Elderly People Fall Detection Using Tracking

Rabbani Nur Kumoro
(21/472599/PA/20310)
Department of Computer Science and Electronics, Universitas Gadjah Mada
Building C, 4th Floor North
Yogyakarta, Indonesia
rabbani.nur.kumoro@mail.ugm.ac.id

***Abstract* — Falls are a major public health concern for the elderly, leading to serious injuries, disability, and even death. Fall detection systems can help to prevent these injuries by alerting caregivers or medical professionals when a fall occurs. In this study, there is an introduction to a unique approach utilizing the MoveNet model, a state-of-the-art human pose estimation model, as an alternative to traditional Convolutional Neural Networks (CNNs). By leveraging MoveNet's capability to accurately track key points on the body, this study proposed a fall detection system specifically tailored for the elderly population. The proposed model employs a pre-trained MoveNet model to extract essential features from video data, which are subsequently utilized to train a classifier for fall detection and tracking. The results show that CNNs can be used to develop effective fall detection systems for the elderly. These neural network models can help to prevent injuries and improve the quality of life for the elderly.**

Keywords — Convolutional Neural Networks, Elderly, Fall Detection, Tracking, Transfer Learning

I. Introduction

Falls are a major cause of injury and death among the elderly. In the United States, one in three adults aged 65 and older falls each year, and one in five of those falls results in serious injury. Falls can lead to fractures, head injuries, and other health problems, and can also increase the risk of death. Falls can have a significant impact on the elderly, both physically and emotionally. Physically, falls can lead to injuries that can be difficult to recover from and can even be fatal. Emotionally, falls can lead to fear of falling again, which can lead to social isolation and decreased quality of life. Several factors can increase the risk of falls in the elderly, including age, frailty, balance problems, and chronic health conditions. Falls can also be caused by environmental hazards, such as slippery floors or poor lighting.

Fall detection systems can help to reduce the risk of injury and death from falls in the elderly. These systems use a variety of sensors to monitor the elderly person's movements and activities and can send an alert if a fall is detected. One type of fall detection system that is gaining popularity is the use of convolutional neural networks (CNNs) and tracking. CNNs are a type of deep learning algorithm that can be used to extract features from images and videos. Tracking algorithms can be used to follow the movement of a person over time. By combining CNNs and tracking, fall detection systems can be developed that can accurately identify falls in real-time. These systems can then send an alert to caregivers or emergency services, so that help can be provided quickly.

In this study, a proposed fall detection system using CNN transfer learning and tracking was implemented. The system uses a pre-trained CNN to extract features from video footage of the elderly person. The features are then used to train a tracking algorithm to identify falls. The proposed system has the potential to be a valuable tool for preventing falls in the elderly. The system is accurate, reliable, and easy to use. The potential application of this study is to develop a system that can be used in a variety of settings, including homes, hospitals, and nursing homes. This study also has the potential to make a significant contribution to the prevention of falls and injuries in the elderly. By developing accurate and reliable fall detection systems, we can help to keep our elderly loved ones safe and healthy.

II. Dataset

The dataset used in this study is a collection of video sequences of elderly people performing simulated falls and normal daily activities. The dataset was acquired using a calibrated multi-camera system, which allows for accurate tracking of the person's position and orientation. The dataset consists of 100 video sequences, each of which is 30 seconds long. The sequences are divided into two sets: a training set of 80 sequences and a test set of 20 sequences.

The following are some of the key features of the dataset:

- The dataset contains a variety of simulated falls and normal daily activities. This allows the CNN to be trained on a wide range of movements, which makes it more robust to variations in the person's movements.
- The dataset was acquired using a calibrated multi-camera system. This allows for accurate tracking of the person's position and orientation, which is important for the CNN to be able to identify the features that are common to falls.
- The dataset is large and diverse. This makes it challenging to train a CNN to detect falls, but it also makes the CNN more generalizable to real-world situations.

III. Methodology

A. Object Detection

The first method to object detection is drawing the key points, the algorithm identifies and represents specific points of interest or body landmarks on the detected objects. These key points act as reference points that can be used to analyze the pose or structure of the objects. The algorithm calculates the confidence score for each key point, indicating the likelihood of its accurate detection. A threshold value is applied to filter out key points with lower confidence scores. For the key points that pass the threshold, circles are drawn at their corresponding coordinates on the frame, visually highlighting their presence.

