Research Article
Interdisciplinary Nursing Research
OPEN

A novel lightweight deep learning fall detection system based on global-local attention and channel feature augmentation

Yuyang Sha, PhD, Xiaobing Zhai, PhD, Junrong Li, MSc, Weiyu Meng, PhD, Henry H.Y. Tong, PhD, Kefeng Li, PhD*

Centre for Artificial Intelligence Driven Drug Discovery, Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, China

Abstract

Background and Objective: Reducing the number of falls in nursing facilities is crucial to prevent significant injury, increased costs, and emotional harm. However, current fall detection systems face a trade-off between accuracy and inference speed. This work aimed to develop a novel lightweight fall detection system that can achieve high accuracy and speed while reducing computational cost and model size.

Methods: We used convolutional neural networks and the channel-wise dropout and global-local attention module to train a lightweight fall detection model on over 10,000 human fall images from various scenarios. We also applied a channel-based feature augmentation module to enhance the robustness and stability of the model.

Results: The proposed model achieved a detection precision of 95.1%, a recall of 93.3%, and a mean average precision of 91.8%. It also had a significantly smaller size of 1.09 million model parameters and a lower computational cost of 0.12 gigaFLOPS than existing methods. It could handle up to 20 cameras, simultaneously with a speed higher than 30 fps.

Conclusion: The proposed lightweight model demonstrated excellent performance and practicality for fall detection in real-world settings, which could reduce the working pressure on medical staff and improve nursing efficiency.

Keywords: Nursing, Fall detection, Convolutional neural networks, Lightweight backbone, Attention, Feature augmentation

Introduction

Falls and related injuries remain a major public health problem all over the world, especially in developing countries[1,2]. Each year, over 68 million falls occur globally, and the number is expected to increase with the aging of the population[3,4]. An estimated 38% of these falls result in injuries, leading to emergency department visits, longer hospital stays, higher rates of transfer to long-term care facilities, and increased financial costs for patients, families, and health facilities[5]. Moreover, falls can have emotional and physical consequences for patients and their caregivers or family members, such as fear of falling, depression, pain, and reduced physical activity. These factors can impair the patient's recovery and independence[6].

Despite efforts to prevent falls, predicting and identifying patients at risk of falling is challenging for caregivers. Previous studies had developed various screening tools based on known risk factors to identify potential fall patients[7]. Several meta-analyses found that common preventative solutions such as chair alarms, bed alarms, and the use of scored risk assessment tools were not associated with significant fall reductions[8–10]. This highlights the need for further research into effective fall prevention strategies to improve patient outcomes and reduce health care costs associated with fall-related injuries.

The etiology of falls and fall-related injuries are multifactorial, with both internal and external factors playing a role[11]. External factors include the physical environment, nursing process, and nurse staffing. Internal factors are patient-specific risk factors such as fainting, weakness, and unsteady gait[12]. Even though patient and staff education can reduce hospital falls, the demand for nursing services is much higher than the hospitals can provide in the existing medical system, especially during the COVID-19 pandemic[13]. It is not practical to rely entirely on the medical staff for patient fall prevention and monitoring. Therefore, automated monitoring of patient conditions through artificial intelligence–related technology represents a promising solution.

With the continuous development of artificial intelligence[14], machine learning[15], and computer hardware, deep learning algorithms have become widely used in computer vision[16], natural language processing[17], speech recognition[18], and other tasks[19]. Compared with traditional methods, deep learning-based approaches may achieve better performance in terms of precision, robustness, and efficiency. In particular, convolutional neural networks (CNNs) algorithms have been successfully applied to image classification and recognition[20]. Unlike traditional methods that require manual feature extraction, CNNs can directly identify features through data[21] and learn highly

E-mail addresses: p2215539@mpu.edu.mo (Y. Sha), zhaixbwh@163.com (X. Zhai), p2213983@mpu.edu.mo (J. Li), mwy_9607@163.com (W. Meng), henrytong@mpu.edu.mo (H.H.Y. Tong), kefengl@mpu.edu.mo (K. Li)

**Corresponding author: Kefeng Li.*

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Interdisciplinary Nursing Research (2023) 2:68–75

Received 15 March 2023; Accepted 2 July 2023

Published online 10 July 2023

http://dx.doi.org/10.1097/NR9.0000000000000026

representative image features in a layered, data-driven manner from sufficient training data. To date, CNNs have been primarily utilized in medical image classification, with 3 main categories of application: using pretrained features from existing models[22], training CNNs from scratch[23], and unsupervised CNN pre-training and supervised fine-tuning for medical images[24].

