Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 20(6): 11238–11259.
DOI: 10.3934/mbe.2023498
Received: 08 March 2023
Revised: 07 April 2023
Accepted: 16 April 2023
Published: 25 April 2023

Research article

Fall detection based on dynamic key points incorporating preposed attention

Kun Zheng[1], Bin Li[1], Yu Li[1], Peng Chang[1], Guangmin Sun[1], Hui Li[1,]* and Junjie Zhang[2,]*

[1] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[2] Smart Learning Institute, Beijing Normal University, Beijing 100875, China

* **Correspondence:** Email: lihui@bjut.edu.cn, 11132022024@bnu.edu.cn; Tel: +8618810899182, +8618911610204.

Abstract: Accidental falls pose a significant threat to the elderly population, and accurate fall detection from surveillance videos can significantly reduce the negative impact of falls. Although most fall detection algorithms based on video deep learning focus on training and detecting human posture or key points in pictures or videos, we have found that the human pose-based model and key points-based model can complement each other to improve fall detection accuracy. In this paper, we propose a preposed attention capture mechanism for images that will be fed into the training network, and a fall detection model based on this mechanism. We accomplish this by fusing the human dynamic key point information with the original human posture image. We first propose the concept of dynamic key points to account for incomplete pose key point information in the fall state. We then introduce an attention expectation that predicates the original attention mechanism of the depth model by automatically labeling dynamic key points. Finally, the depth model trained with human dynamic key points is used to correct the detection errors of the depth model with raw human pose images. Our experiments on the Fall Detection Dataset and the UP-Fall Detection Dataset demonstrate that our proposed fall detection algorithm can effectively improve the accuracy of fall detection and provide better support for elderly care.

Keywords: preposed attention; fall detection; dynamic key points; decision fusion; complementary correction

1. Introduction

Falls, which are unexpected and uncontrolled changes in body posture, commonly occur among the elderly and are a major cause of health problems worldwide. They can lead to various degrees of physical and psychological damage and even death. According to the United Nations World Population Ageing 2020 report, the elderly population is expected to exceed 1.5 billion in 2050, and the percentage of people aged 65 and older is projected to increase from 9.3% in 2020 to approximately 16.0%. With the gradual aging of the global population, more and more elderly people will be exposed to the risk of accidental falls. Fall detection is essential to advanced first aid systems, providing timely assistance to improve the safety of elderly people living alone and avoid injuries caused by falls. The framework proposed by Chen et al. [1] improves the efficiency of first aid for older adults during falls, shortening resuscitation time and reducing the severity of injuries.

In recent years, researchers have conducted numerous studies in the field of fall detection. According to Mubashir [2], fall detection can be mainly divided into three approaches: wearable sensor-based, scene-based sensor-based, and computer vision-based. The wearable sensor-based approach relies on sensors worn by the person to detect falls, while the scene-based sensor-based approach uses sensors installed in the environment to detect falls. The computer vision-based approach uses cameras and computer algorithms to detect falls from video footage.

1) Fall detection using wearable devices is a promising approach, with single-sensor-based systems offering high accuracy and multiple-sensor-based systems offering higher efficiency [3]. Some recent studies have proposed innovative wearable fall detection systems. For example, de Sousa et al. [4] developed a low-power wearable system that integrates a fall detection system with an inflatable airbag to protect the hip. Lin et al. [5] used wireless in-shoe insoles equipped with sensors and a trained reference noise-canceling autoencoder to rapidly identify falls in all four directions. Bet et al. [6] described the latest technology and features for wearable sensors in detecting falls and found that the waist was the most common wearing position. Boudouane et al. [7] proposed a fall detection device using a portable camera worn on the hip, which effectively reduced specificity in fall detection. However, there are challenges to implementing wearable fall detection systems due to the lack of uniform standards for study populations and fall risk assessment, the high cost of sensors, and the cumbersome and forgetful nature of wearing devices for elderly individuals. Additionally, Casilari and Silva [8] found significant differences between simulated falls and real falls when using wearable sensors for fall detection, suggesting that further research is needed in this area.

2) Environmental sensor-based approaches rely on installed sensors that capture information such as sound and pressure to identify fall actions. For instance, Wang et al. [9] proposed a fall detection method that uses Wi-Fi signals based on the theory of indoor propagation of wireless signals, which has several advantages such as non-invasiveness and versatility. Other researchers, such as Madansingh et al. [10], have designed smartphone-based fall detection systems that utilize available embedded sensors such as accelerometer, gyroscope, and magnetometer. Wang et al. [11] proposed a radar-based fall detection method using pattern contour-confined Doppler-time maps that reduce superfluous information to improve detection accuracy. Chaccour et al. [12] summarized the deployment of sensors and proposed a generic fall detection classification scheme. However, such sensors are sensitive to environmental information, such as noise and light, which can result in false alarms and low detection rates.

3) Computer vision-based approaches detect falls by analyzing visual information extracted from

images and videos. In contrast to machine vision, recent research by Gutiérrez [13] has reviewed 81 fall detection systems based on computer vision in the last five years, which have improved from simple characterization to sophisticated classification techniques. However, Gutiérrez also indicated that existing public datasets lack a common reference framework for system performance evaluation, which makes it difficult to compare the performance of different systems.

Compared to the first two approaches, computer vision-based systems offer several practical advantages, including lower equipment costs, greater convenience for the elderly, and richer semantic information contained in the video captured by the camera, which can be retrieved and reviewed at any time. These factors make computer vision-based systems more promising for future applications.

This paper introduces three types of methods for vision-based fall detection: threshold-based, machine-learning-based, and deep learning-based. The threshold analysis method uses real-time acceleration or mean square value data after secondary processing as feature values, comparing them with thresholds derived from numerous fall experiments. When the data exceeds or falls below the threshold, it is deemed abnormal, and this method is effective in detecting and identifying fall events from continuous data [14]. The threshold-based method is often used in combination with other methods to enhance performance [15].

Machine learning-based detection algorithms process and describe underlying visual data using manually designed feature operators such as HOG features [16], ORB features [17], and LBP features [18], and these algorithms incorporate classifiers with multiple functions, providing significant classification advantages. However, these algorithms rely excessively on the manual extraction of action features, and any bias in the feature samples significantly degrades the performance of fall detection. Furthermore, these feature operators have difficulty capturing semantic information in complex images, whereas deep learning-based detection algorithms can compensate for this drawback, with feature values automatically derived from sample data to create a learning strategy with strong feature representation capabilities [19]. Liu et al. [20] proposed an accelerometer signal enhancement model based on deep learning as a front-end processor to improve the detection performance of low-resolution fall detection systems. Yu et al. [21] compared the three types of methods: threshold-based, machine-learning-based, and deep learning-based, and the results showed that the deep learning-based methods outperformed the other two.

In summary, vision-based fall detection methods can be divided into threshold-based, machine-learning-based, and deep learning-based approaches. While the threshold-based method is useful for detecting and identifying fall events from continuous data, machine learning-based methods provide significant classification advantages but can be sensitive to biased feature samples. In contrast, deep learning-based methods automatically extract feature values from sample data, compensating for the shortcomings of manually designed feature operators, and have shown excellent performance in detecting falls.

With the rapid development of deep learning, attention mechanisms have become widely used in image detection. Drawing inspiration from human attention, experts and scholars have proposed attention mechanisms that efficiently allocate information processing resources. The structural model, which includes the attention mechanism, not only records the positional relationships between information but also measures the importance of different information features based on their weights. Lu et al. [22] proposed a graph neural network based on the attention mechanism that holistically solves tasks by formulating them as an iterative information fusion process on the data graph. In another study, Abdulwahab et al. [23] classified and analyzed existing feature selection methods and applied them to reduce dimensions when processing high-dimensional data. These methods help to overcome the inefficiencies and ineffectiveness of processing high-dimensional data.

With the advancements in image processing technology and the widespread application of machine learning algorithms, researchers worldwide have conducted numerous studies related to fall detection algorithms. Mrozek et al. [24] developed a new scalable architecture system for remote monitoring, using Random Forest (RF), SVM, Artificial Neural Network (ANN), and Boosted Decision Trees (BDT) classifiers. Cai et al. [25] proposed a vision-based fall detection method using multi-task hourglass convolutional self-coding. The method involved extracting multi-scale features by extending the perceptual field of neurons through hourglass convolutional layers and completing the detection of fall actions and frame reconstruction tasks in parallel using a multi-task mechanism. Vishnu et al. [26] introduced the fall motion mixture model, which effectively recognizes fall events in various scenarios.

Currently, the majority of fall detection algorithms only detect the original image or key points of the human body in the visual dimension, leaving room for further improvement in accuracy. Additionally, some key points of the human body may not be detectable due to environmental factors, and not all key points of the human body may be relevant for fall detection. To address these issues, this paper aims to make the following contributions.

