Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 3, No. 3, 2022

Fall Detection Algorithm Based on Lightweight Openpose Model with Attention Mechanism

Ruiming Qiu, Wei Teng*, Zihe Wei, Cong Zhang, Yipeng Zhong and Junhao Zhai

School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China

Abstract: At present, the serious consequences caused by the fall of the elderly emerge one after another, and installing cameras at home to prevent emergencies for the elderly has gradually become a choice for more and more people. However, the traditional Openpose model cannot detect in real time, which is inconsistent with the actual demand. Therefore, this paper first proposes to use mobilenetv2 lightweight network to replace the original huge vgg19 backbone network, which makes the model real-time in actual use, and then integrates the attention mechanism module to improve the accuracy of the model without affecting the real-time.

Keywords: Fall detection, Lightweight Openpose, Attention mechanism.

1. Introduction

According to the International Classification of Diseases, falls can be divided into two categories: falls from one plane to another and falls from the same plane. Globally, falls are an important public health problem and have become the leading cause of unintentional injury death. In our country, every year more than 40 million elderly people fall, which has also become an important hidden trouble that causes the old people's physical problems. Therefore, with the fall detection model and algorithm, the home camera can play a good role in detection and reminder. However, the traditional openpose detection model lacks real-time performance in practical application, so we need to use lightweight backbone network Mobilenetv2 to improve the original model, so as to realize real-time monitoring.

2. Introduction and Shortcomings of Openpose

When it comes to algorithms related to pose estimation, Openpose[2] has a considerable influence in this field.

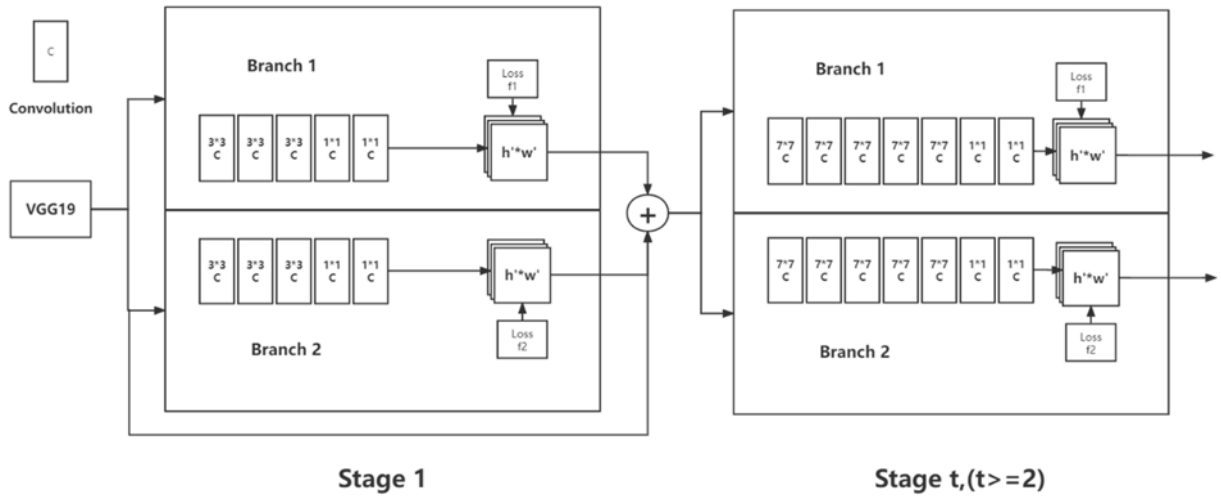

Figure 1. Openpose network structure diagram

After the corresponding features of the picture are extracted through the backbone network, these features will enter into different stage modules, and PCM and PAF in each stage will conduct loss processing and solve. PCM is the thermal map of the key point to represent the position of the key point. PAF is the core of openpose, namely the affinity field of key points, to determine the affinity between different key points.

Makes it impossible for openpose real-time VGG19 is its backbone network, the main reasons for its every layer neural network combined with a layer of output for more complex features extraction, including 16 layer convolution with 3 full connection layer, which makes the feature extraction accuracy while compared with other backbone network has good advantage, However, the complexity and redundancy of the network will be greatly increased accordingly.

