BIBLIOGRAPHIC INFORMATION SYSTEM

Journal Full Title: **Journal of Biomedical Research & Environmental Sciences**

Journal NLM Abbreviation: J Biomed Res Environ Sci

Journal Website Link: https://www.jelsciences.com

Journal ISSN: **2766-2276**

Category: Multidisciplinary

Subject Areas: Medicine Group, Biology Group, General, Environmental Sciences

Topics Summation: 133

Issue Regularity: Monthly

Review Process: Double Blind

Time to Publication: 21 Days

Indexing catalog: IndexCopernicus ICV 2022: 88.03 | GoogleScholar | View more

Publication fee catalog: Visit here

DOI: 10.37871 (CrossRef)

Plagiarism detection software: iThenticate

Managing entity: USA

Language: English

Research work collecting capability: Worldwide

Organized by: SciRes Literature LLC

IndexCopernicus
ICV 2022:
83.03

Manuscript should be submitted in Word Document (.doc or .docx) through

Online Submission

form or can be mailed to support@jelsciences.com

Vision: Journal of Biomedical Research & Environmental Sciences main aim is to enhance the importance of science and technology to the scientific community and also to provide an equal opportunity to seek and share ideas to all our researchers and scientists without any barriers to develop their career and helping in their development of discovering the world.

JOURNAL OF
BIOMEDICAL RESEARCH
issn: 2766-2276 & ENVIRONMENTAL SCIENCES
SUBJECT AREA(S): COMPUTER SCIENCE
DOI: 10.37871

REVIEW ARTICLE

Automated Falls Detection Using Visual Anomaly Detection and Pose-based Approaches: Experimental Review and Evaluation

Yuting Qiu[1], James Meng[2] and Baihua Li[1]*

[1]Department of Computer Science, Loughborough University, UK
[2]Norwich Medical School, University of East Anglia, UK

Abstract

Falls are a pervasive problem facing elderly populations, associated with significant morbidity and mortality. Prompt recognition of falls, especially in elderly people with cognitive or physical impairments who cannot raise the alarm themselves, is a challenge. To this end, wearable sensors can be used to detect fall behaviour, including smartwatches and wristbands. These devices are limited by their intrusiveness, require user compliance and have issues around endurance and comfort, reducing their effectiveness in elderly populations. They can also only target patients already recognised as falls risks, and cannot apply to non-identified patients. Leveraging state of the art AI deep learning, we introduce two types of automated fall detection techniques using visual information from cameras: 1) self-supervised autoencoder, distinguishing falls from normal behaviour as an anomaly detection problem, 2) supervised human posture-based fall activity recognition. Five models are trained and evaluated based on two publicly available video datasets, composed of activities of daily living and simulated falls in an office-like environment. To test the models for real-world fall detection, we developed two new datasets, including videos of real falls in elderly people, and more complex backgrounds and scenarios. The experimental results show autoencoder detectors are able to predict falls directly from images where the background is pre-learned. While the pose-based approach uses foreground body pose only for AI learning, better targeting complex scenarios and backgrounds. Video-based methods could be a potential for low-cost and non-invasive falls detection, increasing safety in care environments, while also helping elderly people retain independence in their own homes.

Introduction

The prevalence of falls in the older population has significant implications for both individual well-being and healthcare systems. According to the World Health Organization (WHO), falls are the second leading cause of death from accidental or unintentional injuries worldwide, with adults aged 65 years and older being the most affected [1]. In addition to the direct physical harm caused by falls, cognitive and physical impairments mean elderly people may not be able to alert others of their fall. Even in institutionalised settings like care homes, an

***Corresponding author(s)**

Baihua Li, Department of Computer Science, Loughborough University, UK

Email: b.li@lboro.ac.uk

DOI: 10.37871/jbres1872

Submitted: 30 December 2023

Accepted: 23 January 2024

Published: 24 January 2024

Keywords

- Falls detection
- Anomaly detection
- Deep learning
- Autoencoder
- Pose estimation
- Action recognition
- Deep neural networks

GENERAL SCIENCE GROUP
COMPUTER SCIENCE
VOLUME: 5 ISSUE: 1 - JANUARY, 2024

How to cite this article: Qiu Y, Meng J, Li B. Automated Falls Detection Using Visual Anomaly Detection and Pose-based Approaches: Experimental Review and Evaluation. J Biomed Res Environ Sci. 2024 Jan 24; 5(1): 055-063. doi: 10.37871/jbres1872, Article ID: JBRES1872, Available at: https://www.jelsciences.com/articles/jbres1872.pdf

elderly person who falls in their room at night may not be found until the next morning. The delay in raising the alarm and prompt medical treatment can contribute significantly to morbidity and mortality [2]. Thus prompt recognition of falls can ensure help arrives sooner, as well as helping elderly people retain independence safely at home.