1) A new set of dynamic key points based on high-frequency human key points are proposed for fall detection. These key points are identified during a fall state and automatically labeled to implement an attention capture mechanism using the original data-driven attention mechanism of the depth model. The attention is distinguished by the size of pixel values.

2) The paper also demonstrates that the depth model, which incorporates both human pose and key points, can yield strong complementarity in fall detection results. To this end, a fall detection model is proposed that incorporates dynamic key points and the preposed attention capture mechanism.

The main structure of this paper is as follows: Section 2 introduces the method utilized in this paper, Section 3 shows the experimental results and analysis, Section 4 contains the discussion and outlook, and finally concludes the work of this paper.

2. Materials and methods

2.1. Deep learning model based on attention mechanism

In this paper, we utilized Swin Transformer [27] to train the model, as it has demonstrated excellent performance in the field of image analysis. Unlike the original Transformer [28] that employs self-attention as a layer in the network structure and uses self-attention and feedforward networks for encoding and decoding, Swin Transformer introduces the hierarchical construction method typically used in CNNs [29] and conducts self-attention computation in the window region without overlap, which enhances efficiency and reduces computational complexity. The whole model comprises four stages, each of which reduces the resolution of the input feature map and expands the perception field layer by layer like CNN. Figure 1 depicts the system structure. Model-a is the deep learning model without the preposed attention, while Model-b is the deep learning model with the preposed attention, which is trained using labeled key points.

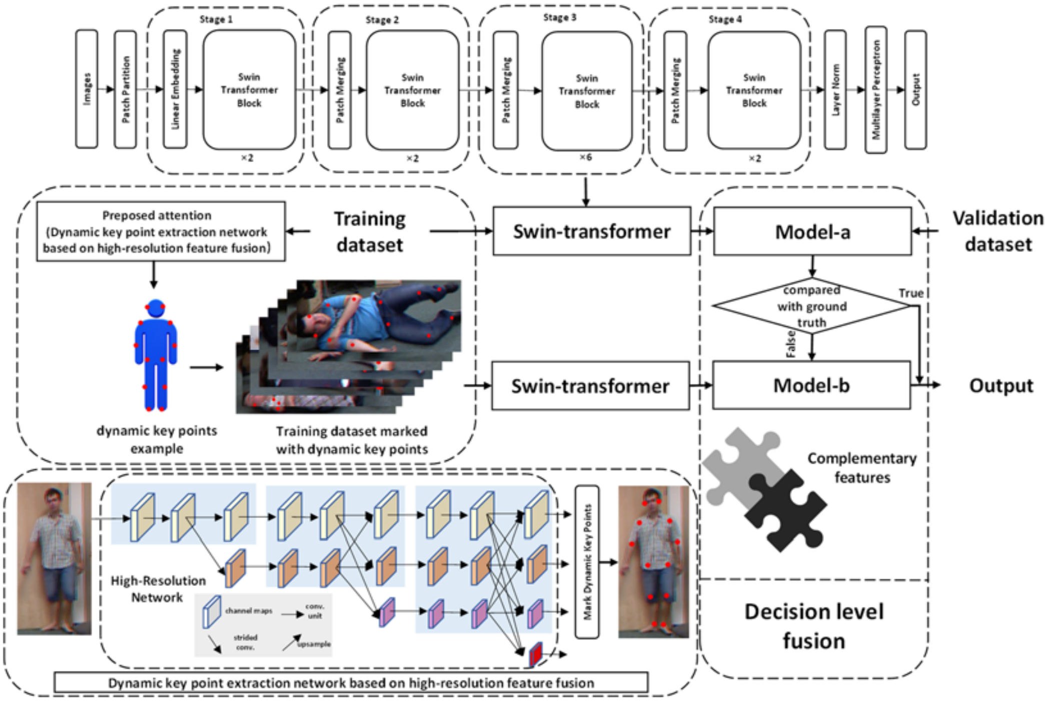


Figure 1. System structure diagram.

2.2. Preposed attention

At present, most fall detection algorithms only rely on a single dimension of information, such as visual images or key points. To improve detection accuracy, it is necessary to expand the dimensionality of information properly. In this paper, we propose a preposed attention mechanism to enhance the images fed into the Swin Transformer training network. We achieve this by adding dynamic key points, visually stimulating elements that can attract more attention and cause unconscious attention shifting [30]. Our experiments explore whether there is a complementary fusion between the dynamic key points and the unmarked images, and the results will be presented in this paper.

For labeling the key points, we selected the High-Resolution Network (HRNet) [31], which is based on a multi-resolution feature fusion network. This network changes the link between high and low resolutions from series to parallel, maintains high-resolution representation throughout the network structure, and introduces interaction between high and low resolutions to improve model performance. Some commonly used key point detection datasets include the COCO dataset [32], which represents human key points as 17 joints with up to 300 K samples; the MPII Human Pose Dataset [33], which contains 25 K samples and defines 16 key point coordinates; and the Leeds Sports Pose Dataset contains 2 K samples, with 14 key points per sample [34]. The HRNet network follows the heat map estimation framework and applies it to human pose estimation, and has experimentally demonstrated superior pose estimation performance on the COCO key points detection dataset. Therefore, we trained our model on the COCO key points detection dataset, which can detect a total of 17 human key points.

We calculated the frequency of key points on the UR Fall Detection Dataset (URFD) [35], which includes 30 video segments captured during various daily activities. HRNet was used to label the fall state map for each segment. Due to the fall posture, not all key points may contribute to fall detection. Therefore, we counted the frequency of occurrence of each of the 17 key points in the fall state map and ranked them according to their frequency, as shown in Figure 2.

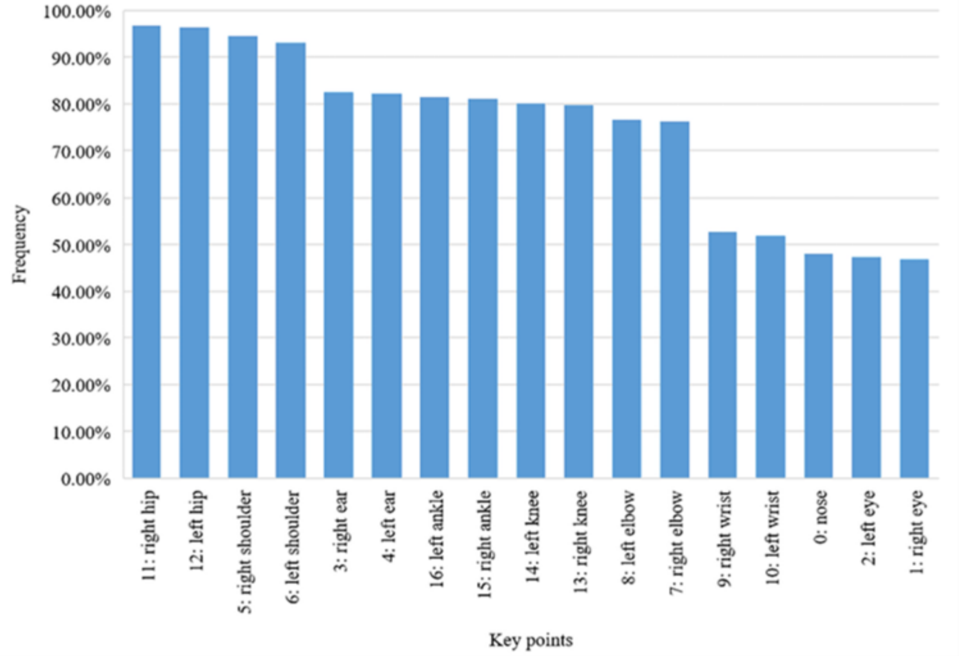
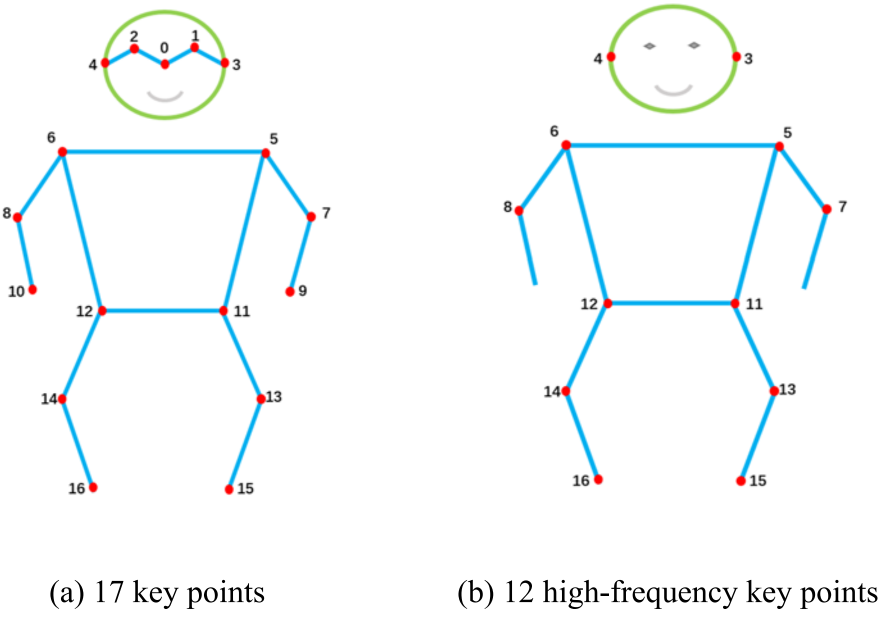


Figure 2. Human body key points frequency diagram.