The main feature of the VGG network model is to use convolution kernels of 3*3 size to replace the original convolution kernels of larger size and accumulate them to form a neural network. In this way, the depth of the whole neural network can be deepened and the overall performance of the model can be improved. Although the structure is simple, it is difficult to carry out real-time detection and classification because of the large number of network layers.

3. MobileNetV2 Lightweight Network

In order to enable the fall detection task to be carried out in real time and maintain a relatively good detection accuracy, we use the classic MobileNet[3] lightweight network series. Although the accuracy is slightly reduced compared with the vgg network series, the number of parameters and the number of network layers are significantly reduced, making it able to complete the real-time detection task well. In this series, we use the MobileNetV2 lightweight network. The MobileNet v2 network was proposed by the google team in 2018. Compared with the MobileNet V1 network, the accuracy is higher and the model is smaller.

The MobileNetV2 lightweight network has two main highlights:

The first point is that it USES the residual structure, the first use of 1 * 1 size convolution to l d, then by the size of 3 * 3 DW convolution (by channel convolution) to extract the corresponding feature, finally using convolution to achieve dimension reduction, 1 * 1 size relative to the residual structure, it will lift d in the order of exchange, and use the DW convolution instead of 3 * 3 standard convolution, The new activation function ReLU6 is used to replace the original ReLU activation function.

The second point is that for the inverted residual structure in the first point, the linear activation function is used to replace the original ReLU activation function in the last convolutional layer. The reason for the substitution is that ReLU activation function will cause large instantaneous loss for low-dimensional data feature information, while it will cause small instantaneous loss for high-dimensional data feature information. Because of the spindle structure of the inverted residual itself, the final output information is a low-dimensional feature information, so choosing the linear activation function can reduce the feature loss of the output feature information as much as possible.

Table 1. Network structure of MobileNetV2

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

As shown in Table 1, t is the expansion factor and the expansion multiplier of the convolution kernel in the first 1x1 convolution layer. c is the depth channel of the output characteristic matrix, n is the number of repetitions and s is the step distance (for the first layer, others are 1, similar to ResNet, and the size changes by the step length of the first layer). It can be seen that the network structure is very lightweight.

By replacing the original vgg19 network model with MobileNet v2, the number of parameters is reduced, which makes the real-time goal of the detection task be achieved while keeping the accuracy of the detection task excellent.

4. Mechanism of Attention

Although real-time detection can be achieved by using the lightweight MobileNetV2 network, the recognition accuracy is taken as the corresponding cost. Although the decrease of the accuracy is not much, we still need to add some small modules to make up for it. In this paper, attention mechanism[4] is added to process the recognition features. Thus, the estimation accuracy of the model can be effectively improved.

Attention mechanism is a commonly used technique in the field of deep learning. It can make the corresponding network pay attention to the places that should be paid more attention, so as to deepen the main features and ignore the secondary features, thus making the processing more efficient and effective. This method does not need manual adjustment, which is a way to achieve network adaptation. It comes in many forms, but the core is attention.

The attention mechanism adopted in this paper is implemented by SENet, which focuses on obtaining the input feature values and the processed channel weights. Its specific implementation is shown in Figure 2:

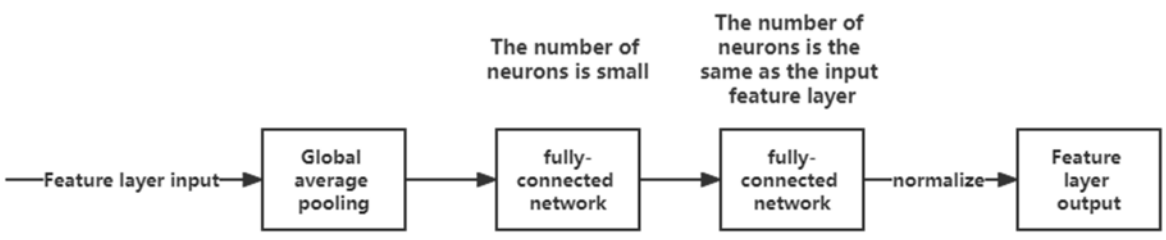

Figure 2. Specific implementation of SENet

(1) Global average pooling is performed on the input feature layer.