Efforts to detect falls have historically focused on wearable devices [3,4]. While these solutions show some promise, they are not without significant drawbacks. Traditional wearables include emergency alarm devices with buttons that rely on in-person operation, or motion sensors such as accelerometers and gyroscopes that can be embedded in wearable devices. These devices are intrusive, bulky and need to be carried at all times, leading to discomfort, fatigue or inability to wear them for prolonged periods. Thus user compliance, particularly for older people with health and cognitive problems, is a major problem. Battery life is another practical issue, with users forgetting to charge or replace batteries, leading to failure of fall detection. Most importantly, only people already identified as high falls risk after medical assessment will likely be recommended for monitoring through wearable sensors.

Given these limitations, video-based fall detection techniques offer a non-intrusive method which can offer comprehensive coverage of multiple people without issues surrounding compliance. For example, they have the potential to be used in care homes to automatically monitor elderly people from afar. Early video-based activity recognition relies heavily on hand-crafted features and rule-based determination, such as the use of optical flow analysis, trajectory tracking and temporal difference [5]. Such methods are often sensitive to lighting changes, dynamic backgrounds and environmental conditions. They cannot effectively handle the variability in falling patterns and complexity in ambient background where multiple movements may occur simultaneously.

The most recent and effective deep learning-based techniques for activity recognition from videos can be classified into four categories: Convolutional Neural Network (CNN) based systems, Long Short-Term Memory (LSTM) based systems, autoencoder based systems and pose-based action recognition. CNNs are effective for finding patterns and shapes. They could address fall detection as a classification or detection task in images. The main advantage of LSTMs is their capability to deal with sequential data, such as fall detection from wearable sensors. Combinations of CNN and LSTM can be used for fall detection from time series videos. The combined CNN and LSTM could better address general problems in visual systems such as image noise, occlusion and camera perspective by incorporating temporal information. They are used to capture spatiotemporal features simultaneously, which allow adaptation of AI learning to individuals with unique movement characteristics as well as relationships between foreground movements and complex scenes [6]. CNN and LSTM architectures are mostly trained through supervised learning. As highlighted in [6], very limited work was observed in abnormality detection for action recognition including falls.

To bridge this gap, we explore autoencoder-based fall detectors and pose-estimation based approaches. Autoencoders [7] can learn normal behaviour features from images directly and thus distinguish contrasting anomalies. They can treat fall detection as an anomaly detection problem, distinguishing a variety of falls from normal behaviours. Autoencoders are trained in a self-supervised manner using plenty of data, without requiring hand-labelled images. Posture estimation is a key aspect of understanding human motion [8]. These techniques aim to accurately recognise and track the position and orientation of the body, providing the necessary contextual information for fall detection algorithms. In the common pose estimation algorithms, the human body pose is described in a graphic format. This means using coordinates to represent the nodes of the body and connecting them into a form that describes the posture.

A pose-based approach addresses fall detection as a supervised activity recognition task. The main advantage of the pose-based approach is that it only uses the extracted postures, removing the impact of background complexity, thus increasing fall detection robustness.

The effectiveness of visual fall detection is highly dependent on the quality and representativeness of the data used for training and evaluation. Publicly available datasets used to train AI models mainly consist of falls behaviours simulated by young people, due to a scarcity of data from older people, leading to a lack of generalisability. Thus the unique movement patterns, postures and behaviours associated with falls in older adults may not be adequately captured by datasets and the models trained from them.

Contributions

1) Our research highlights the effectiveness of state-of-the-art deep learning-based methods for fall detection from CCTV videos.

2) We trained and evaluated four autoencoder-based fall detectors in a self-supervised manner. The autoencoders aim to reconstruct an input image as the network output. During network training, the learning algorithm aims to minimise the reconstruction loss. When performing an anomaly fall detection, fluctuation of large reconstruction errors detects an anomalous situation.

3) We employed a Transfer Learning (TL) technique in the autoencoder training process, and validated model performance improvement on more complex video scenarios of elderly falls. Benefiting from pre-training on simple scenarios e.g. office environment and imitated falls, transfer learning largely reduced training data requirement for complex backgrounds, multi-people and real-world elderly falls which are difficult to obtain.

4) A pose estimation-based fall detector is built and compared with autoencoder-based architectures. This approach utilises extracted pose from key points of the body skeleton as input for activity recognition, eliminating background interference. The experimental results demonstrated better detection accuracy on elderly falls in real-world scenarios compared with autoencoders.

5) We developed two real-world datasets, specifically including examples of falls in older adults in real-world environments. This approach aims to enhance the realism and relevance of the data used for fall detection in training and evaluating AI models.

We conducted extensive evaluation on different real-world datasets, including two datasets are publicly available (UR Falls detection dataset [9] and Multiple Camera Fall Dataset [10]), and two datasets generated by ourselves. The inclusion of falls data from different environments, perspectives, and age groups, including the dataset created solely from falls of elderly people, allowed us to assess the generalisability and robustness of the models. The models we trained using our bespoke real-world dataset show promising results. In addition, we discuss the merits and limitations of the autoencoder and pose estimation-based approaches, suggesting viable future directions.