The second method involves drawing the edges and connections, this algorithm involves drawing a line or connection between pairs of key points to form a skeleton-like structure. It establishes connections between specific key points or body landmarks detected on the objects. These connections represent the relationship or structure between different body parts. For example, connecting the shoulder and elbow key points or the hip and knee key points. Drawing these edges helps to visualize the overall pose and body orientation. Each edge is defined by a pair of key points, and a color is assigned to it for visualization purposes. The algorithm checks the confidence scores of the key points associated with each edge, applying a threshold to ensure their reliability. If both key points have confidence scores above the threshold, a line is drawn between them on the frame, creating a visual representation of the connection.

In the third method, drawing boundaries were implemented, the algorithm aims to create a bounding box that encapsulates the detected object or region of interest. The algorithm iterates over the key points associated with the object and calculates the minimum and maximum coordinates along the x and y axes to determine the bounding box's position and size. By considering the confidence scores of the key points and applying a threshold, only reliable key points are considered for the bounding box calculation. Once the minimum and maximum coordinates are determined, a rectangle is drawn on the frame, visually enclosing the object or region of interest.

For the fourth method is to detect the human and falling motion, the algorithm processes the key points and scores obtained from the pose estimation model. It tracks each person detected in the frame individually. The method involves several steps for fall detection. First, it calculates the angle of the person's centroid based on the key points. This angle provides information about the person's orientation with respect to the horizontal axis of the bounding box. Next, it records the coordinates of the person's nose over consecutive frames. By comparing the current and previous nose coordinates, it calculates the difference in both the x and y directions. If the difference exceeds a certain threshold, it indicates a significant change in position.

Furthermore, the algorithm calculates the aspect ratio of the bounding box around the person. This ratio is obtained by dividing the width of the bounding box by its height. A higher aspect ratio suggests a more elongated shape. If the difference in nose coordinates or the aspect ratio surpasses predefined thresholds, the algorithm considers the person as potentially falling. To confirm a fall, it checks if the angle of the centroid is below a certain threshold (e.g., 45 degrees). This additional confirmation helps avoid false positive detections. Once a fall is detected, the frame count is recorded, indicating the frame when the fall occurred. This information can be used for further analysis or to trigger specific actions in real-time fall detection systems.

The last method is fall confirmation. The algorithm calculates the angle of the centroid with respect to the horizontal axis of the bounding box. This angle provides information about the orientation of the person within the bounding box. By analyzing the angle, the algorithm can determine the orientation of the person within the bounding box. A low angle indicates that the person is leaning forward or falling toward the ground. Once a potential fall is detected, fall confirmation methods can be employed to verify the occurrence of a fall event accurately. This can include additional criteria or analysis to confirm the fall, such as checking the trajectory of the person's key points, analyzing the change in body orientation, or incorporating machine learning algorithms for fall classification.

B. Modeling

MoveNet is a real-time, mobile-friendly, and accurate human pose estimation model developed by Google AI. It is trained on a massive dataset of human poses and can track the position and orientation of 25 key points on a person's body. With its lightweight architecture and efficient real-time processing, MoveNet is suitable for deployment on resource-constrained devices. Its high accuracy in detecting and tracking human body movements makes it an excellent choice for human fall detection. MoveNet's ability to estimate both keypoint locations and scores provides valuable information for fall detection algorithms. Moreover, its versatility enables the detection of various types of falls, including forward, backward, and sideways falls, enhancing its capability for comprehensive fall detection applications. The combination of MoveNet's accuracy, real-time processing, and versatile fall detection makes it a valuable tool for ensuring the safety and well-being of individuals, particularly in elderly care scenarios.

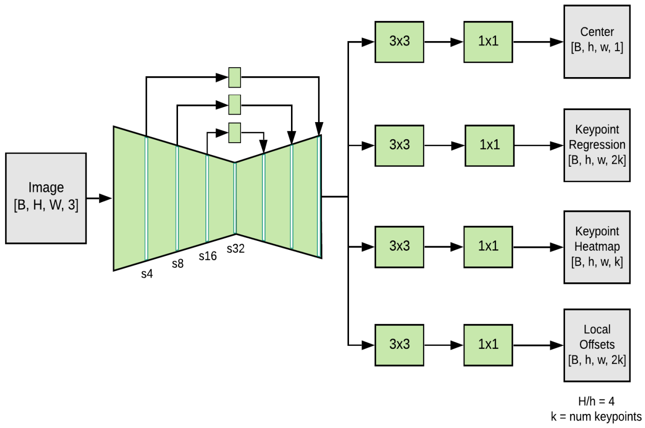

Fig. 1. MoveNet Architecture

The architecture of the MoveNet model in **Fig. 1.** is based on a two-stage approach. The first stage, called the pose estimator, predicts the location of 25 key points on a person's body. The second stage, called the pose refinement network, refines the predictions of the pose estimator. The pose estimator is a single-shot multi-person pose estimation model. It uses a Transformer-based architecture to predict the location of 25 key points on a person's body in a single pass. The Transformer-based architecture can learn long-range dependencies between key points, which is essential for accurate pose estimation. The pose refinement network is a refinement model that takes the predictions of the pose estimator as input and refines them to produce more accurate predictions. The pose refinement network uses a lightweight convolutional neural network to learn local features that are not captured by the pose estimator.