(2) Two full connections were made, the number of neurons in the first time was small, and the number of neurons in the second time was the same as that in the feature layer.

(3) Normalize, take Sigmoid and fix its value between 0 and 1 to obtain the weight of each channel in the input feature layer.

(4) After the weights are obtained, the output of the feature layer is obtained by multiplying the weights with the original input feature layer.

In order to make the feature more effective, SENet attention block is placed after each convolution block of MobileNetV2. After the feature layer of the upper convolutional network is processed by the attention block, the feature is fused and then transferred to the lower network. This makes the features extracted by the network more representative, has more advantages in the extraction of key points, and improves the success rate of project fall detection.

5. Algorithm Design

Based on the characteristics of Openpose, we judge fall detection by comparing before and after frames.

First of all, we pass the video stream to the lightweight openpose model. Through the position recognition and judgment of different joints of the human body, the coordinates of different joint nodes are obtained and written into the corresponding frame json file. After writing, we first initialize the counter count=0 and calculate the ordinate change of head and body between every two frames. If the ordinate difference of the previous frame minus the next frame between head and body is greater than the threshold, it means that the interval of these two frames meets the condition of falling, count+1. If more than four consecutive intervals meet the condition of falling, that is, count>4, it is judged as falling. Otherwise, re-initialize count to 0 and continue the calculation. The purpose of making the four intervals conform to the conditions is to prevent misjudgment of the behavior of non-fall but rapid decline in ordinate, so as to improve the accuracy of fall detection. The specific process is shown in Figure 3.

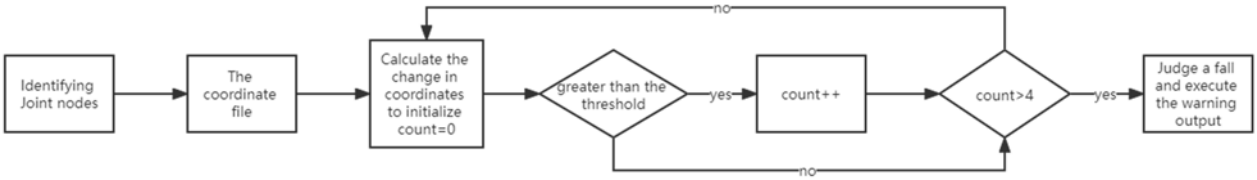

Figure 3. Fall detection algorithm design

6. Experimental Application and Results

In this experiment, a total of 84 videos were collected on the network, and 9 videos were shot by myself for testing from different angles, and a total of 93 videos were verified. Among them, 84 videos were recognized well under the model and algorithm, with an accuracy rate of 90.32%. The model and algorithm mentioned in this paper can have a good detection effect on the state of falling in daily life. The actual detection effect is shown in Figure 4, the left figure is the non-fall state, and the right figure is the detected fall state.

Figure 4. Demonstration of Fall Detection

7. Epilogue

In this paper, the original openpose model is improved by MobileNetV2 lightweight network, so that the model has real-time fall detection, and on this basis, attention mechanism is integrated to re-fuse features and improve the accuracy of actual detection. In the future, the model algorithm will be considered to adapt to the home camera, so as to realize the fall detection and recognition in practical applications.

Acknowledgment

College Students' Innovation and Entrepreneurship Training Plan of University of Science and Technology Liaoning in 2022.

References

[1] Liu Xiaohong, Wu Miao, Niu Qian:Risk Factors of Falls in the Elderly, Beijing Med, Vol. 43 (2021) No.6,p.533-534+538.

[2] Zhang Zezheng, Wang Jun, Dong Mingli, Wang Lei, Yan Bixi: Rapid Key Point Detection Method for Humanoid Robot Based on Improved OpenPose, Laser Journal, [2022-10-18] ,p.1-7.

[3] Yang Ye, Jie Qiang Zhang: Research on Maize Disease Recognition based on Lightweight network MobileNetV2, Modern Computer, Vol. 28 (2022) No.11,p.46-50.

[4] Ji Guangkai, Wang Rong, Peng Shufan: Pedestrian re-recognition method based on attention mechanism and conditional convolution, Journal of Beijing University of Aeronautics and Astronautics , [2022-10-18],p.1-10.