Methodology

Two deep learning-based fall detection approaches are investigated, namely, the autoencoder-based and the pose estimation-based fall detectors. In particular, four autoencoder based models adapting from Deep AutoEncoder (DAE) [11], Convolutional AutoEncoder (CAE) [12], Convolutional LSTM AutoEncoder (ConvLSTM-AE) [13], Deep Spatio-Temporal Convolutional AutoEncoder (DSTCAE) [14], and one pose-based model formed by Tiny-YOLO [15], AlphaPose [16] and Spatial Temporal Graph Convolutional Networks (ST-GCN) [17]. Autoencoder detectors are trained with normal Activities of Daily Living (ADL) and analyse the reconstruction error to detect anomaly falls. For the pose-based falls detector, the efficient Yolo model is utilised for human detection, the well-established AlphaPose is used for skeleton recognition, and a SpatioTemporal Graph Convolutional Neural network incorporating both spatial information from the skeleton and temporal movement features across frames is used for activity recognition.

Fall detection models

As shown in figure 1, an autoencoder consists of an encoder and decoder, each comprising multilayer Convolutional Neural Networks (CNNs). The encoder maps image input to a lower dimensional latent space (bottleneck), capturing salient underlying visual features. The decoder takes this compressed representation and reconstructs the original input. The autoencoder is trained to minimise reconstruction error between input image and reconstructed image, thus ensuring the output faithfully approximates the input. The model training is achieved in a manner of self-supervised learning without human labelling.

The encoder maps the input $\mathbf{x}$ into the code as a compressed latent representation. The encoder consists of multiple layers (e.g. CNNs), performing a nonlinear transformation δ. These successive layers systematically reduce the dimensionality of the input. We define this encoding process in Eq.1, h is the latent variables in the information bottleneck, $\mathbf{W}$ and $\mathbf{b}$ represent encoder weights and bias, respectively.

$$h = \delta\left(Wx + b\right) \qquad (1)$$

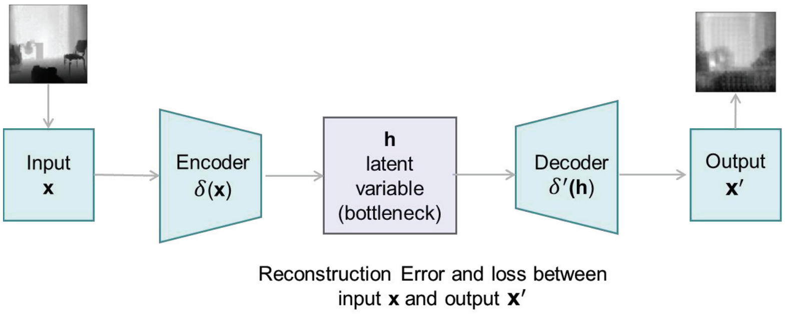

Figure 1 Autoencoder-based image reconstruction for anomaly detection.

The decoder maps the latent variables **h** to a reconstruction x' of the input. The decoder is usually the mirror opposite δ' of the encoder. Latent variables obtained from the encoder are re-transformed back to form a reconstructed output. The decoder is defined in Eq.2 with W' and b' representing the learnt weights and bias of the decoder, respectively.

$$x' = \delta'\left(W'h + b'\right) \tag{2}$$

The goal of the autoencoder is to optimize the difference (namely reconstruction error) between the original input and the reconstructed output using a loss function. The mean-square-error loss is defined in Eq. 3. When performing an anomaly detection task, normal behaviour can be faithfully reproduced by the model. Whereas fluctuations in reconstruction error above a threshold detects an anomalous situation (e.g. a fall).

$$\mathcal{L}\left(x,x'\right) = x - x'^{2} = x - \delta'\left(W'(\delta(Wx+b)) + b'\right)^{2} \tag{3}$$

Model 1: Deep AutoEncoder based on the sparsity (DAE)

Figure 2 shows the pipeline of the DAE model [11] for falls detection. Normal ADL videos are used for model training. The input video images are split into cubes of 10 x10 pixel patches over 5 successive frames. The Sparsity Value (SV) of each 10x10x5 block is calculated based on the similarity of pixels between the 5 patches, representing the proportion of zero elements. Thus a low sparsity hints that pixel information is changing among frames, thus containing more useful information.

When the sparsity value is less than a defined threshold τ^{SV}, the block is considered to be a key block. Consequently, nine 30 × 30 × 10 blocks around the key block are extracted for model learning and computation of Reconstruction Error (RE). The input is considered anomalous only if the RE value is greater than a specific threshold τ^{RE}. This process of finding key points saves time cost for training.