(a) 17 key points (b) 12 high-frequency key points

Figure 3. Example diagram of human body key points. Which: 0: nose, 1: right eye, 2: left eye, 3: right ear, 4: left ear, 5: right shoulder, 6: left shoulder, 7: right elbow, 8: left elbow, 9: right wrist, 10: left wrist, 11: right hip, 12: left hip, 13: right knee, 14: left knee, 15: right ankle, 16: left ankle.

We selected the 12 points with frequencies higher than 75%: namely 3–8 and 11–16, and defined them as high-frequency key points. A comparison of 17 key points detected by the original HRNet network and the 12 key points is shown in Figure 3.

Fall detection can suffer from incomplete information due to environmental occlusion and obstacles, which often results in missing some of the 12 high-frequency key points identified in the URFD. To obtain as much accurate information as possible, we propose the use of dynamic key points, which correspond to the maximum possible detection of the 12 unobstructed high-frequency key points. The number of dynamic key points varies dynamically based on the detection of the obscured parts, up to a maximum of 12. We present the pseudo-code for dynamic key points extraction in Algorithm 1.

Algorithm 1. Marking dynamic key points with HRNet

```
Input: Original images
Load the trained 17 human key points models
Adopt the top-down detection method, detect the human body first, then mark the key points
repeat
    Extract the features of key points in the human image
    Locate 17 key points // Use the extracted features to locate the 17 key points on the human body.
    for i ∈ (3,8) & (11,16) do // Iterating on our selected high-frequency key points.
        if len(pred_boxes) >= 1 then // Check if the model detected the body part
            corresponding to the key point from the picture
            marked key points
    end
until all 12 key points are detected // Repeat the process until all 12 key points are detected.
Output: Dynamic key points picture marked
```

By adding these dynamic key points, we aim to provide additional information that can aid in fall detection. Specifically, we add the dynamic key points to the image before it enters the Swin Transformer network, and we make them more visually stimulating, so they are likely to attract more attention than the unmarked images alone. The Swin Transformer network can then use this information to make more accurate predictions about the fall state.

Overall, the preposed attention mechanism with the dynamic key points aims to provide a complementary fusion between the dynamic key points and the unmarked images. This fusion enhances the images' quality and makes them more informative, leading to improved fall detection performance.

2.3. Decision fusion

In this paper, we propose the use of dynamic key points for fall detection, which enhances the attentional capture of the model and improves detection accuracy despite incomplete information caused by environmental occlusion and obstacles. To achieve complementary effects, the results of two models are fused at the decision level. Model-a is used to validate the validation set and output error images separately. Then, the error images are validated using Model-b, to achieve decision fusion of the two models. The proposed approach is illustrated in Figure 4.

As shown in Figure 4, Model-a is trained on the data layer composed of original images through the Swin-transformer. Meanwhile, Model-b is trained on the data layer composed of images marked with dynamic key points through the same architecture. Compared with Model-a, the data layer of Model-b added the size and coordinate information of dynamic feature points. Due to being trained on different data, the two models have different feature layers, which allow them to capture different aspects of the data.

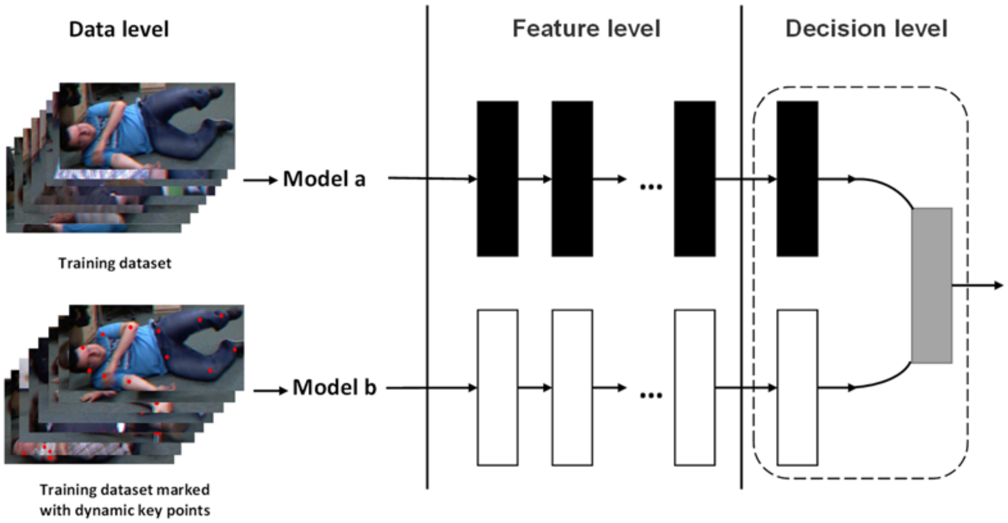


Figure 4. Demonstration of information fusion at different levels.

At the decision level, Model-a and Model-b are fused to achieve complementary effects. Due to the differences in the data layer, the feature layers of Model-a and Model-b contain different network structure features and weights, which results in different decisions being output for the same data to be verified. In our method, specifically, the error images that Model-a fails to detect are sent to Model-b for further detection, and the final result is based on the detection results of Model-b.

By combining the strengths of both models, our proposed approach achieves improved accuracy in fall detection, even in scenarios with incomplete information caused by environmental occlusion and obstacles.

2.4. Missing Detection Correction Rate (MDCR) and Error Detection Correction Rate (EDCR)

In this paper, we define missing detection as the model fails to detect images with ground truth as fall, and the MDCR is expressed as (1):

$$MDCR = \frac{FN-MV_m}{FN} \quad (1)$$

where FN represents the total number of false negatives in the dataset that were missed by Model-a. MV_m, on the other hand, represents the number of false negatives that were corrected after being processed by Model-b.

Similarly, error detection is defined as the model fails to detect images with safe ground truth, and the EDCR is expressed as (2).

$$EDCR = \frac{FP - MV_e}{FP} \tag{2}$$

where FP represents the total number of false positives in the dataset that were missed by Model-a. MV_e, on the other hand, represents the number of false positive that were corrected after being processed by Model-b.

The results of EDCR and MDCR directly reflect the corrective effects of our proposed dynamic key points on the original image, further demonstrating the complementarity between Model-b and Model-a.

2.5. Related terms

In order to compare the effectiveness of different fall detection methods from various aspects, we define the key terms related to fall detection by a confusion matrix as shown in Table 1.

Table 1. Terms related to the confusion matrix of fall detection.

Test results	Positive (factual falls)	Negative (factual safe)	Total
Positive (Detecting falls)	True Positive (TP)	False Positive (FP)	TP + FP
Negative (Detecting safe)	False Negative (FN)	True Negative (TN)	FN + TN
Total	TP + FN	FP + TN	TP + FP + FN + TN

Note: TP: Samples predicted by the model as fall and ground truth as fall. FN: Samples predicted by the model as safe but ground truth as fall. FP: Samples predicted by the model as fall but ground truth as safe. TN: Samples predicted by the model as fall and ground truth as safe.

Sensitivity is defined as the proportion of TP to all factual positives of the participating tests. In this paper, sensitivity reflects the sensitivity of the applied method to falls, and experiments are conducted on two datasets.

The specificity of a fall detection method is defined as the proportion of TF to all actual negatives in the tested dataset. In this study, specificity reflects the accuracy of the proposed method for detecting falls safely.

3. Results

To investigate the potential complementarity between Model-a and Model-b in fall detection, we conducted separate experiments on the Fall Detection Dataset (FDD) [36] and the UP-Fall Detection Dataset (UFDD) [37]. We evaluated the missed detection correction rate, false detection correction rate, and corrected accuracy. The Swin Transformer-based model was trained on the URFD, and the best model was tested on the FDD and UFDD datasets, with the detected erroneous images saved as output. We then labeled all key points and dynamic key points defined in the training set images of the model and categorized them into groups No.1–No.5, as shown in Figure 5, with all 17 key points labeled as an example.