The task of detecting patient falls can also be simplified as an object detection problem. Therefore, a CNN-based model can be used to analyze surveillance camera footage, identify patient movements, and issue alerts for special situations. Recently, a few falling detection systems based on CNNs and long short-term memory have been proposed[25–27]. However, current falling detection systems encounter several challenges that affect their performance, such as extreme poses, partial occlusion, data imbalance, and poor image quality[28]. Moreover, the bounding boxes generated by these systems often do not well with the actual falls, resulting in false positives in fall detection tasks[29]. Some of the issues related to power consumption, real-time operations, sensing limitations, privacy, and record of real-life falls are still open and require further research.

To address the aforementioned limitations, we presented a novel fall detection system in this paper. Our system incorporated a lightweight backbone, a novel attention module, and a channel-based feature augmentation module. To ensure robust model training, we generated and labeled a comprehensive dataset of over 10,000 fall images, covering individuals of all age groups. We then evaluated the proposed fall detection system using the generated dataset, comparing its performance, model parameters, computational cost, and running speed with common object detection algorithms.

Methods

Database construction

In this study, we collected a large number of high-definition image samples from open-source datasets and media such as Microsoft COCO (https://cocodataset.org/)[30], UR Fall Detection Dataset (URFD) (http://fenix.ur.edu.pl/~mkepski/ds/uf.html), UMAFall dataset (https://doi.org/10.6084/m9.figshare.4214283.v7)[31], and YouTube (www.youtube.com). The dataset comprised images of individuals across all age groups (children, adults, and the elderly), captured in both indoor and outdoor settings. The indoor scenes consisted of bedrooms, living rooms, kitchens, bathrooms, wards, corridors, staircases, elevators, and other common environments. The outdoor scenes include squares, lawns, parks, and other natural settings. The location of a human is typically represented by a bounding box[30]. We used the labelme tool for image annotation (https://github.com/wkentaro/labelme)[32]. During the analysis, we implemented 5-fold cross-validation to reduce the impact of partition randomness on the obtained results. In each fold, we randomly selected 80% of the data for training and 20% for testing.

Implementation details

We implemented the deep learning model in PyTorch and trained it on a single Nvidia A100 GPU. For model optimization, we used Adam with a weight decay of 10^{-6} and a momentum of 0.9. All input images were resized to 640×480 during the training stage.

Lightweight feature extraction model

We used depthwise separable convolution (ConvDW) to replace part of the standard convolution layer in our model structure. Moreover, we employed the 3×3 convolution layer with a stride of 2 for feature down-sampling, which improved training speed and stability. To solve the dead neurons and bias shift caused by ReLU during model training, we chose the Leaky ReLU as the activation function for the proposed lightweight model. The details of the parameters for the proposed backbone are shown in Table 1. ConvDW represents the depthwise separable convolution, and ConvDW-down represents feature down-sampling by convolution layer. The model structure is illustrated in Figure 1.

Global-Local attention module (GLAM)

We introduced an attention-based module for feature fusion, which includes a local attention module (LAM) and a global attention module (GAM).

The structure of the LAM is illustrated in Figure 2A. Given a local feature $F \in R^{C \times H \times W}$, 2 groups of new features U and V are generated with the same dimensions as the original input. Then we reshape the new features and map their dimension to $F \in R^{C \times N}$, where $N = H \times W$ represents the total number of features. Finally, a softmax operation is used to calculate the attention matrix with the results of matrix multiplication between U and V. LAM's attention matrix $S \in R^{N \times N}$ can be calculated as follows:

$$s_{ij} = \frac{exp(U_i \cdot V_j)}{\sum_{i=1}^{N} exp(U_i \cdot V_j)}, \quad (1)$$

where S_{ij} represents the impact of the ith position on the jth position. Simultaneously, we can get the feature matrix D and reshape them into$D \in R^{C \times N}$. Then, we employ the matrix multiplication between D and attention matrix S to get the final output $E_j^{LAM} \in R^{C \times H \times W}$, which is defined as follows:

$$E_j^{LAM} = \eta \sum_{i=1}^{N} \left(S_{ij} \cdot D_i\right) + F_i, \quad (2)$$

Table 1

The structure of our developed lightweight fall detection model.