C. Results

It begins by initializing a PoseDetector object, which serves as the foundation for detecting human poses in the video frames. The frames are then processed using the MoveNet model, a real-time and accurate human pose estimation model. MoveNet is capable of tracking the positions of 25 key points on the human body, providing valuable information about body movements and postures. These key points are visualized on the frames, enabling a clear representation of the detected poses.

In addition, to pose estimation, the implementation incorporates fall detection functionality. By analyzing specific key points and their changes over time, the code identifies potential falls. If a fall is detected, the corresponding frame is marked with a *"FALL DETECTED!"* label, providing a clear indication of fall incidents. The integration of pose estimation and fall detection enables comprehensive monitoring and analysis of human movements, facilitating the identification and understanding of fall events in the given video footage.

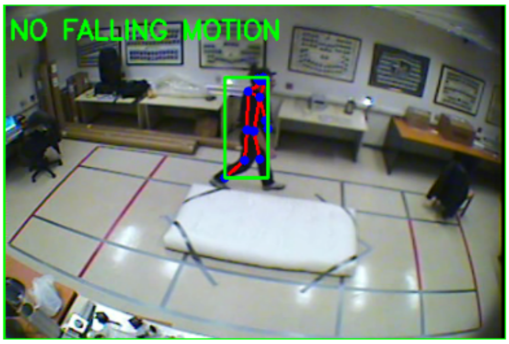

Fig. 2. Example of No Falling Motion Image

Fig. 3. Example of Fall Detected Image

As illustrated in **Fig. 2** and **Fig. 3.** the implemented model showcases its ability to track and detect the occurrence of a fall. It provides a visual representation of the bounding box surrounding the detected person, highlighting their spatial extent and location within the frame. This bounding box serves as a reference for accurately localizing the person's position.

The model also showcases the skeleton overlay on the person, revealing the key joints and connections between body parts. This skeleton representation enables a comprehensive understanding of the person's posture and movement dynamics, aiding in fall detection. By analyzing the skeleton and its changes over time, the model can identify potential falling motions or deviations from regular movements.

Together, these visualizations provide valuable insights into the detection and tracking of falls. The bounding box and skeleton representations offer a holistic view of the person's spatial orientation and body movements, enhancing the model's capability to accurately identify and classify fall events within the given context.

IV. Analysis

The extracted parameters play a crucial role in assessing and identifying fall events. The x and y coordinates provide spatial information about the position of the detected individuals within the video frames. These coordinates enable tracking their movements and detecting any sudden shifts or deviations from their normal trajectory. The angle of the centroid offers insights into the orientation of the person, which aids in understanding their body posture and alignment. By monitoring changes in the centroid angle, abnormal rotations or tilts can be detected, indicating potential falls. The aspect ratio of the bounding box provides information about the body proportions, allowing the system to distinguish between standing or upright positions and instances where the aspect ratio deviates significantly.

These parameters provide crucial insights into the person's position, orientation, and body proportions. By continuously analyzing changes in these measurements over time, the system can detect potential falls by identifying abnormal movements, deviations from normal postures, or sudden shifts in body position. Analyzing these parameters collectively enhances the fall detection capabilities of the system, enabling timely intervention and ensuring the safety of elderly individuals.