Model 2: Convolutional AutoEncoder (CAE)

Another issue that affects the performance of autoencoder-based anomaly detection methods is that "meaning" is not clearly defined during the learning process of normal behaviours. Background scenes can be highly diverse and chaotic, making it difficult to extract anomalous visual behaviour features from videos. The CAE model [12] addresses this by learning temporal regularities and detecting objects associated with irregular motion, incorporating past and future

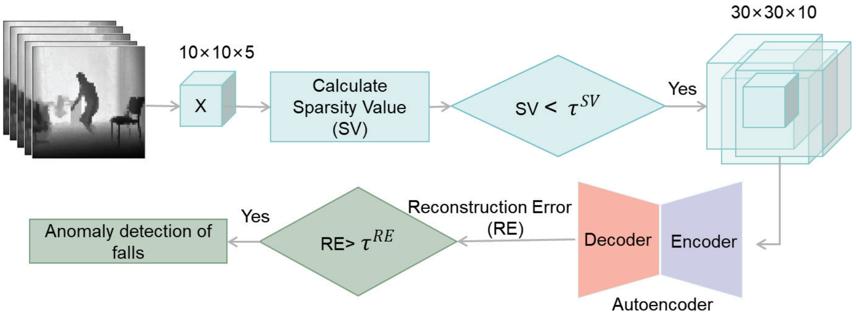

Figure 2 Pipeline of Model 1-DAE.

frames. Model 2-CAE learns motion features end-to-end by stacking sliding windows as a model input. The encoder consists of three convolutional layers and two pooling layers. The decoder consists of three inverse convolutional layers and two inverse pooling (upsampling) layers.

Model 3: Convolutional LSTM AutoEncoder (ConvLSTM-AE)

Long Short-Term Memory (LSTM) structure is employed to learn the temporal evolution of spatial features, extracted from several convolutional layers. This hybrid spatiotemporal architecture of LSTM and CNN is beneficial for efficient anomaly detection in time series data [13].

Model 4: Deep Spatio-Temporal Convolutional AutoEncoder (DSTCAE)

A deep spatio-temporal convolutional autoencoder employing 3D CNNs is used to learn spatial and temporal features [14]. A new anomaly scoring method that combines the reconstruction score of frames across a temporal window is also used to detect falls.

Model 5: Pose-based fall detector

Pose-based fall detection involves a combination of human detection, pose estimation and activity recognition, as summarised in figure 3. To achieve this, Tiny-YOLO [15] is used to detect the human body bounding box in images. The state-of-the-art AlphaPose [16] is then employed to calculate the location the key points (joints) of the human body, thus determining the spatial configurations of body poses. Finally, the Spatio-Temporal Graph Convolutional Network ST-GCN [17] combines the body position, joints and pose structure information to achieve action recognition. The ST-GCN calculates the probability that a frame fits each of 7 activities (standing, walking, sitting, lying down, standing up, sitting down, fall down), with the highest probability activity as its output.

Fall datasets

Four main datasets were used to train and evaluate the above five falls detection techniques, as shown in table 1. Two datasets are publicly available (UR Falls detection dataset [9] and Multiple Camera Fall Dataset [10]). A downside of these datasets is that they remain in the same scenario of office and imitate the fall of middle-aged people. To increase the real-world applicability of the models, we created two additional datasets, the Self-recorded Dataset and YouTube collected Dataset, to consider different age groups, environments, camera angles and number of people.

Office-1 - UR Fall Detection Dataset: The UR Falls dataset comprises 40 videos of middle-aged adults simulating ADLs (e.g. walking, sitting, squatting, standing), and an additional 30 videos of middle-aged adults simulating falls. The camera position and room remain the same in each video, remaining at the level of a person (i.e. in-plane view). Only one person ever appears in each video.

Office-2 - Multiple Cameras Fall Dataset: In the

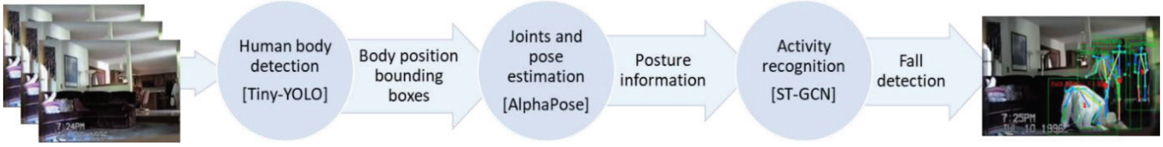
Human body detection [Tiny-YOLO]
Body position bounding boxes
Joints and pose estimation [AlphaPose]
Posture information
Activity recognition [ST-GCN]
Fall detection

Figure 3 Pipeline for the pose-based fall detector.

Table 1: Summary of characteristics of falls datasets.