Stage	Input	Operation	Repeat
1	H × W × C	Conv3x3	1
	H × W × 16	ConvDW	1
2	H × W × 16	ConvDW-Down	1
	$\frac{H}{2} \times \frac{W}{2} \times 32$	ConvDW	1
3	$\frac{H}{2} \times \frac{W}{2} \times 32$	ConvDW-Down	1
	$\frac{H}{4} \times \frac{W}{4} \times 64$	ConvDW	3
4	$\frac{H}{4} \times \frac{W}{4} \times 64$	ConvDW-Down	1
	$\frac{H}{8} \times \frac{W}{8} \times 128$	ConvDW	2
5	$\frac{H}{8} \times \frac{W}{8} \times 128$	ConvDW-Down	1
	$\frac{H}{16} \times \frac{W}{16} \times 256$	ConvDW	1
6	8 × 8	AvgPool	—
	1 × 1	FC	—

C indicates channel; FC, fully connected layer; H, height; W, width.

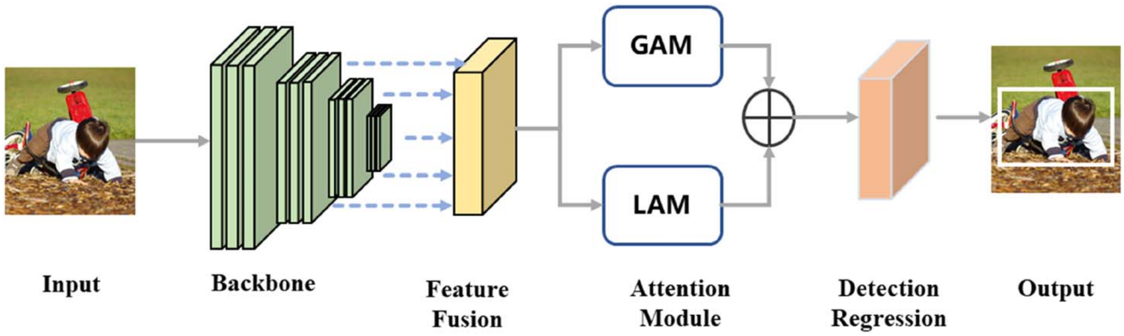

Figure 1. Flowchart of the algorithm. GAM indicates global attention module; LAM, local attention module.

