A Fall Detection Alert System Based on Lightweight Openpose and Spatial-Temporal Graph Convolution Network

Weijian Mai¹, Fengjie Wu²*, Ziqian Guo¹, Yuhan Xiang³, Gensheng Liu¹, and Xiaobin Chen¹

¹ School of Electronics and Communications Engineering, Guangzhou University, Guangzhou, Guangdong, 510006, China
² Laboratory Center of Guangzhou University, Guangzhou University, Guangzhou, Guangdong, 510006, China
³ School of Mathematics and Information Sciences, Guangzhou University, Guangzhou, Guangdong, 510006, China
*Corresponding author’s e-mail: wufengjie501@gzhu.edu.cn

Abstract. To solve the home safety problem of the solitary group, we have developed a system based on Lightweight Openpose and Spatial-Temporal Graph Convolution Network (ST-GCN). We use Lightweight Openpose to extract the human skeleton sequence from the video stream in real-time, and use ST-GCN to extract the action time and space features from it to obtain an efficient fall detection model. For further application, we develop an email alert system using STMP protocol. This work integrates fall detection into the surveillance camera to realize real-time intelligent monitoring and detection at home. When the family member falls, the picture will be captured, and the alarm email will be sent to the owner or guardian, to provide a safer guarantee for the family member, especially the empty-nesters and young people living alone.

1. Introduction
As the aging of society continues to intensify, the proportion of the elderly population has risen, and the number of elderly people at home on their own has increased. The young and middle-aged people are often busy with work and have no time to take care of the elderly who are alone at home. Elderly people who are alone at home often cannot respond in time to illegal intrusions. In addition, the elderly are prone to fall at home due to the decline of their physical functions or their diseases. However, most of the cameras only record but do not recognize, which requires human detection and identification of abnormal conditions. It is hard to avoid negligence while consuming human resources and failing to report abnormal conditions on time.

Abnormal behavior monitoring based on video streams requires a compromise between detection and detection effects and complex calculations. Openpose [1], the current open-source framework for detecting skeletons, has good performance but requires a large amount of calculation, which requires a powerful GPU. The Lightweight Openpose [2] used in this article improves its framework and greatly reduces the amount of calculation. The example proves that the lightweight Openpose can run up to 26 frames in the CPU environment, which can meet the requirements of our system. Based on the extracted human skeleton, we use the ST-GCN [3] for feature extraction and train an efficient fall detection model.

For further application, we developed an email alert system using Simple Mail Transfer Protocol (STMP)
2. Overview

The overview of our system is shown in Figure 1. The system extracts the skeleton of humans from the video stream and performs fall detection. We define the “fall” state as abnormal, and other states as normal. An email alert is triggered when an abnormal state is detected.

3. Methodology

3.1. Fall Detection System

The overview of the fall detection system is shown in Figure 2. First, we use the Lightweight Openpose method to extract skeletons from video streams with multiple frames. Next, the Spatial-Temporal Graph Convolutional Networks (ST-GCNs) are utilized to extract features from the obtained skeleton sequences. Finally, fall detection is realized by classifying whether the action in the video stream is “fall” (abnormal).

3.1.1. Lightweight Openpose for Skeleton Extraction.

Lightweight openpose is an improvement of the original Openpose framework, which reduces the accuracy rate by less than 1% while greatly improving the running speed. This method simplifies the computational complexity of the original framework and enables the 26 fps video stream monitoring program to run on the CPU, which provides great support for the real-time fall detection system.