	Dataset	Video number	In-plane view	CCTV-like view	Multiple people	Multiple environments	Video quality	Fall behavior
1	UR Fall Detection Dataset(Office-1)[9]	40 ADLs 30 Falls	Y				Same	Imitation
2	Multiple Cameras Fall Dataset(Office-2)[10]	176 ADLs 16 Falls		Y	Y		Same	Imitation
3	Self-recorded Dataset(Dormitory)	10 Falls	Y			Y	Same	Imitation
4	YouTube collected Dataset(Elderly)	9 Falls	Y	Y	Y	Y	Varied	Real falls in elderly

Multiple Camera Falls dataset, there are 176 videos containing varying numbers of people, simulating ADLs by middle-aged adults. An additional 16 videos are of middle-aged adults simulating falls. The camera is at an overhead angle, akin to CCTV, the recordings are of a single office room from 8 camera angles.

Dormitory - Self-recorded Dataset: We recorded this dataset in varying environments (e.g. different corridors and rooms in the dormitory) and shows a younger person simulating falls in 10 videos.

Elderly - YouTube collected Dataset: The second new dataset is the Elderly Dataset, which is comprised of videos sourced from YouTube of real recorded falls in elderly people. This dataset is the most realistic, as it shows real falls (all 3 other datasets have simulated falls), a range of environments, and camera angles, multiple people may be in the frame (or only one). The video quality is also varied between sequences, reflecting real-world differences in video recordings.

In the UR Fall Detection Dataset, fall events are already labelled with the exact time and frame at which the fall events start and end. Thus for the other three datasets, the timing of the start and end of fall events was added through manual labelling.

Evaluation matrix

The start time and end time for fall events, either pre-included in public datasets or added by us, were used as ground truth, indicating each frame's label as fall or not fall. The autoencoder-based fall detectors determine falls if the anomaly value exceeds a certain threshold. Therefore we can obtain the Receiver Operating Characteristic (ROC) curve of false positive vs. true positive at varying anomaly threshold values. The Area Under Curve (AUC) for ROC is therefore obtained as an accuracy value for each model which can be compared.

Results and Discussion

For the four autoencoder-based models, we used the 40 ADLs in the Office-1 dataset for model training (20 epochs). The four generated fall detectors are tested on the 30 fall videos from the Office-1 Dataset, with table 2 showing the comparative results. The best performing model was model 4 – DSTCAE using deep 3D CNNs, achieving 83% AUC accuracy.

Table 2: Area under the ROC Curve (AUC) comparison of the four autoencoder-based models training on ADLs and testing on falls in the Office-1 dataset.

Model	AUC
Model 1: DAE	62%
Model 2: CAE	71%
Model 3: ConvLSTM-AE	80%
Model 4: DSTCAE	83%

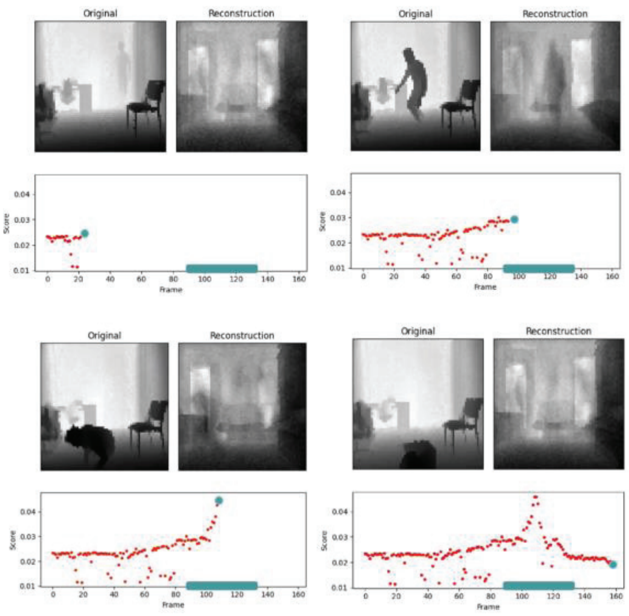

Figure 4 Successful fall detection using Model 1-DAE tested on a fall video sequence in Office-1. Four frames in the video are shown. For each frame, top left is the original image, top right is the reconstructed image, bottom is the anomaly score. The green dot indicates the time of the corresponding frame, and the green bar denotes the ground truth of fall period.

Figure 4 shows an example of successful fall detection by Model 1: DAE on a fall video from Office-1 dataset. The video starts with an empty room, into which a person enters and subsequently has a fall, remaining on the ground after the fall. We observe in the first frame that the background of the reconstructed image is not the same as the original, explaining the consistently low level of anomaly value throughout the video. When the person in the video performs some normal ADLs, the anomaly score remains steady with minimal fluctuations. When the falls event occurs, the anomaly value starts to fluctuate and reaches a high peak value. The anomaly value then falls back to its baseline once the fall event is over. This highlights the temporal relationship that is being used by the model to detect falls events, beyond just analysing the spatial appearance of each individual frame on its own.

Model 2 also performs well when tested on the Office-1 falls dataset. However, when we evaluate the generalisation of Model 1 and Model 2 (trained on Office-1 dataset) using the quite different Dormitory dataset, they both perform poorly on the Dormitory dataset. An example is shown in figure 5, where Model 1 fails to predict the fall, and Model 2 showed incorrect false peaking after predicting a fall (Figure 6).