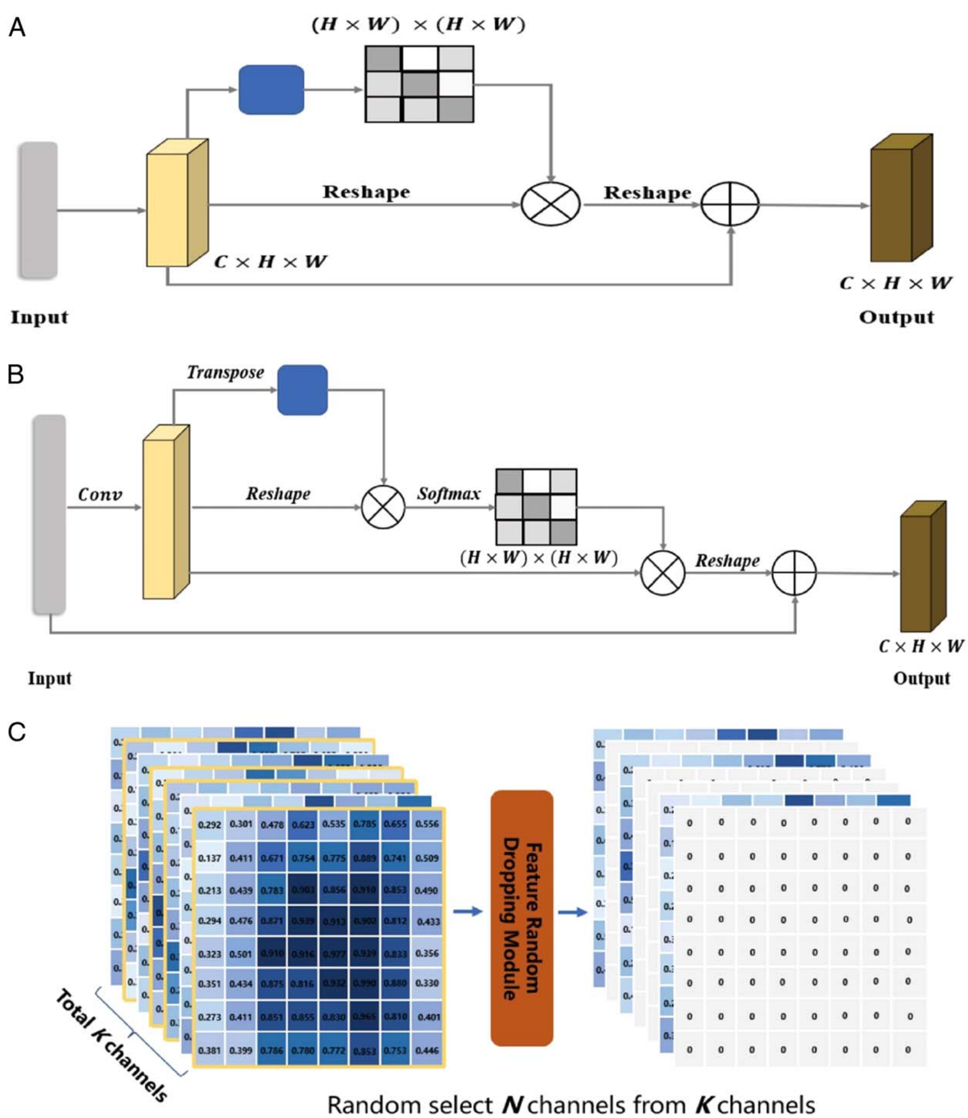

Figure 2. The schematic diagram of global-local attention module and feature augmentation module (Channel). A, Local attention module. B, Global attention module. C, Channel. C indicates channel; H, height; W, width.

Table 2

The performance of the developed fall detection system compared with other models.

Model	Backbone	Precision (%)	Recall (%)	mAP (%)	Speed (ms)
SSD	MobileNet V2	85.2	84.6	87.3	5.1
SSD	VGG16	89.4	87.5	89.6	9.7
YOLOv3	ShuffleNet V2	86.4	87.3	88.2	4.8
YOLOv3	DarkNet-19	95.4	94.9	90.6	8.6
YOLOX	CSPDarkNet	96.7	94.6	92.3	9.2
Faster RCNN	ResNet-50	96.2	92.2	93.1	32.1
Our model	—	95.1	93.3	91.8	1.8

mAP indicates mean average precision.

where η is a constant factor that balances the weight between the local detail feature and the original input feature. E_j^{LAM} contains local details and global semantic information, which is beneficial in improving the performance significantly for small objects and unconstrained environments.

The structure of GAM is illustrated in Figure 2B. Different from LAM, the GAM directly uses the features $F \in R^{C\times H\times W}$ output from deep neural networks to calculate the attention matrix $M \in R^{C\times C}$. We reshape the feature into $F \in R^{C\times N}$. Next, we calculated the global attention map between F and M. Finally, the attention map was then normalized using the softmax operation. Specifically, the attention map was defined as follows:

$$m_{ij} = \frac{exp(F_i \cdot F_j)}{\sum_{i=1}^{C} exp(F_i \cdot F_j)}, \quad (3)$$

where m_{ij} represents the impact of the ith channel on the jth channel. In GAM, the attention matrix needs to be multiplied with the original input feature to obtain the relation matrix between other channels. Finally, the relation matrix was added with the original input feature F to obtain the final output $E_j^{GAM} \in R^{C\times H\times W}$:

$$E_j^{GAM} = \alpha \sum_{i=1}^{C} \left(m_{ij} \cdot F_i\right) + F_i, \quad (4)$$

where α is a learnable parameter. The Equation 4 exhibits the final output is weighted sum of all channels and original input

Figure 3. The accurate differentiation of different types of movement using our developed fall detection system. Note: 0 with blue frame indicated normal stage and 1 with red frame indicated falling stage.

Table 3

A developed lightweight convolutional neural network model for fall detection.

Backbone	Parameters (million)	GFLOPS
MobileNet V2	3.43	0.31
ShuffleNet V2	3.51	0.28
VGG16	20.01	15.51
DarkNet-19	20.02	7.39
Our model	1.09	0.12

GFLOPS indicates gigaFLOPS.

features. The global attention matrix contains long-range relationship between feature maps, which can further enhance the performance of the model.

Feature augmentation module

We developed a feature augmentation module using channel-wise dropout to improve the ability of the backbone to extract human features and their corresponding dependencies. The schematic diagram of the proposed module is shown in Figure 2C. This module can be embedded into the existing deep learning model so that the model can explore the dependency relationship between different parts of the human body during training[33,34]. For a given set of feature maps (F = {f_1, f_2, …, f_n}), where f_i represents the *i*th layer in the feature set. The designed module will randomly select the *K* layer and then set all parameters to zero. The subsequent deep neural network will then recover the related information based on the association between feature maps. This learning process can aid the model in exploring the long-range dependence and geometric relationship between different parts of the human body.