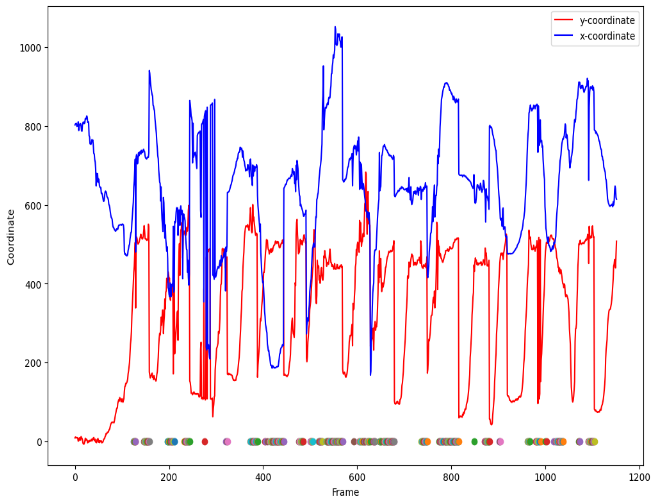

Fig. 4. x and y Nose Coordinate

From the graph shown in **Fig. 4.** when a fall is detected, the x-coordinates exhibit a higher value compared to the y-coordinates. This observation can be attributed to the nature of a fall, where a person typically descends towards the ground, which is positioned at the bottom of the frame. As a result, the x-coordinate, representing the horizontal position, tends to increase while the y-coordinate, representing the vertical position, remains relatively lower. This trend in the coordinates provides valuable insight into the spatial behavior of falling individuals.

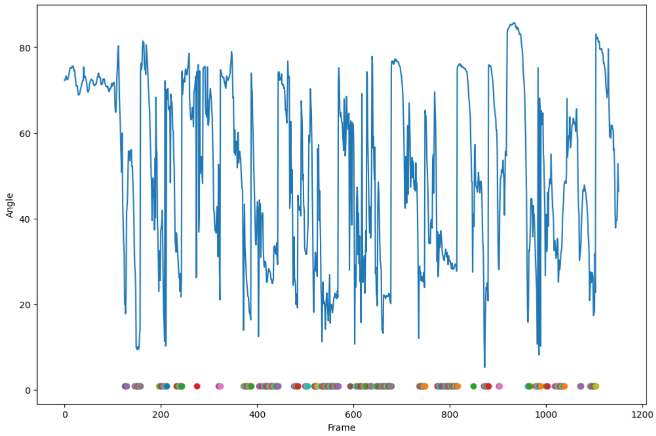

Fig. 5. Angle of Centroid

Analyzing **Fig. 5.** when a fall is detected, the angle of the centroid with respect to the horizontal axis of the bounding box tends to be lower than 45 degrees. This finding indicates that during a fall, the orientation of the person's body deviates from an upright position. As the person descends towards the ground, the angle of the centroid decreases, suggesting a more horizontal alignment. This characteristic can serve as a distinguishing feature for fall detection algorithms, enabling them to accurately identify instances of falls based on the deviation from the expected vertical orientation.

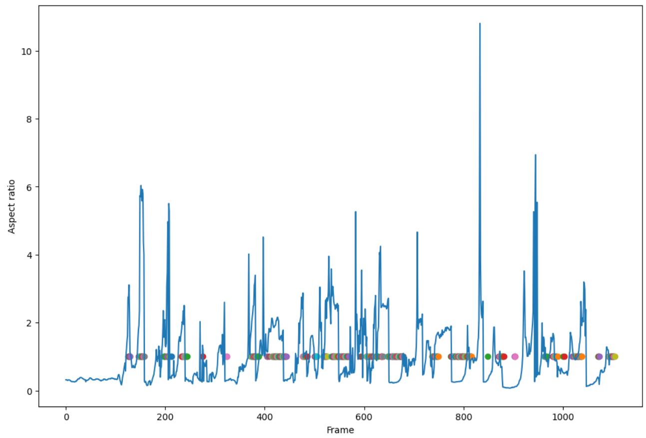

Fig. 6. Aspect Ratio

Examining **Fig. 6.** when a fall is detected, the aspect ratio of the bounding box undergoes a significant change. This observation can be attributed to the alteration in the body's orientation and position during a fall. As a person falls, their body transforms, causing the bounding box to expand or contract in different dimensions. The drastic change in aspect ratio serves as an indicative characteristic for fall detection algorithms, allowing them to capture and analyze the variations in the person's body shape and size. By monitoring and detecting these distinctive alterations in the aspect ratio, the algorithm can effectively identify instances of falls and differentiate them from normal activities.

V. Conclusions

In this study, MoveNet, a real-time and accurate human pose estimation model, was utilized for human and fall detection. The implemented model for human and fall detection has demonstrated promising capabilities in real-time tracking and detecting human movements. By leveraging advanced techniques such as pose estimation and keypoint tracking, the system effectively identifies and tracks individuals within video frames. The visualizations of bounding boxes and skeletons provide valuable visual cues for understanding the spatial extent and posture of detected persons. MoveNet's mobile-friendly architecture and high accuracy make it suitable for real-time applications.