This can be explained by anomaly detection relying heavily on the model's video reconstruction capability rather than the recognition of falls. Figure 7 shows Model2-CAE reconstructing the video of the Office-1 dataset much better for the background than Model1-DAE, resulting in lower absolute anomaly values. Of note, when the Dormitory dataset is used for testing, it has a different environment from the training set. Model-1-DAE still reconstructs the background similar to the training set, which directly leads to the fact that the model does not detect the fall, but rather considers the background to be more similar to the original video when the people fall and the lowest anomaly score peak appears.

We then tested the generalisability of Model 3-ConvLSTM-AE and Model4-DSTCAE (trained on Office-1 ADLs) on all four datasets, as shown in table 3. Both models achieved an average AUC accuracy of around 80% on the two office datasets. An example of multi-falls detection by Model 3 on Office-2 is shown in figure 6A. In addition, the two models also perform well on the Dormitory dataset, reaching respectively 71% and 76%, and a representative example is shown in figure 6C.

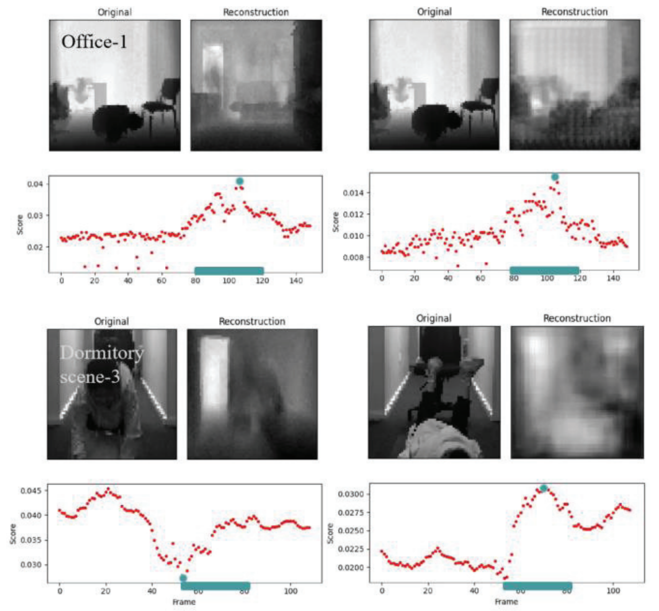

Figure 5 Top: Model 1 (left) and Model 2 (right) tested on Office-1, both good performance.

Bottom: Model 1 (left) and Model 2 (right) tested on Dormitory, showing degraded performance.

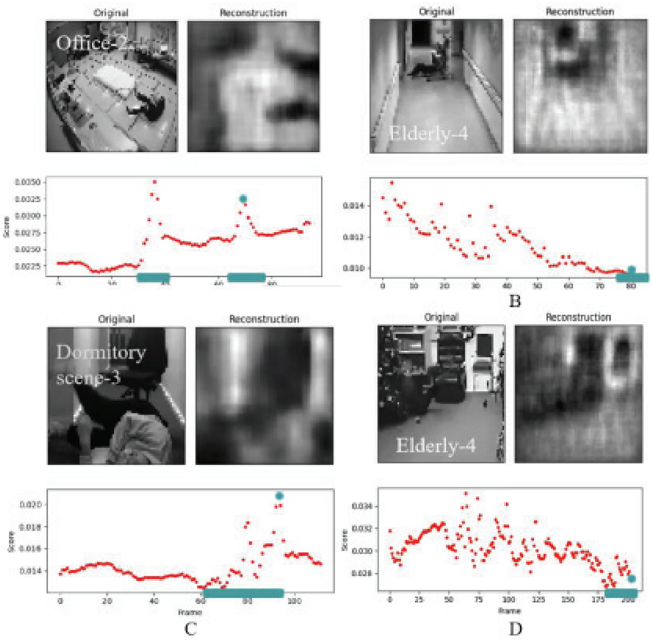

Figure 6 A) Model 3-ConvLSTM-AE on Office-2 showing good multi-fall detection.

B) Model 3 failing to detect in Elderly dataset.

C) Model 4-DSTCAE successful detection in Dormitory example.

D) Model 4 failing to detect in Elderly.

Table 3: AUC comparison of Model3-ConvLSTM-AE and Model 4-DSTCAE training and testing on four datasets.

Training Dataset	Test Dataset	Model 3: ConvLSTM-AE (AUC)	Model 4: DSTCAE (AUC)
Office-1 40 ADLs	Office-1: 30 falls	80 %	83%
	Office-2: 16 falls	78%	80%
	Dormitory: 10 falls	71%	76%
	Elderly: 9 falls	57%	60%

However, Models 3 and 4 struggled when applied to the more realistic Elderly dataset, only achieving 57% and 60% respectively. To improve detection accuracy, we leverage transfer learning technology. Model 3-ConvLSTM-AE was firstly pre-trained on the Office-1 dataset of 40 ADLs, then it was continually trained and fine-tuned on the Office-2 dataset with an additional 10 ADL videos. The Office-2 dataset includes more examples of different scenarios, angles and numbers of people. Through this transfer learning, Model-3 gained an improved detection accuracy of 3% on the Elderly dataset. We would argue that transfer learning on larger representative datasets will further improve model performance significantly.