To assess the effectiveness of dynamic key points in improving fall detection accuracy, we created four different training sets comprising different types of data: 1) all key points, 2) all key points and original images, 3) dynamic key points, and 4) dynamic key points and original images. We then trained models on each of these sets and used them to validate the output error images. This allowed us to determine the gain effect of dynamic key points and evaluate their impact on the accuracy of fall detection.

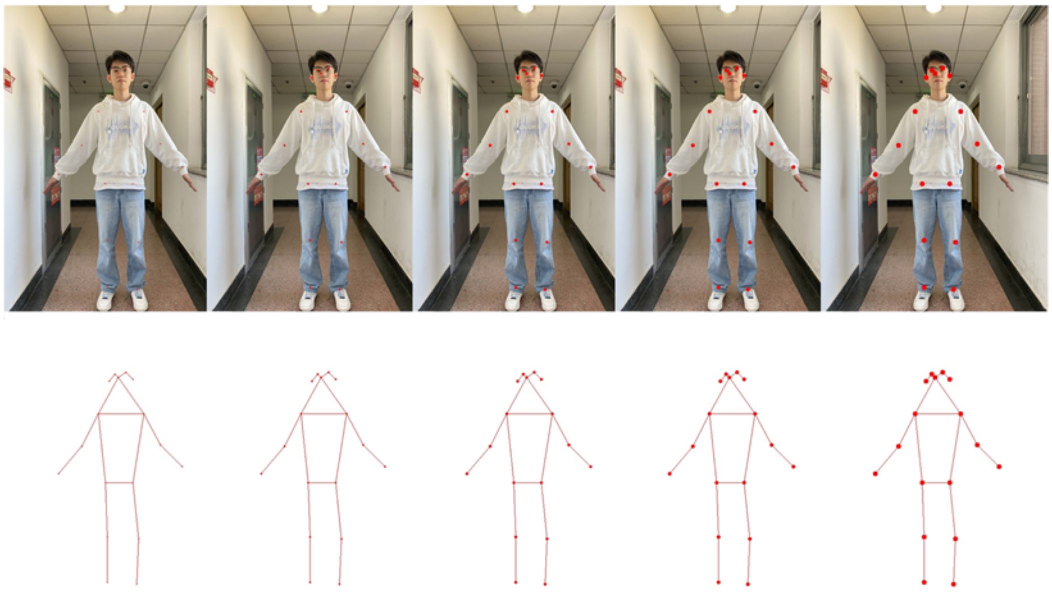


Figure 5. Demonstration of key points groups.

3.1. MDCR

We conducted test experiments on four methods regarding the MDCR are tested. The test results on FDD are shown in Figure 6.

From Figure 6, it can be seen that the MDCR of the No.3 and No.4 key points groups is better than that of the other three groups. In terms of methods, the dynamic key points perform better.

In the field of fall detection, determining the accuracy of the result is important in addition to the final test result. We construct the same dataset for the experiments.

The average correction rate of key points group No.4 was found to be the highest among the four methods. Therefore, key points group No.4 was taken as an example to count the recognition measurement accuracy of the four different methods used to correct false negative (FN) samples. Figure 7 shows the accuracy distribution of the four methods. A–D represent the following respectively: All key points, All key points and original images, Dynamic key points, Dynamic key points and original images. The red line represents the median of the box, and we have labeled the median and interquartile range for each box.

From the box plot, it can be seen that the dynamic key points method has the highest overall distribution of data, as evidenced by both the interquartile range and the median.

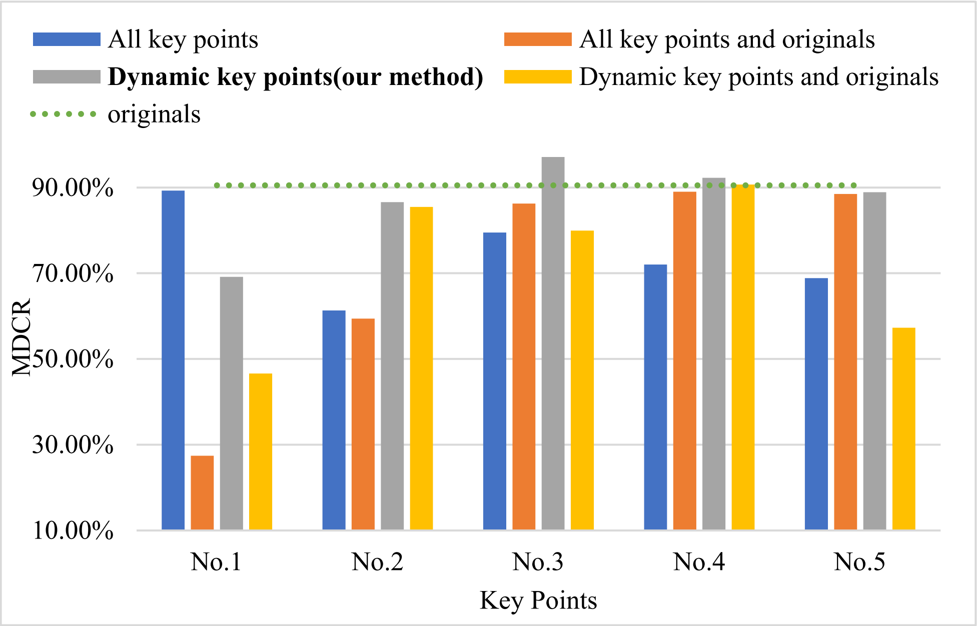


Figure 6. Comparison of the MDCR by four methods on the FDD dataset.

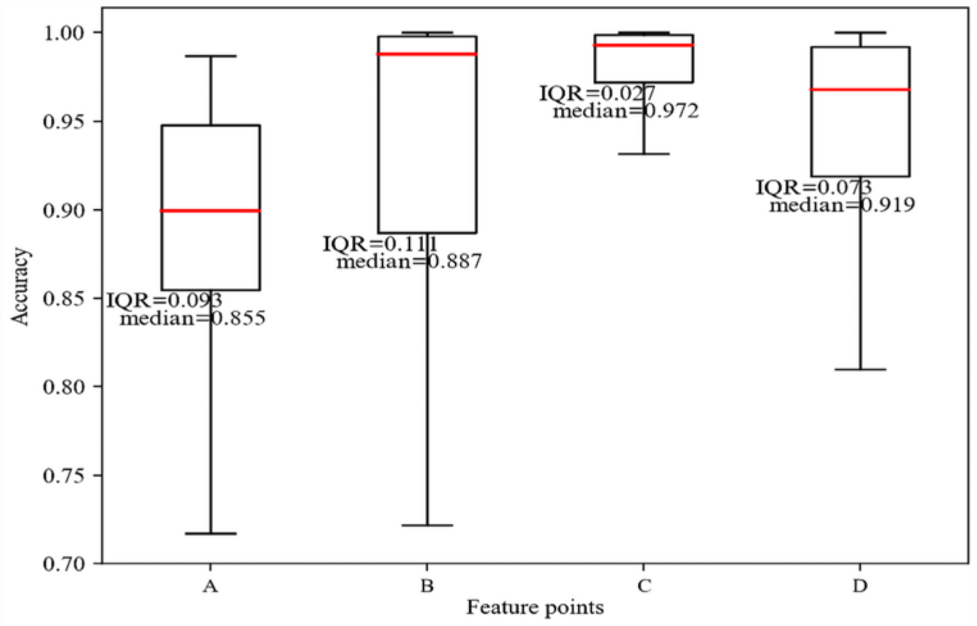


Figure 7. Comparison of the Accuracy of groups No.4 key points on the FDD dataset.

3.2. EDCR

The test results of the four methods on UFDD are shown in Figure 8.

From Figure 8, it can be seen that the average EDCR of the No.4 key points group is superior to the other four groups, and the dynamic key points method in No.4 achieves the highest accuracy.

We selected the best group of 5 key points based on the MDCR results, and then determined that the dynamic key points method in Group No.4 had the highest accuracy rate in correcting FN data. In the EDCR section, we found that the Dynamic key points method yielded the highest accuracy. Thus, we used this method as an example and assessed the recognition accuracy of corrected FP samples using five different key point groups. We also marked the median and interquartile range of the box plot in Figure 9. From the box plot, it can be seen that the No.4 key points group yielded the best results.