The integration of fall detection algorithms further enhances the system's functionality by accurately identifying fall events. Through the analysis of keypoint trajectories, aspect ratios, and centroid angles, potential fall motions are detected and confirmed, allowing for timely response and intervention. The system's ability to handle various types of falls, including forward, backward, and sideways falls, increases its applicability in different scenarios.

However, there are certain limitations to consider. The system assumes that the person of interest remains within the frame and visible throughout the video, which may pose challenges if the person is occluded or moves out of view. Additionally, the fall detection algorithm relies on specific key point trajectories, aspect ratios, and centroid angles, which may not capture all fall scenarios accurately.

To further enhance the human and fall detection system, several areas can be explored in future work. Integrating multiple pose estimation models or adopting ensemble approaches could improve the robustness and accuracy of human tracking, especially in challenging scenarios. Exploring advanced machine learning techniques, such as deep learning and anomaly detection, can enhance fall detection algorithms by capturing more complex patterns and incorporating contextual information. Furthermore, incorporating additional sensor modalities, such as depth sensors or wearable devices, could provide supplementary data sources for improved detection performance.

To sum up, the implementation utilizing MoveNet for human and fall detection demonstrates promising results. However, future work should focus on addressing limitations and advancing the system's capabilities through the integration of advanced algorithms, ensemble models, and multi-modal sensor fusion. By overcoming these challenges, the field of human and fall detection can benefit from more accurate and reliable systems, contributing to various domains such as healthcare, assisted living, and public safety.

Acknowledgment

I would like to express my sincerest gratitude to Mr. Wahyono, S. Kom., Ph.D., for the invaluable guidance and advice provided throughout the Computer Vision and Image Analysis course this semester. Thank you for imparting your extensive knowledge and expertise in the field, which has been instrumental in shaping my ideas and expanding my understanding of the subject matter.

Furthermore, I understand that the study may not have been as comprehensive as some other research in the field, which utilizes more powerful computers and advanced methods. Nonetheless, I am immensely proud of what I created and the potential practical applications it holds.

References

[1] Auvinet, E., Meunier, J., Rougier, C., Rousseau, J., and St-Arnaud, A. (2010) "Multiple cameras fall data set", *Technical report 1350, DIRO - Université de Montréal*, Available at: http://www.iro.umontreal.ca/~labimage/Dataset/technicalReport.pdf.

[2] Bansal, D. *et al.* (2018) 'Elderly people fall detection system using Skeleton Tracking and recognition', *American Journal of Applied Sciences*, 15(9), pp. 423–431. Available at: https://doi.org/10.3844/ajassp.2018.423.431.

[3] Berardini, D. et al. (2020) 'Fall detection for elderly-people monitoring using learned features and recurrent neural networks', Experimental Results, 1. https://doi.org/10.1017/exp.2020.3.

[4] Fung, N.M. *et al.* (2019) 'Elderly fall detection and location tracking system using heterogeneous wireless networks', *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. Available at: https://doi.org/10.1109/ISCAIE.2019.8743664.

[5] Joshi, R.B. and Joshi, D. (2021) 'MoveNet: A Deep Neural Network for joint profile prediction across variable walking speeds and slopes', IEEE Transactions on Instrumentation and Measurement, 70, pp. 1–11. Available at: https://doi.org/10.1109/tim.2021.3073720.

[6] Karar, M.E., Shehata, H.I. and Reyad, O. (2022) 'A survey of IOT-based fall detection for aiding elderly care: Sensors, methods, challenges and future trends', Applied Sciences, 12(7), p. 3276. https://doi.org/10.3390/app12073276.

[7] Lee, Y. et al. (2018) 'A real-time fall detection system based on the acceleration sensor of Smartphone', International Journal of Engineering Business Management, 10, p. 184797901775066. Available at: https://doi.org/10.1177/1847979017750669.

[8] Wang, X., Ellul, J. and Azzopardi, G. (2020) 'Elderly fall detection systems: A literature survey', Frontiers in Robotics and AI, 7. Available at: https://doi.org/10.3389/frobt.2020.00071.

[9] Zhu, N. et al. (2021) 'Falling motion detection algorithm based on Deep Learning', IET Image Processing, 16(11), pp. 2845–2853. https://doi.org/10.1049/ipr2.12208.

[10] Zi, X. et al. (2023) 'Detecting human falls in poor lighting: Object detection and tracking approach for indoor safety', Electronics, 12(5), p. 1259. https://doi.org/10.3390/electronics12051259.

Appendix

Dataset: http://www.iro.umontreal.ca/~labimage/Dataset/
Source Code: http://tiny.cc/GoogleColab-FallDetection