Similar to all bottom-up approaches, lightweight Openpose pipeline comprises two parts (Figure 3):

- Neural network inference provides two tensors: the key point heat map and its paired relationship (partial affinity field, pafs).
- The key points are grouped by character instances. It includes upsampling the tensor to the original image size, extraction of key points at the peak of the heat map, and grouping them by instance.
3.1.2. Spatial-Temporal Graph Convolutional Network for Feature Learning. By extending the graph convolution network to the spatial-temporal graph model, a general representation of the skeleton sequence for fall detection is designed, which is called the spatial-temporal graph convolution network (ST-GCN). As shown in Figure 2, the model is formulated on a skeletal diagram sequence, where each node corresponds to a joint of the human body. There are two types of edges in the graph, namely, spatial edges that conform to the natural connections of joints (spatial edges) and temporal edges that connect the same joints in successive time steps. On this basis, a multi-layer spatial-temporal graph convolution is constructed, which allows information to be integrated along the two dimensions of space and time. The hierarchical nature of ST-GCN eliminates the need to manually divide parts or traverse rules. This not only can obtain stronger expression ability and higher performance.

3.2. Experiments

3.2.1. Dataset Construction. We utilized lightweight Openpose to extract human skeletons in the video to build the dataset of the ST-GCN model. As shown in Figure 4, the "Abnormal" dataset consists of each frame of skeleton images during the fall. Considering the camera position difference, this dataset includes images from various angles of the falling state, making the system performance more robust. Meanwhile, as shown in Figure 5, the "Normal" dataset contains skeleton images of normal human walking or other non-falling actions.
3.2.2. Model Training. The whole dataset is divided into the training set and test set according to the ratio of 8:2 for further training. The training goal is to optimize the cross-entropy loss function. Furthermore, we utilize Adam [5] with a learning factor of $10^{-4}$ and the first and second decay rates of 0.9 and 0.999, respectively. The batch size was set to 32. Dropout [6] with a probability of 0.2 is adopted to solve the overfitting issue. This system is run on an Intel i7-6850K, 3.6GHz CPU with 26 fps.

3.2.3. Model Evaluation. To apply in different scenarios, the proposed system sets the fall detection threshold with a range in [0,1]. The lower the threshold, the higher the fall detection sensitivity, which is easy to cause false detection (high Recall rate). The higher the threshold, the ratio of false detections will decrease, but it is easy to cause missed testing (high Precision rate). In real life, the threshold can be selected according to the application scenario. In this work, we use the F1 score metric to evaluate the model.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

The result of the precision and recall rate of our model is shown in Figure 6. The F1 score value over training epochs is presented in Figure 7.

3.3. Email Alert System

Simple Mail Transfer Protocol (SMTP) is a protocol that provides reliable and effective email transmission. SMTP is a mail service based on the File Transfer Protocol (FTP) file transfer service. It is mainly used to transfer mail information between systems and provide notifications about incoming letters. As shown in Figure 8, through the server designated by the SMTP protocol, we can immediately
send e-mails to the receiver's server. The receiver can be set as family members or other important persons to be timely aware of the fall phenomenon occurring in the family.

3.4. Results
As shown in Figure 9, when the system detects a fall, it will capture the current image and send it to the designated user's mailbox as an alert. In this way, the system can provide safety protection after falls to groups such as young people living alone or elderly people at home.

4. Conclusions
The proposed system uses a small amount of calculation to achieve higher performance, enabling it to run on the CPU, making it more widely used. This work integrates fall detection into the surveillance camera to realize real-time intelligent monitoring and detection at home. When the family member falls, the picture will be captured, and the alarm email will be sent to the owner or guardian, to provide a safer guarantee for the family member, especially the empty-nesters and young people living alone.
Acknowledgments
This work was supported in part by the Climbing Program of Guangdong Province, China, under Grant pdjh2021a0387, and in part by the National College Student Innovation and Entrepreneurship Training Program, China, under Grant S202111078012.

References
[1] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[2] Osokin, Daniil. "Real-time 2d multi-person pose estimation on CPU: Lightweight OpenPose." arXiv preprint arXiv:1811.12004 (2018).
[3] Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." Thirty-second AAAI conference on artificial intelligence. 2018.
[4] Klensin, John, et al. SMTP service extensions. RFC 2846, November, 1995.
[5] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
[6] Hinton, Geoffrey E., et al. "Improving neural networks by preventing co-adaptation of feature detectors." arXiv preprint arXiv:1207.0580 (2012).