Figures 6B,D show where the Models 3 and 4 fail to detect in a not-so-complex real-life hospital and care home scenario respectively. This demonstrates a limitation of anomaly detection, where the model must learn as much normal appearances of both foreground human activity and background environment as possible, in order to distinguish anomalies. In actuality, model complexity is inherently constrained in order to achieve processing speeds required for real-time monitoring. Such constrained AI models would struggle to learn the diversity of activities, video angles, and background environments from varying video qualities.

For the above challenges, pose estimation is a potential solution. Model 5 Pose-based detector is generated using 5 fall videos from Office-1 dataset. First, Tiny-YOLO was applied to detect the human bounding box, shown as a green box in figure 7. Next, AlphaPose was used to detect skeleton and key points. Finally, the sequential skeleton data was used as input of ST-GCN and each frame was given a ground truth label for fall or not, through supervised learning.

The model calculates the percentage confidence of the pose fitting each of seven activities, and outputs the highest confidence activity as recognition result. Figure 7 shows some sample results across all 4 datasets, where the recognition result accompanied by percentage confidence is displayed next to the bounding box. We note that in figure 6D, we get the best detection results for the real-world elderly fall video, which fails in all anomaly detection approaches using autoencoder-based (Figure 6B).

Conclusion

We present five deep learning-based models to detect falls from video images. We trained the models and evaluated their performance over two publicly available datasets and two self-developed datasets. Our two new datasets increased the challenge by encompassing varying and complex backgrounds, camera angles and real-life falls in elderly people. Our results showed that while the autoencoder-based models may work on well-trained background scenarios, they generally performed worse on more dynamic and challenging backgrounds as presented in our two new datasets. Model 4 – DSTCAE which employs deep 3D CNNs outperformed the other three autoencoder-based models with an AUC accuracy of 60% on the most challenging Elderly dataset. The pose-based detector was able to detect falls when autoencoder-based methods failed.

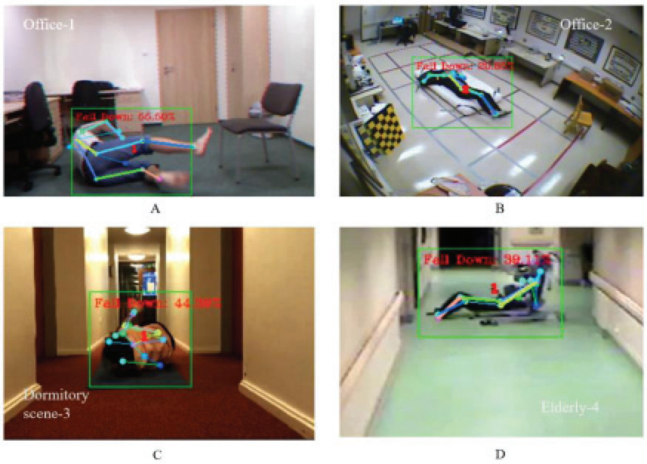

Figure 7 Examples of successful fall detection using Model-5 Pose-based detector on all four datasets.

We note that as autoencoder models are trained on ADLs, many "normal" behaviours not present in the training dataset could be recognised as anomalous. This major drawback inherently requires larger datasets to cover a greater range of "normal" human behaviours. In essence, it is defining falls by learning through "exclusion". They usually perform well in scenarios similar to training data, however they would struggle to reconstruct images (background and foreground) in transformed different scenarios. In contrast, a posture-based detector learns by "inclusion" - trained specifically on fall data to effectively recognise a fall. It is thus less likely to be confounded by previously unseen normal human behaviours. Thus they require less training data than autoencoder-based models while also potentially being more accurate. However, this comes with increased computational cost associated with pose estimation.

For future research and development of visual fall detection, efforts would focus on system robustness and real-time performance for complex real-world scenarios, e.g. occlusion, lighting and presence of multiple people, in both indoor and outdoor environments. For the deployment of a camera-based fall detection system, privacy should also be considered. To this end, thermal and infrared cameras could be a potential option. Edge computing and embedded systems, which allow the processing of data on board, will also reduce related security issues.

There is not yet a comprehensive benchmark dataset for a systematical evaluation of fall detection algorithms. This presents a hurdle in advancing the current research and technology for healthcare. Nearly all published datasets use middle-aged subjects to simulate falls, while there are significant differences in falls between the elderly, middle-aged subjects and children. Rather than merely focusing on elderly people, falls of children, "accidental" and "non-accidental" falls should be studied to cover a wide scope.