Results

Comparative analysis of the performance of the developed fall detection system

In this paper, we evaluated the performance of the proposed fall detection system by comparing it with common object detection algorithms, such as SSD[35], YOLOv3[36], YOLOX[37], and Faster RCNN[38]. Following previous studies[35,37], we trained these methods using our collected dataset of over 10,000 fall images and evaluated their performance using mean average precision (mAP), precision, and recall. The developed model achieved a precision of 95.1%, recall of 93.3%, mAP of 91.8, and the inference speed of 1.8 ms (Table 2). Taking precision, recall, mAP, and speed into consideration, our proposed fall detection system outperformed the existing object detection algorithms. Although our detection precision was slightly lower than that of YOLOX (95.1% for our model versus 96.7% for YOLOX), our method achieved significantly faster inference speed than YOLOX (1.8 ms for our model and 9.2 ms for YOLX), making it suitable for real-time fall detection tasks. Faster RCNN showed slightly better precision (96.2% vs. 95.1%) and mAP (91.8% vs. 93.1%) compared with our proposed system (Table 2). Figure 3 displays the examples for detecting different types of movement using our developed fall detection system.

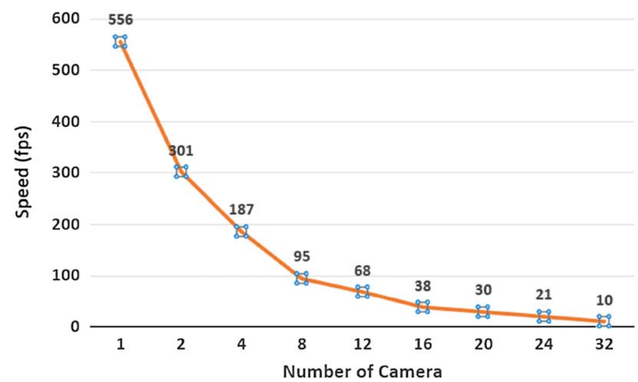

Figure 4. The correlations between the algorithm speed of the developed fall detection system and the number of cameras. The general acceptable speed is > 30 fps with as many as the cameras.

Model efficiency analysis

To be applicable in real-world scenarios, the proposed fall detection system must be capable of simultaneously analyzing multiple videos, necessitating a high level of efficiency and accuracy from the model. To evaluate the system's performance, we conducted a comparative analysis of its model parameters and computational cost against those of existing methods. As shown in Table 3, our lightweight CNN model for fall detection had only 1.09 million parameters and 0.12 gigaFLOPS, which were significantly smaller than those of current methods. These results indicate the potential of our model as a practical solution for fall detection in real-world settings.

Data parallel analysis

The proposed patient fall system is based on a deep learning model and requires simultaneous processing and analysis of video information from multiple cameras. To achieve real-time performance, the algorithm needs to operate at a speed greater than 30 fps. We evaluated the algorithm's scalability by varying the number of cameras and measuring the running speed. We used TensorRT to quantify and optimize the model and deployed it on an Nvidia GTX 2080 GPU[39]. As shown in Figure 4, the running speed of the algorithm decreased as the number of cameras increased. The algorithm can handle up to 20 cameras with a speed higher than 30 fps, which meets real-time processing requirements[40]. For larger-scale scenarios, we can consider using more powerful GPU platforms such as Nvidia V100 or A100, or applying more effective pruning or quantization techniques.

Table 4

Effects of LAM and GAM.

Module			
Baseline	LAM	GAM	mAP (%)
✓			85.6
✓	✓		87.3
✓		✓	89.2
✓	✓	✓	91.8

GAM indicates global attention module; LAM, local attention module; mAP, mean average precision.

Interaction analysis between modules

The patient fall system proposed in this paper is based on a deep learning model and consists of several modules. To demonstrate the efficacy of these modules, we conducted a series of ablation studies to investigate their individual contributions. The experimental results are presented in Table 4, which indicate that each proposed module can boost the model performance in terms of accuracy and robustness.

Discussion

Patient falls pose a serious safety challenge in health care settings, particularly for vulnerable populations such as older adults and children. Despite the implementation of various fall prevention strategies, their effectiveness remains limited and resource-intensive[41]. There is an urgent need for more interdisciplinary approaches to address the complex and multifaceted nature of fall prevention. In this study, we developed a novel fall detection system by the combination of a lightweight feature extraction model, GLAM, and channel-based feature augmentation. Our fall detection system outperformed most of the existing object detection algorithms in terms of precision, speed, model parameters, and computational cost.