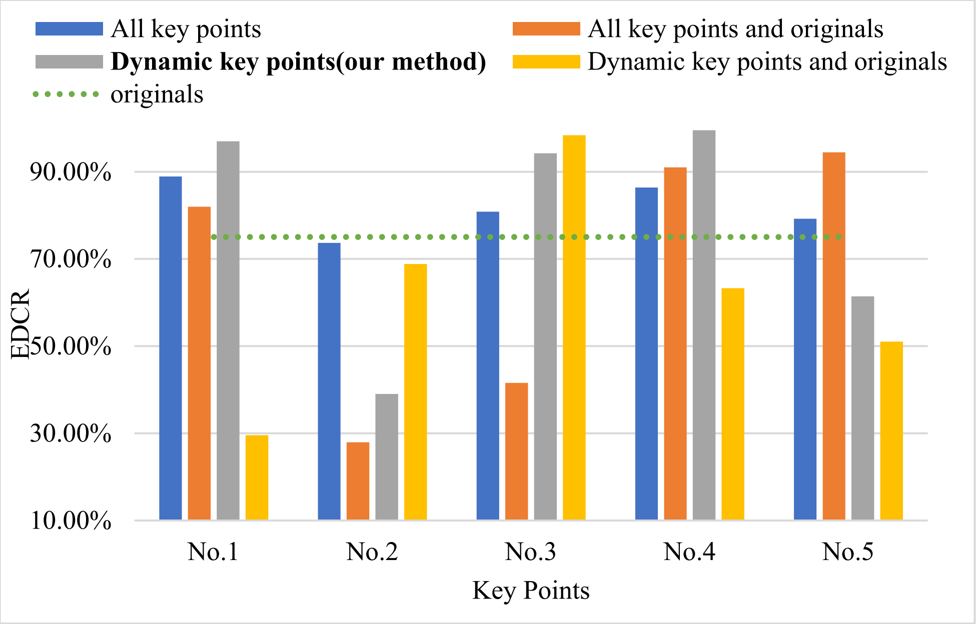


Figure 8. Comparison of the EDCR by four methods on the UFDD dataset.

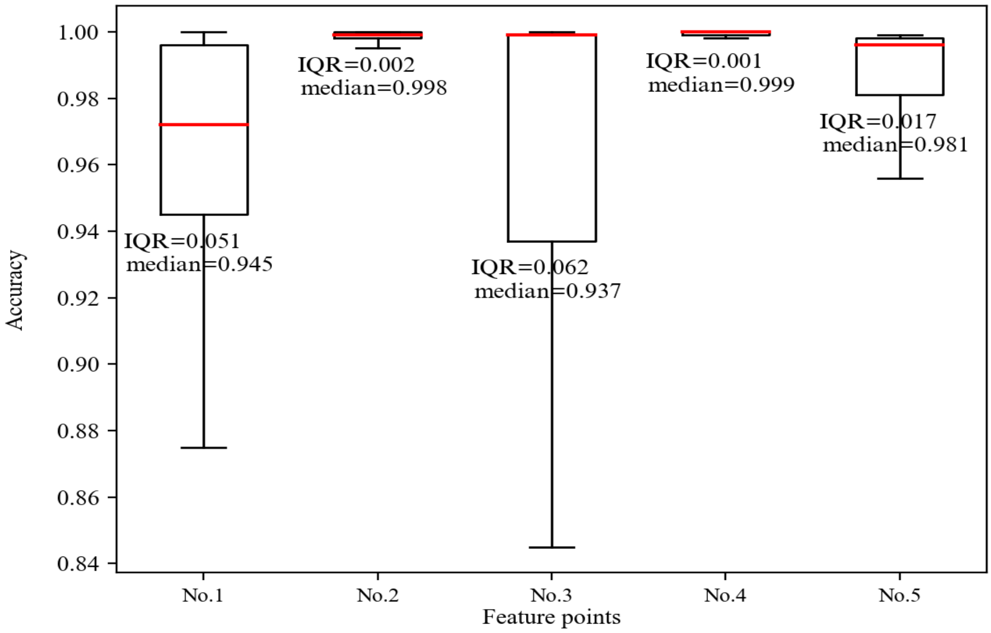


Figure 9. Comparison of the Accuracy of 5 groups Dynamic key points on the UFDD dataset.

3.3. Accuracy

In this paper Accuracy is defined as the accuracy corrected by Model-b, expressed as (3):

$$Accuracy = \frac{A-FN-FP+MV_M+MV_E}{A} \tag{3}$$

where A is All images of the test.

The experimental results are shown in Figures 10 and 11.

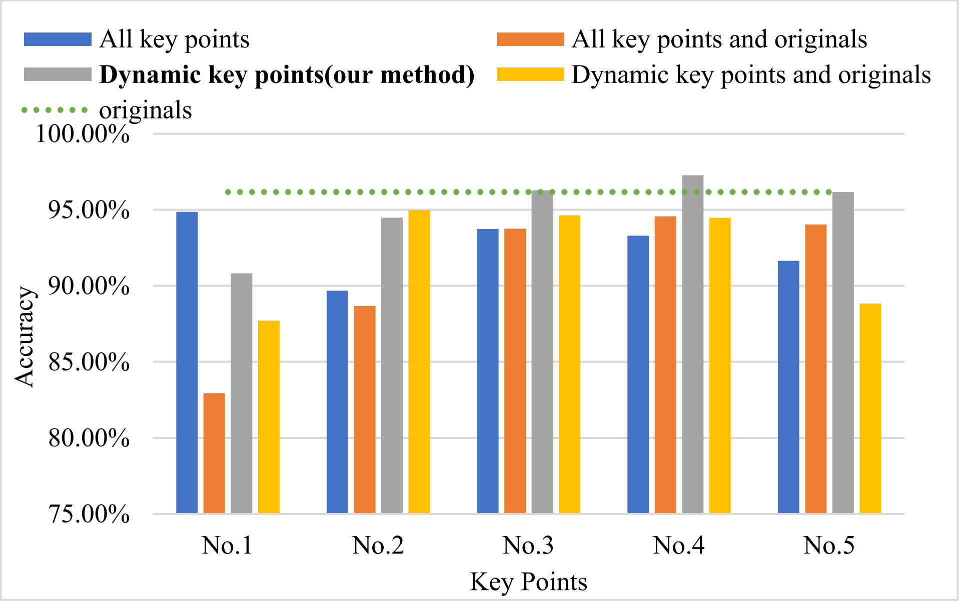


Figure 10. Accuracy obtained by four methods on the FDD dataset.

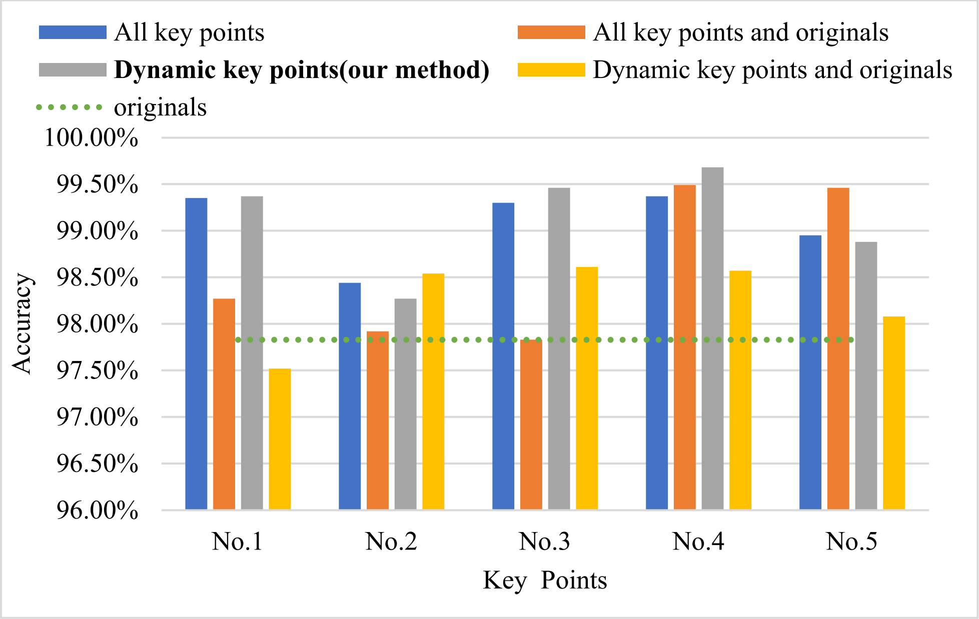


Figure 11. Accuracy obtained by four methods on the UFDD dataset.

From Figures 10 and 11, it can be seen that among the four methods and 5 key point groups, the Dynamic key points method of the No.4 group has the highest accuracy.

3.4. Related terms

The test results of sensitivity on FDD are shown in the Figure 12. From the figure it demonstrated that only the Dynamic key points method of groups No.3 and No.4 outperforms the correction effect of origin.

