolutional neural network model," Procedia Computer Science. Vol. 149, pp.2-11, 2019.

[37] J. C. Eraso Guerrero, E. Muñoz España, y M. Muñoz Añasco, "Human Activity Recognition via Feature Extraction and Artificial Intelligence Techniques: A Review," Tecnura, vol. 26, n.º 74, pp. 213–236, oct. 2022. https://doi.org/10.14483/22487638.17413

[38] J. Gutiérrez, V. Rodríguez y S. Martin, "Comprehensive Review of Vision-Based Fall Detection Systems," Sensors 2021, 21, 947. 2021. https://doi.org/10.3390/s21030947

[39] M. Kepski, B. Kwolek, "Embedded system for fall detection using body-worn accelerometer and depth sensor," In Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Warsaw, Poland, 24–26 September 2015; Volume 2, pp. 755–759, 2015.

[40] E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, "Multiple Cameras Fall Data Set," Technical Report Number 1350; University of Montreal: Montreal, QC, Canada, 8 July, 2011.

[41] I. Charfi, J. Miteran, J. Dubois, M. Atri, R. Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: Comparison of support vector machine and Adaboost-based classification," J. Electron. Imaging 2013, 22, 041106,2013.

[42] K. Adhikari, H. Bouchachia, H. Nait-Charif, "Activity recognition for indoor fall detection using convolutional neural network," In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May, 2017.

[43] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, C. Peñafort-Asturiano, "UP-Fall Detection Dataset: A Multimodal Approach", Sensors 19(9), 1988: 2019. doi:10.3390/s19091988.

[44] MOT Dataset. Available online: https://motchallenge.net/ (accessed on 11 November 2023).

[45] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. Zitnick, "Microsoft coco: Common objects in context," In Computer Vision-

ECCV 2014, ECCV 2014. Lecture Notes in Computer Science; Springer: Cham, Swizerland, 2014; pp. 740–755

[46] CENTRE FOR DIGITAL HOME—MMU. Available online: http://foe.mmu.edu.my/digitalhome/FallVideo.zip (accessed on 11 November 2023)

[47] A. Shahroudy, J. Liu, T. Ng, G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1010–1019, 2016.

[48] F. Harrou, N. Zerrouki, Y. Sun, A. Houacine, "Vision-based fall detection system for improving safety of elderly people," IEEE Instrum. Meas. Mag. 2017, 20, 49–55, 2017. [CrossRef]

[49] S. Ali, R. Khan, A. Mahmood, M. Hassan, M. Jeon, "Using Temporal Covariance of Motion and Geometric Features via Boosting for Human Fall Detection," Sensors 2018, 18, 1918, 2018. [CrossRef]

[50] D. Kumar, A. Ravikumar, V. Dharmalingam, V. Kafle, "Elderly Health Monitoring System with Fall Detection Using Multi-Feature Based Person Tracking," In Proceedings of the 2019 ITU Kaleidoscope: ICT for Health: Networks, Standards and Innovation (ITU K), Atlanta, GA, USA, 4–6 December 2019. [CrossRef]

[51] F. Harrou, N. Zerrouki, Y. Sun, A. Houacine, "An Integrated Vision-Based Approach for Efficient Human Fall Detection in a Home Environment," IEEE Access 2019, 7, 114966–114974, 2019. [CrossRef]

[52] Y. Fan, M.D. Levine, G. Wen, S. Qiu, " A deep neural network for real-time detection of falling humans in naturally occurring scenes," Neurocomputing 2017, 260, 43–58.2017. [CrossRef]

[53] W. Min, H. Cui, H. Rao, Z. Li, L. Yao, "Detection of Human Falls on Furniture Using Scene Analysis Based on Deep Learning and Activity Characteristics," IEEE Access 2018, 6, 9324–9335. 2018. [CrossRef]

[54] C. Ma, A. Shimada, H. Uchiyama, H. Nagahara, R. Taniguchi, "Fall detection using optical level anonymous image sensing system," Opt. Laser Technol. 2019, 110, 44–61. 2019. [CrossRef]

[55] R. Espinosa, H. Ponce, S. Gutiérrez, L. Martínez-Villaseñor, J. Brieva, E. Moya-Albor, "vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection Dataset," Comput. Biol. Med. 2019, 115, 103520. 2019. [CrossRef]

[56] B. Wang, J. Yu, K. Wang, X. Bao, K. Mao, "Fall Detection Based on Dual-Channel Feature Integration," IEEE Access 2020, 8, 103443–103453. 2020. [CrossRef]

[57] S. Htun, T. Zin y H. Hama, "Virtual Grounding Point Concept for Detecting Abnormal and Normal Events in Home Care Monitoring Systems," Appl. Sci. 2020, 10, 3005; 2020. doi:10.3390/app10093005

[58] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement,". 8 Apr, 2018 https://arxiv.org/abs/1804.02767

[59] E. De la Hoz Domínguez, T. Fontalvo Herrera., y A. Mendoza Mendoza, "Machine Learning and SMEs: Opportunities for an improved decision -

making process", Investigación e Innovación en Ingenierías, vol. 8, n°. 1, 2020. DOI: https://doi.org/10.17081/invinno.8.1.3506

[60] Z. Cao, G. Hidalgo, T. Simon, S. Wei y Y. A. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. https://github.com/CMU-Perceptual-Computing-Lab/openpose