In this study, we proposed a lightweight feature extraction network based on ConvDW-down, a modified convolution layer inspired by classical deep learning models such as ResNet[42], DarkNet[36], and MobileNet[43]. Our network aimed to improve the performance of YOLOv3[36], a common object detection algorithm that can handle various tasks, for the specific application of patient fall detection. YOLOv3 uses DarkNet-19 as its backbone, which consists of 19 convolutional layers and 5 pooling layers. However, DarkNet-19 has a large number of parameters and requires significant computational resources, which makes it impractical for fall detection. Compared with our proposed method, Faster RCNN offers better model accuracy but falls in real-time performance. Since fall detection has high requirements on model inference speed, it is hard for Faster RCNN to apply in this task. While human pose detection is a more complex task that requires identifying semantic points in human body images, compared with human detection, it often requires more powerful backbones to learn feature representations. This can result in higher computational costs and slower performance on embedded devices. In contrast, fall detection is a relatively simple task that can be handled with lightweight object detection methods and does not require complex algorithms such as human pose detection or semantic segmentation.

The object detection algorithm can struggle with complex situations such as extreme postures and partial occlusion due to the loss of information during feature extraction in deep neural networks[44]. To address this issue, in this study, LAM was introduced to incorporate contextual information into local features and enhance the model's representation capability[45]. Combined with the feature pyramid network[45,46], LAM can significantly improve the detection ability of the proposed fall detection system in unconstrained environments. In addition, each channel in the feature map generated by the deep neural network can be regarded as the response to a specific region in the input sample, and there is often a correlation between different channels, which is referred to as global information. Therefore, we designed a global attention mechanism to explore the relationship between different channels, which can enhance the model to capture global dependence explicitly. In addition, we conducted extensive experiments to evaluate the performance of each module in this paper. The results illustrate that both the feature augmentation and attention modules are essential for improving detection accuracy. In particular, the feature augmentation module has a more considerable impact on the model's performance and is only applied during training. Consequently, it does not increase the number of parameters or computational costs during deployment.

In the task of fall detection, there are also challenging scenarios that pose difficulties for deep learning-based models, such as large variations in pose, expression, shape, and extreme lighting conditions[47]. These challenges can significantly reduce the accuracy of detection and thus impact the effectiveness of medical staff in responding to falls. Compared with common object detection tasks, fall detection in patients can leverage certain prior knowledge, such as the symmetry of the human body and the large range of the target. Channel-wise dropout is a technique that exploits the characteristics of CNN[48]. In this study, we investigated the distinctive features of fall detection and applied channel-wise dropout to enhance the performance of our proposed system. This technique enables the model to infer the occluded parts of the human body from the visible information and increase the detection accuracy. For instance, when a person falls and some body parts are hidden from the camera, the model can still recognize the fall event based on the remaining information. However, the lack of depth information can limit the accuracy of fall detection using a monocular camera. Due to practical constraints, obtaining direct-depth information is not always feasible. Therefore, we addressed this issue by enriching our training dataset with a new class of annotated samples of humans lying in bed or on a sofa. This approach helped our model learn to distinguish between lying down and falling actions, thereby reducing false detections. Moreover, post-processing operations were applied to the detection results to improve the model's accuracy, and we implemented a threshold filter to avoid false detections. In practical use, we integrated multiobject tracking into our fall detection system to leverage information between video frames, further improving accuracy.

The limited size of the fall dataset may potentially result in overfitting when employing deep learning techniques for fall detection. To address this issue and ensure the robustness of the model training, we had generated and labeled a comprehensive dataset of over 10,000 fall images that had been collected from open sources. Notably, these images encompass individuals of all age groups, including children and elderly individuals. In comparison to previously reported fall detection systems[49,50], this dataset represented one of the most extensive collections of fall images that have been utilized for classifier training to our knowledge.

In our project, we proposed a fall detection system based on object detection algorithms. While classification-based methods are another potential solution, they may only be effective in simple scenarios, such as when an image contains only one person or a single movement. In contrast, if the input sample contains complex movements or multiple objects, the performance of a classification-based method may dramatically decrease[51].