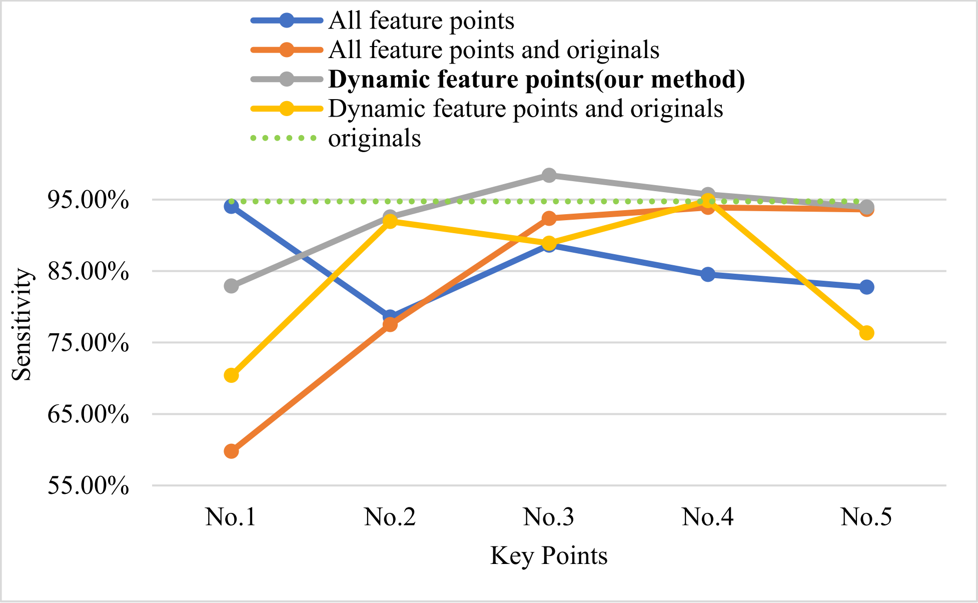


Figure 12. Comparison of the Sensitivity by four methods on the FDD dataset.

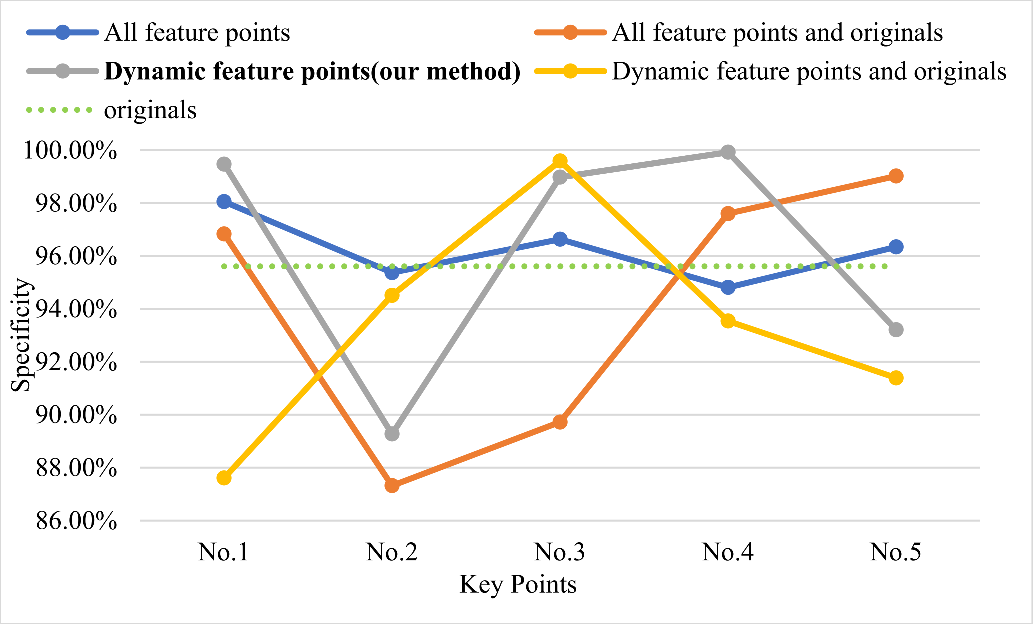


Figure 13. Comparison of the Specificity by four methods on the UFDD dataset.

We evaluated the specificity of our method on both datasets, and the results for the UFDD are shown in Figure 13. From the figure, it can be observed that half of the methods outperform the baseline, and the Dynamic key points method using key points group No.4 achieves the highest specificity.

We compared the training and validation speed of adding preposed attention and not adding preposed attention in Tables 2 and 3. As can be seen from the tables, with preposed attention, the training and validation speed of the data only decreased slightly. Therefore, it can be concluded that our proposed preposed attention did not increase the computation burden.

Table 2. Comparison of iteration speeds of data training processes with or without preposed attention.

Model	Average iterations per second (it/s)
Model-a	3.07
Model-b	3.01

Table 3. Comparison of iteration speeds of data validation processes with or without preposed attention.

Model	Average iterations per second (it/s)
Model-a	8.77
Model-b	8.61

Table 4 compares the accuracy of different algorithms on the dataset URFD, and it can be indicated that our method outperforms most algorithms in terms of detection accuracy, slightly lower than the hybrid algorithm [42,43] and Deep network [44] with higher detection cost and time.

Table 4. Comparison of fall detection methods in URFD.

Methods	Accuracy (%)
CNN [38]	96.43
1D CNN [38]	92.72
OF CNN [40]	83.02
OF CNN [41]	91.40
multi-stream DNN [42]	99.72
3D CNN, LSTM [43]	99.27
Deep network [44]	99.67
Our method	99.20

Table 5. Results of our method for cross dataset detection.

Dataset	Accuracy (%)
FDD	97.27
UFDD	99.68

In order to verify the applicability and robustness of the method in this paper, we did cross-library detection. The model is trained on URFD and detected on FDD and UFDD respectively, and the results are shown in Table 5. Based on the accuracy it can be indicated that the robustness of the method in this paper is strong.

After analyzing the experimental results, it has been demonstrated that each method has different accuracy improvement effects for different indexes. For instance, the Dynamic key points method of groups No.3 displays the most significant improvement for the correction of missed detections, but the correction effect for false detections is mediocre. On the other hand, the Dynamic key points method of groups No.4 achieves the best improvement in both datasets for the final accuracy index, which represents the comprehensive performance of the proposed method. These results further validate the effectiveness of the proposed method.

4. Discussion

Fall detection algorithms play an important role in fall detection and fall prevention. Each fall detection algorithm is capable of achieving very high accuracy, and applying fall detection to real life is an urgent issue to be considered at present. Through the above experiments, it was shown that the performance of the Model-b, compared with the Model-a, did not improve significantly, and even decreased slightly. The Model-b trained with Dynamic key points groups No.4, which has the best results in the above experiments, was selected for comparison, as shown in Table 6.

Table 6. Accuracy of the two models on different datasets.

Model	URFD	FDD	UFDD
Model-a	0.987	0.926	0.931
Model-b	0.982	0.938	0.986

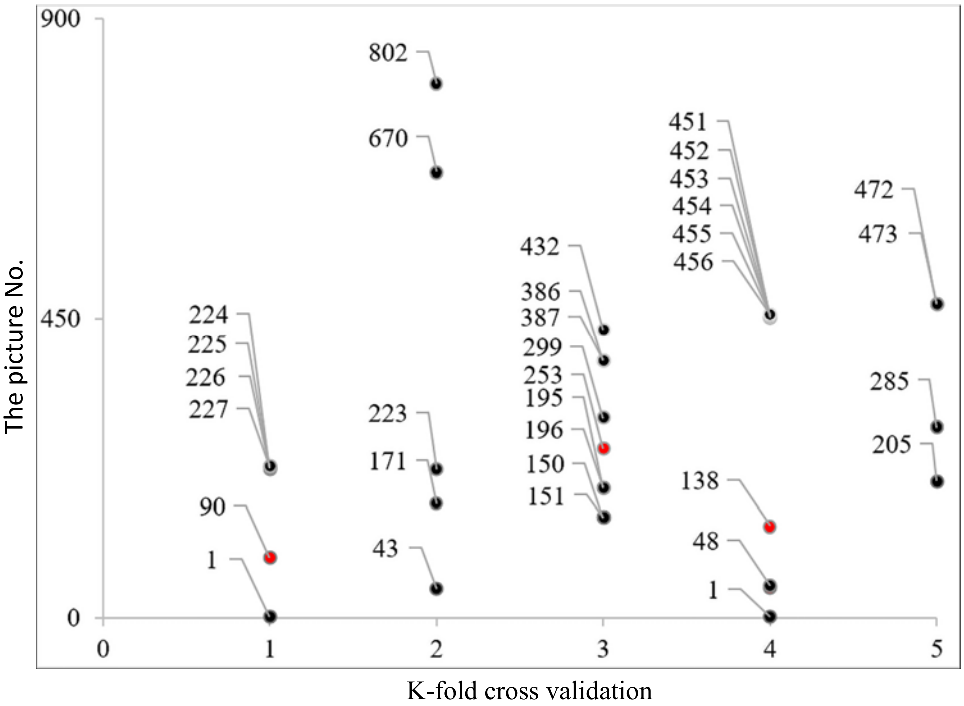


Figure 14. Correction results of Model-a.