References

1. Moncada LVV, Mire LG. Preventing Falls in Older Persons. Am Fam Physician. 2017 Aug 15;96(4):240-247. PMID: 28925664.
2. Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, Aguilar-Navarro S, Alexander NB, Becker C, Blain H, Bourke R, Cameron ID, Camicioli R, Clemson L, Close J, Delbaere K, Duan L, Duque G, Dyer SM, Freiberger E, Ganz DA, Gómez F, Hausdorff JM, Hogan DB, Hunter SMW, Jauregui JR, Kamkar N, Kenny RA, Lamb SE, Latham NK, Lipsitz LA, Liu-Ambrose T, Logan P, Lord SR, Mallet L, Marsh D, Milisen K, Moctezuma-Gallegos R, Morris ME, Nieuwboer A, Perracini MR, Pieruccini-Faria F, Pighills A, Said C, Sejdic E, Sherrington C, Skelton DA, Dsouza S, Speechley M, Stark S, Todd C, Troen BR, van der Cammen T, Verghese J, Vlaeyen E, Watt JA, Masud T; Task Force on Global Guidelines for Falls in Older Adults. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing. 2022 Sep 2;51(9):afac205. doi: 10.1093/ageing/afac205. Erratum in: Age Ageing. 2023 Sep 1;52(9): Erratum in: Age Ageing. 2023 Oct 2;52(10): PMID: 36178003; PMCID: PMC9523684.
3. Kristoffersson A, Lindén M. A Systematic Review of Wearable Sensors for Monitoring Physical Activity. Sensors (Basel). 2022 Jan 12;22(2):573. doi: 10.3390/s22020573. PMID: 35062531; PMCID: PMC8778538.
4. Zhang S, Li Y, Zhang S, Shahabi F, Xia S, Deng Y, Alshurafa N. Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. Sensors (Basel). 2022 Feb 14;22(4):1476. doi: 10.3390/s22041476. PMID: 35214377; PMCID: PMC8879042.
5. Jobanputra C, Bavishi J, Doshi N. Human activity recognition: A survey. Procedia Comput Sci. 2019;155:698-703. doi: 10.1016/j.procs.2019.08.100.
6. Gupta N, Gupta SK, Pathak RK, Jain V, Rashidi P, Suri JS. Human activity recognition in artificial intelligence framework: a narrative review. Artif Intell Rev. 2022;55(6):4755-4808. doi: 10.1007/s10462-021-10116-x. Epub 2022 Jan 18. PMID: 35068651; PMCID: PMC8763438.
7. Bank D, Koenigstein N, Giryes R. Autoencoders. In: Machine learning for data science handbook: data mining and knowledge discovery handbook. 2023;353-374.
8. Zheng C, Wu W, Chen C, Yang T, Zhu S, Shen J, Kehtarnavaz N, Shah M. Deep learning-based human pose estimation: A survey. ACM Comput Surv. 2023;56(1):1-37. doi: 10.48550/arXiv.2012.13392.
9. Kwolek B, Kepski M. Human fall detection on embedded platform using depth maps and wireless accelerometer. Comput Methods Programs Biomed. 2014 Dec;117(3):489-501. doi: 10.1016/j.cmpb.2014.09.005. Epub 2014 Oct 2. PMID: 25308505.
10. Auvinet E, Rougier C, Meunier J, St-Arnaud A, Rousseau J. Multiple cameras fall dataset. DIRO-Université de Montréal, Tech. Rep. 2010;8:1350(24).
11. Sabokrou M, Fathy M, Hoseini M. Video anomaly detection and localization based on the sparsity and reconstruction error of auto-encoder. Electron Lett. 2016;52(13):1122-1124. doi: 10.1049/el.2016.0440.
12. Hasan M, Choi J, Neumann J, Roy-Chowdhury AK, Davis LS. Learning temporal regularity in video sequences. IEEE Conf. Computer Vision and Pattern Recognition. 2016;733-742. doi: 10.48550/arXiv.1604.04574.
13. Nogas J, Khan SS, Mihailidis A. Fall detection from thermal camera using convolutional LSTM autoencoder. Workshop on Aging, Rehabilitation and Independent Assisted Living, IJCAI Workshop. 2018.
14. Nogas J, Khan SS, Mihailidis A. DeepFall: Non-Invasive Fall Detection with Deep Spatio-Temporal Convolutional Autoencoders. J Healthc Inform Res. 2019 Dec 18;4(1):50-70. doi: 10.1007/s41666-019-00061-4. PMID: 35415435; PMCID: PMC8982799.
15. Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018 Apr 8. doi: 10.48550/arXiv.1804.02767.
16. Fang HS, Xie S, Tai YW, Lu C. Rmpe: Regional multi-person pose estimation. IEEE Int. Conf. Computer Vision. 2017;2334-2343.
17. Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. AAAI Conf. Artificial Intelligence. 2018;32(1). doi: 10.1609/aaai.v32i1.12328.