Limitations

Although our proposed fall detection system demonstrates promising results, there are several challenges that need to be addressed to enhance its practicality and performance. One of the challenges is accurate identifying small and occluded regions of the human body. In addition, the potential gap between the training dataset and real-world scenarios could lead to decreased performance. To overcome these challenges, future research could explore the use of Transformer models, contrastive learning, and vision-language pretrained models. Integrating depth information could also improve model performance, particularly in complicated cases such as humans lying in bed. In the future, we plan to explore more advanced methods to obtain an accurate depth map from the RGB image.

Conclusions

In this work, we introduced a novel patient fall detection system based on CNNs. Our system employed a single-stage object detection algorithm with a lightweight feature extraction model. We also integrated GLAM and channel-based feature augmentation modules to enhance its accuracy and robustness. We evaluated our system on a collected dataset of over 10,000 human fall images and showed that it surpassed the existing models in performance while being computationally efficient. Our system has the potential to be a valuable tool in clinical settings for patient fall detection.

Sources of funding

This study was supported by a fund from Macao Polytechnic University (RP/FCSD-02/2022). The sponsor had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Conflicts of interest disclosure

The authors declare that they have no financial conflict of interest with regard to the content of this report.

References

1 Moreland B, Kakara R, Henry A. Trends in nonfatal falls and fall-related injuries among adults aged ≥65 years—United States, 2012-2018. MMWR Morb Mortal Wkly Rep 2020;69(27):875–81.
2 Zhang H, Li K, Zhao Y, *et al*. Long-term use of fluoxetine accelerates bone loss through the disruption of sphingolipids metabolism in bone marrow adipose tissue. Transl Psychiatry 2020;10(1):138.
3 Li F, Harmer P, Fitzgerald K, *et al*. Effectiveness of a therapeutic tai ji quan intervention vs a multimodal exercise intervention to prevent falls among older adults at high risk of falling: a randomized clinical trial. JAMA Intern Med 2018;178(10):1301–10.
4 King B, Pecanac K, Krupp A, *et al*. Impact of fall prevention on nurses and care of fall risk patients. Gerontologist 2018;58(2):331–40.
5 Miake-Lye IM, Hempel S, Ganz DA, *et al*. Inpatient fall prevention programs as a patient safety strategy: a systematic review. Ann Intern Med 2013;158(pt 2):390–6.
6 El Sayed A, Said MT, Mohsen O, *et al*. Falls and associated risk factors in a sample of old age population in Egyptian community. Front Public Health 2023;11:1068314.
7 Ek S, Rizzuto D, Xu W, *et al*. Predictors for functional decline after an injurious fall: a population-based cohort study. Aging Clin Exp Res 2021;33(8):2183–90.
8 Oliver D, Hopper A, Seed P. Do hospital fall prevention programs work? A systematic review. J Am Geriatr Soc 2000;48(12):1679–89.
9 Morris ME, Webster K, Jones C, *et al*. Interventions to reduce falls in hospitals: a systematic review and meta-analysis. Age Ageing 2022;51(5): afac077.
10 Tricco AC, Thomas SM, Veroniki AA, *et al*. Comparisons of interventions for preventing falls in older adults: a systematic review and meta-analysis. JAMA 2017;318(17):1687–99.
11 Heng H, Jazayeri D, Shaw L, *et al*. Hospital falls prevention with patient education: a scoping review. BMC Geriatr 2020;20(1):140.
12 Zhao YL, Bott M, He J, *et al*. Evidence on fall and injurious fall prevention interventions in acute care hospitals. J Nurs Adm 2019;49(2):86–92.
13 Li J, Yang Z, Qiu H, *et al*. Anxiety and depression among general population in China at the peak of the COVID-19 epidemic. World Psychiatry 2020;19(2):249–50.
14 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
15 Greener JG, Kandathil SM, Moffat L, *et al*. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022;23(1):40–55.
16 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ICLR; 2015. 1–14.
17 Devlin J, Chang M-W, Lee K, *et al*. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv; 2018.
18 Sainath TN, Weiss RJ, Wilson KW, *et al*. Multichannel signal processing with deep neural networks for automatic speech recognition. IEEE/ACM Trans 2017;25(5):965–79.
19 Cui S, Li L, Zhang Y, *et al*. Machine learning identifies metabolic signatures that predict the risk of recurrent angina in remitted patients after percutaneous coronary intervention: a multicenter prospective cohort study. Adv Sci 2021;8(10):2003893.
20 Chen B, Chen C, Hu J, *et al*. Computer vision and machine learning-based gait pattern recognition for flat fall prediction. Sensors (Basel) 2022;22 (20):7960.
21 Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59.
22 Ginneken Bv, Setio AAA, Jacobs C, *et al*. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. presented at: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI); 2015.
23 Menze BH, Jakab A, Bauer S, *et al*. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 2015;34 (10):1993–2024.
24 Li R, Zhang W, Suk H-I, *et al*. Deep learning based imaging data completion for improved brain disease diagnosis. presented at: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014: 17th International Conference; 2014.
25 Maitre J, Bouchard K, Gaboury S. Fall detection with UWB radars and CNN-LSTM architecture. IEEE J Biomed Health Inform 2021;25(4): 1273–83.
26 Tian Z, Zhang L, Wang G, *et al*. An RGB camera-based fall detection algorithm in complex home environments. Interdiscip Nurs Res 2022;1 (1):14–26.
27 Alanazi T, Muhammad G. Human fall detection using 3D multi-stream convolutional neural networks with fusion. Diagnostics (Basel) 2022;12 (12):3060.
28 Usmani S, Saboor A, Haris M, *et al*. Latest research trends in fall detection and prevention using machine learning: a systematic review. Sensors (Basel) 2021;21(15):5134.
29 Igual R, Medrano C, Plaza I. Challenges, issues and trends in fall detection systems. Biomed Eng Online 2013;12:66.
30 Lin T-Y, Maire M, Belongie S, *et al*. Microsoft COCO: Common Objects in Context. Computer Vision—ECCV 2014; 2014:740–55.
31 Casilari E, Santoyo-Ramón JA, Cano-García JM. UMAFall: a multisensor dataset for the research on automatic fall detection. Proc Comput Sci 2017;110:32–9.
32 Russell BC, Torralba A, Murphy KP, *et al*. LabelMe: a database and web-based tool for image. Int J Comput Vision 2005;77(1):1–10.
33 Huang G, Sun Y, Liu Z, *et al*. Deep networks with stochastic depth. Computer Vision—ECCV 2016. 2016:646–61.
34 Ghiasi G, Lin T-Y, Le QV. Dropblock: a regularization method for convolutional networks. NeurIPS; 2018: 1–11.