Although the improvement in accuracy with Model-b alone was small, further analysis revealed strong complementarity between Model-a and Model-b. To investigate this, we used the highest-accuracy Model-a trained on URFD in a 5-fold cross-validation approach (defined as K1–K5), where we labeled the validation set images of each fold, sequentially validated and outputted the images with detection errors. We then used Model-a and Model-b in turn to detect the output images. Black dots indicate the error detection result, red dots indicate the corrected error result, and more red dots represent stronger complementarity between the models. The results are shown in Figures 14 and 15.

As an example, we take K1 of Figure 14: in the K1 fold, the images incorrectly detected by the first CNN model are No.1, No.90, No.224, No.225, No.226, and No.227. The No.90 image is corrected after being detected by the Model-a with the next highest accuracy. ●: error detection result, ●: corrected error result.

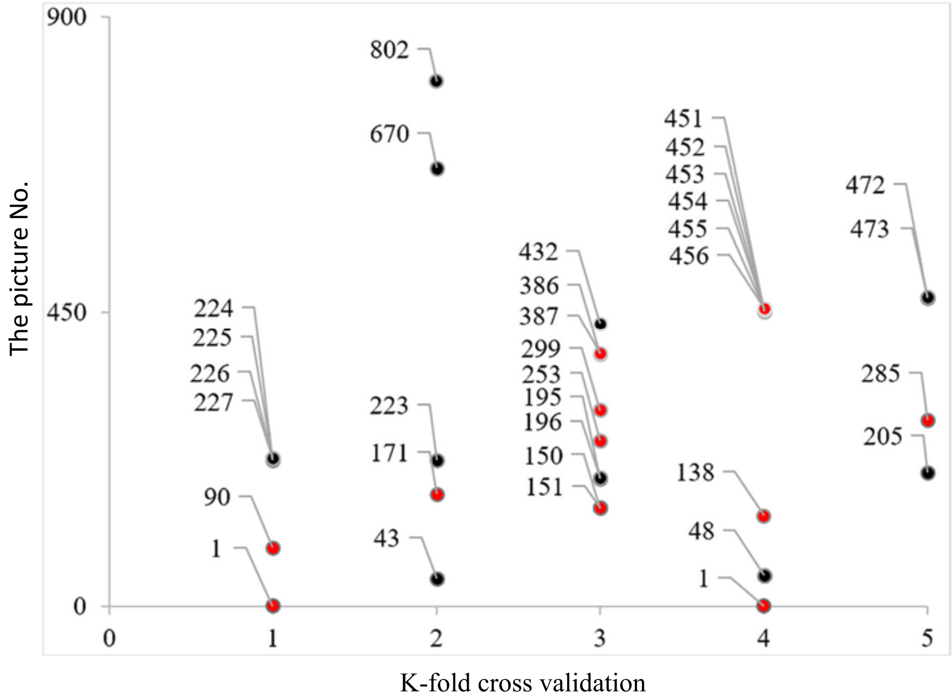


Figure 15. Correction results of Model-b.

The experiments described above have demonstrated the complementarity between the models trained using dynamic key points and original images. To explore whether training the models with dynamic key points first and then using original images for complementary correction would yield similar results, we conducted additional experiments. The results showed that original images do have a correction effect, and the final accuracy is similar. This provides further evidence for the complementarity between the two approaches. However, the correction effect is lower compared to the method proposed in this paper. The results of these additional experiments are summarized in Tables 7 and 8.

Table 7. Data on FDD in two orders.

FDD	MDER	EDCR	Accuracy	sensitivity	specificity
Exchange Order	0.5238	0.5	0.97	0.971	0.9687
Our method	0.9225	0.847	0.973	0.9571	0.9852

Table 8. Data on UFDD in Two Orders.

UFDD	MDER	EDCR	Accuracy	sensitivity	specificity
Exchange Order	0.7673	0.867	0.997	0.9964	0.9992
Our method	0.9268	0.995	0.997	0.9848	0.9939

In the field of biomedical imaging, Optical Coherence Tomography (OCT) is a commonly used non-invasive detection technology. Among them, Swept-Source OCT (SS-OCT) system uses a rapidly scanning laser for imaging in the 1300 nm bio-imaging window [45]. Various automatic calibration schemes have been widely studied to ensure the high-precision imaging of the SS-OCT system. Ratheesh et al. [46] proposed an automatic calibration scheme based on spectral phase. Meleppat et al. [47] proposed an effective wavenumber linearization method that can help SS-OCT systems achieve more accurate imaging. At the same time, dynamic key points detection technology has become an important research direction for real-time monitoring of small structural changes in biological tissue.

In cutting-edge research in this field, Meleppat et al. [48] successfully achieved imaging of the retina and choroid in different types of vertebrate eyes through experiments. Murukeshan et al. [49] introduced a method for quantitative detection of biofilm thickness using an SS-OCT system.

Based on these cutting-edge studies, we can consider new application scenarios, such as whether marking key points at the sclera or iris of the pupil, and combining dynamic key points detection technology, can help the SS-OCT system detect eye or retina structural changes earlier or more quickly, thus achieving more accurate diagnosis and treatment of eye diseases. This is a new direction worth further research.

5. Conclusions

This paper presents a novel approach for fall detection using dynamic key points and a preposed attention mechanism that fuses this information with original human posture images. We propose a deep learning model that incorporates this attention mechanism, which improves the accuracy of fall detection when compared to models trained with original images and all key points. Our extensive experiments on multiple datasets demonstrate the complementarity of dynamic key points and original images at the decision level, resulting in improved detection accuracy. In the future, we plan to develop more attention capture methods to fuse complementary information from different modalities.

Acknowledgments

This work was supported in part by Scientific Research Plan of Beijing Education Commission (SZ202110005002), CN.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Chen, Y. Zhang, B. Xiao, H. Li, A framework for the elderly first aid system by integrating vision-based fall detection and BIM-based indoor rescue routing, *Adv. Eng. Inf.*, **54** (2022), 101766. http://doi.org/10.1016/j.aei.2022.101766
2. M. Mubashir, L. Shao, L. Seed, A survey on fall detection: Principles and approaches, *Neurocomputing*, **100** (2013), 144–152. http://doi.org/10.1016/j.neucom.2011.09.037
3. S. Nooruddin, M. Islam, F. A. Sharna, H. Alhetari, M. N. Kabir, Sensor-based fall detection systems: a review, *J. Ambient Intell. Hum. Comput.*, **2009** (2009), 1–17. http://doi.org/10.1109/biocas.2009.5372032
4. F. A. S. F. de Sousa, C. Escriba, E. G. A. Bravo, V. Brossa, J. Y. Fourniols, C. Rossi, Wearable pre-impact fall detection system based on 3D accelerometer and subject's height, *IEEE Sens. J.*, **22** (2022), 1738–1745. http://doi.org/10.1109/biocas.2009.5372032
5. Z. Lin, Z. Wang, H. Dai, X. Xia, Efficient fall detection in four directions based on smart insoles and RDAE-LSTM model, *Expert Syst. Appl.*, **205** (2022), 117661. http://doi.org/10.1016/j.eswa.2022.117661
6. P. Bet, P. C. Castro, M. A. Ponti, Fall detection and fall risk assessment in older person using wearable sensors: A systematic review, *Int. J. Med. Inf.*, **130** (2019), 103946. http://doi.org/10.1016/j.ijmedinf.2019.08.006
7. I. Boudouane, A. Makhlouf, M. A. Harkat, M. Z. Hammouche, N. Saadia, A. R. Cherif, Fall detection system with portable camera, *J. Ambient Intell. Hum. Comput.*, **11** (2019), 2647–2659. http://doi.org/10.1007/s12652-019-01326-x
8. E. Casilari, C. A. Silva, An analytical comparison of datasets of Real-World and simulated falls intended for the evaluation of wearable fall alerting systems, *Measurement*, **202** (2022), 111843. http://doi.org/10.1016/j.measurement.2022.111843
9. C. Wang, L. Tang, M. Zhou, Y. Ding, X. Zhuang, J. Wu, Indoor human fall detection algorithm based on wireless sensing, *Tsinghua Sci. Technol.*, **27** (2022), 1002–1015. http://doi.org/10.26599/tst.2022.9010011
10. S. Madansingh, T. A. Thrasher, C. S. Layne, B. C. Lee, Smartphone based fall detection system, in *2015 15th International Conference on Control, Automation and Systems*, ICCAS, (2015), 370–374. https://doi.org/10.1109/ICCAS.2015.7364941
11. B. Wang, Z. Zheng, Y. X. Guo, Millimeter-Wave frequency modulated continuous wave radar-based soft fall detection using pattern contour-confined Doppler-Time maps, *IEEE Sens. J.*, **22** (2022), 9824–9831. http://doi.org/10.1109/jsen.2022.3165188
12. K. Chaccour, R. Darazi, A. H. El Hassani, E. Andres, From fall detection to fall prevention: A generic classification of fall-related systems, *IEEE Sens. J.*, **17** (2017), 812–822. http://doi.org/10.1109/jsen.2016.2628099
13. J. Gutiérrez, V. Rodríguez, S. Martin, Comprehensive review of vision-based fall detection systems, *Sensors*, **21** (2021), 947. https://pubmed.ncbi.nlm.nih.gov/33535373
14. C. Y. Hsieh, K. C. Liu, C. N. Huang, W. C. Chu, C. T. Chan, Novel hierarchical fall detection algorithm using a multiphase fall model, *Sensors*, **17** (2017), 307. https://pubmed.ncbi.nlm.nih.gov/28208694
15. L. Ren, Y. Peng, Research of fall detection and fall prevention technologies: A systematic review, *IEEE Access*, **7** (2019), 77702–77722. http://doi.org/10.1109/access.2019.2922708

16. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2005), 886–893. http://doi.org/10.1109/cvpr.2005.177
17. E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in *2011 International Conference on Computer Vision*, (2011), 2564–2571. http://doi.org/10.1109/iccv.2011.6126544
18. X. Wang, T. X. Han, S. Yan, An HOG-LBP human detector with partial occlusion handling, in *2009 IEEE 12th International Conference on Computer Vision*, (2009), 32–39. http://doi.org/10.1109/iccv.2009.5459207
19. M. Islam, S. Nooruddin, F. Karray, G. Muhammad, Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges and future prospects, *Comput. Biol. Med.*, **149** (2022), 106060. http://doi.org/10.1109/iccv.2009.5459207
20. K. C. Liu, K. H. Hung, C. Y. Hsieh, H. Y. Huang, C. T. Chan, Y. Tsao, Deep-learning-based signal enhancement of low-resolution accelerometer for fall detection systems, *IEEE Trans. Cognit. Dev. Syst.*, **14** (2022), 1270–1281. http://doi.org/10.1109/tcds.2021.3116228
21. X. Yu, B. Koo, J. Jang, Y. Kim, S. Xiong, A comprehensive comparison of accuracy and practicality of different types of algorithms for pre-impact fall detection using both young and old adults, *Measurement*, **201** (2022), 111785. http://doi.org/10.2139/ssrn.4132951
22. X. Lu, W. Wang, J. Shen, D. J. Crandall, L. Van Gool, Segmenting objects from relational visual data, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2022), 7885–7897. http://doi.org/10.1109/tpami.2021.3115815
23. H. M. Abdulwahab, S. Ajitha, M. A. N. Saif, Feature selection techniques in the context of big data: taxonomy and analysis, *Appl. Intell.*, **52** (2022), 13568–13613. http://doi.org/10.1007/s10489-021-03118-3
24. D. Mrozek, A. Koczur, B. Małysiak-Mrozek, Fall detection in older adults with mobile IoT devices and machine learning in the cloud and on the edge, *Inf. Sci.*, **537** (2020), 132–147. http://doi.org/10.1016/j.ins.2020.05.070
25. X. Cai, S. Li, X. Liu, G. Han, Vision-based fall detection with multi-task hourglass convolutional auto-encoder, *IEEE Access*, **8** (2020), 44493–44502. http://doi.org/10.1109/access.2020.2978249
26. C. Vishnu, R. Datla, D. Roy, S. Babu, C. K. Mohan, Human fall detection in surveillance videos using fall motion vector modeling, *IEEE Sens. J.*, **21** (2021), 17162–17170. http://doi.org/10.1109/jsen.2021.3082180
27. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., Swin transformer: Hierarchical vision transformer using shifted windows, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), 10012–10022. https://doi.org/10.48550/arXiv.2103.14030
28. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al, Attention is all you need, in *Advances in Neural Information Processing Systems*, **30** (2017).
29. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE*, **86** (1998), 2278–2324. http://doi.org/10.1109/5.726791
30. H. Pashler, J. C. Johnston, E. Ruthruff, Attention and performance, *Ann. Rev. Psychol.*, **52** (2001), 629. https://doi.org/10.1146/annurev.psych.52.1.629
31. J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, et al., Deep high-resolution representation learning for visual recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, **43** (2021), 3349–3364. https://doi.org/10.1109/TPAMI.2020.2983686

32. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft coco: Common objects in context, in *European Conference on Computer Vision*, (2014), 740–755. http://doi.org/10.1007/978-3-319-10602-1_48
33. M. Andriluka, L. Pishchulin, P. Gehler, B. Schiele, 2d human pose estimation: New benchmark and state of the art analysis, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 3686–3693. http://doi.org/10.1109/cvpr.2014.471
34. B. Sapp, B. Taskar, Modec: Multimodal decomposable models for human pose estimation, in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, (2013), 3674–3681. http://doi.org/10.1109/cvpr.2013.471
35. B. Kwolek, M. Kepski, Human fall detection on embedded platform using depth maps and wireless accelerometer, *Comput. Methods Programs Biomed.*, **117** (2014), 489–501. http://doi.org/10.1016/j.cmpb.2014.09.005
36. K. Adhikari, H. Bouchachia, H. Nait-Charif, Activity recognition for indoor fall detection using convolutional neural network, in *2017 Fifteenth IAPR International Conference on Machine Vision Applications*, MVA, (2017), 81–84. http://doi.org/10.23919/mva.2017.7986795
37. 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** (2019), 988. https://pubmed.ncbi.nlm.nih.gov/31035377
38. H. Yhdego, J. Li, S. Morrison, M. Audette, C. Paolini, M. Sarkar, et al., Towards musculoskeletal simulation-aware fall injury mitigation: transfer learning with deep CNN for fall detection, in *2019 Spring Simulation Conference (SpringSim)*, (2019), 1–12. http://doi.org/10.22360/springsim.2019.msm.015
39. H. Sadreazami, M. Bolic, S. Rajan, Fall detection using standoff radar-based sensing and deep convolutional neural network, *IEEE Trans. Circuits Syst. II Express Briefs*, **67** (2020), 197–201. http://doi.org/10.1109/tcsii.2019.2904498
40. A. Núñez-Marcos, G. Azkune, I. Arganda-Carreras, Vision-based fall detection with convolutional neural networks, *Wireless Commun. Mobile Comput.*, **2017** (2017). https://doi.org/10.1155/2017/9474806
41. S. Chhetri, A. Alsadoon, T. Al-Dala'in, P. W. C. Prasad, T. A. Rashid, A. Maag, Deep learning for vision-based fall detection system: Enhanced optical dynamic flow, *Comput. Intell.*, **37** (2020), 578–595. http://doi.org/10.1111/coin.12428
42. C. Khraief, F. Benzarti, H. Amiri, Elderly fall detection based on multi-stream deep convolutional networks, *Multimedia Tools Appl.*, **79** (2020), 19537–19560. http://doi.org/10.1007/s11042-020-08812-x
43. N. Lu, Y. Wu, L. Feng, J. Song, Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data, *IEEE J. Biomed. Health Inf.*, **23** (2019), 314–323. http://doi.org/10.1109/jbhi.2018.2808281
44. H. Li, C. Li, Y. Ding, Fall detection based on fused saliency maps, *Multimedia Tools Appl.*, **80** (2020), 1883–1900. http://doi.org/10.1007/s11042-020-09708-6
45. R. K. Meleppat, M. V. Matham, L. K. Seah, Optical frequency domain imaging with a rapidly swept laser in the 1300 nm bio-imaging window, in *International Conference on Optical and Photonic Engineering*, (2015), 721–729. http://doi.org/10.1117/12.2190530

46. K. M. Ratheesh, L. K. Seah, V. M. Murukeshan, Spectral phase-based automatic calibration scheme for swept source-based optical coherence tomography systems, *Phys. Med. Biol.*, **61** (2016), 7652–7663. http://doi.org/10.1088/0031-9155/61/21/7652
47. R. K. Meleppat, M. V. Matham, L. K. Seah, An efficient phase analysis-based wavenumber linearization scheme for swept source optical coherence tomography systems, *Laser Phys. Lett.*, **12** (2015), 055601. http://doi.org/10.1088/1612-2011/12/5/055601
48. R. K. Meleppat, C. R. Fortenbach, Y. Jian, E. S. Martinez, K. Wagner, B. S. Modjtahedi, et al. In Vivo imaging of retinal and choroidal morphology and vascular plexuses of vertebrates using swept-source optical coherence tomography, *Transl. Vision Sci. Technol.*, **11** (2022), 11. https://pubmed.ncbi.nlm.nih.gov/35972433
49. V. M. Murukeshan, L. K. Seah, C. Shearwood, Quantification of biofilm thickness using a swept source based optical coherence tomography system, in *International Conference on Optical and Photonic Engineering*, (2015), 683–688. http://doi.org/10.1117/12.2190106