35 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. In: *Computer Vision–ECCV 2016: 14th European Conference*; 2016: 21–37.
36 Redmon J, Farhadi A. Yolov3: an incremental improvement. arXiv; 2018.
37 Ge Z, Liu S, Wang F, *et al*. Yolox: exceeding yolo series in 2021. arXiv; 2021.
38 Ren S, He K, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39(6):1137–49.
39 Pacal I, Karaboga D. A robust real-time deep learning based automatic polyp detection system. Comput Biol Med 2021;134:104519.
40 Huang J, Rathod V, Sun C, *et al*. Speed/accuracy trade-offs for modern convolutional object detectors. CVPR; 2017:7310–11.
41 Turner K, Staggs VS, Potter C, *et al*. Fall prevention practices and implementation strategies: examining consistency across hospital units. J Patient Saf 2022;18(1):e236–42.
42 He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. CVPR; 2016:770-8.
43 Sandler M, Howard A, Zhu M, *et al*. Mobilenetv2: inverted residuals and linear bottlenecks. CVPR; 2018:4510-20.
44 Zhang T, Wang J, Liu P, *et al*. Fall detection by embedding an accelerometer in cellphone and using KFD algorithm. Int J Comput Netw Secur 2006;6(10):277–84.
45 Hu H, Zhang Z, Xie Z, *et al*. Local relation networks for image recognition. Proc IEEE Comput Soc Conf Comput Vis 2019;26:3464–73.
46 Lin T-Y, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. CVPR; 2017:2117-25.
47 Alam E, Sufian A, Dutta P, *et al*. Vision-based human fall detection systems using deep learning: a review. Comput Biol Med 2022;146:105626.
48 He M, Zhang J, Shan S, *et al*. Locality-aware channel-wise dropout for occluded face recognition. IEEE Trans Image Process 2022;31:788–98.
49 Lu N, Wu Y, Feng L, *et al*. Deep learning for fall detection: three-dimensional CNN combined with LSTM on video kinematic data. IEEE J Biomed Health Inform 2019;23(1):314–23.
50 Liu W, Liu X, Hu Y, *et al*. Fall detection for shipboard seafarers based on optimized BlazePose and LSTM. Sensors (Basel) 2022;22(14):5449.
51 Feichtenhofer C, Fan H, Malik J, *et al*. Slowfast networks for video recognition. ICCV; 2019:6202–11.