IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Sitting Posture Recognition Based on OpenPose

To cite this article: Kehan Chen 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **677** 032057

IOP Conf. Series: Materials Science and Engineering **677** (2019) 032057 doi:10.1088/1757-899X/677/3/032057

Sitting Posture Recognition Based on OpenPose

Kehan Chen

Shanghai Jiaotong University, Shanghai, China

hzhc_wjdi@163.com

Abstract. Sedentary and poor sitting posture can damage the health of adolescents. Therefore, it is very practical to effectively detect the sitting posture of students in the classroom and to warn the bad sitting posture. This paper proposed an in-class student sitting posture recognition system based on OpenPose, which uses the monitor in the classroom to detect the sitting posture of the students, and uses OpenPose to extract the posture feature. Keras deep learning framework is used to construct the convolutional neural network, which is used to train the datasets and recognize sitting posture of students. Experiments show that the accuracy is more than 90% after 100 epoch training.

1. Introduction

Nowadays, cervical spondylosis and other symptoms of lumbar spondylosis have been younger and younger [1]. The correct sitting posture is an important means to prevent such diseases. It is very important for children to develop a good sitting posture from a young age for future growth. At present, almost every classroom has monitoring, which can identify the student's sitting posture by monitoring the captured video, and feedback the student situation of the bad sitting posture to the inspector, and then use it to correct the sitting posture of the student.

The research of sitting posture recognition technology belongs to a branch of human body posture estimation research, and human body gesture recognition is also a hot spot in recent years. Before 2015, the coordinates of the joints were directly returned [2], then the literature [3] proposed CPM method which uses a sequential convolutional architecture to express spatial information and texture information. The literature [4] proposed an hourglass-type network structure and most of the single-person pose estimation algorithms that appeared after 2016 are based on this model structure, such as the literature [5] and the literature [6]. Multi-person pose estimation methods are generally divided into two categories, in which Top-down is to frame people first, and then use a single method to locate human joints [7]. The Bottom-up method first makes all the joint positions out, and then distinguishes who the joint belongs to [8].

The sitting recognition method can be divided into two types: a sensor-based method and an image-based method. The literature [9] uses the Astra3D sensor to perform depth image acquisition on the sitting posture of the human body. Literature [10] proposed a sitting posture recognition system based on Kinect sensor. The literature [11] installed a pressure sensor on the seat to detect sitting postures by pressure detection. References [12] and [13] used the pressure sensor and the inertial sensor to achieve the sitting position detection. Although these traditional methods have the advantage of high accuracy of measurement data, they also have limitations such as single data, inconvenient use, and high cost, and all require additional hardware sensor support. Therefore, computer vision-based sitting posture estimation can be a good solution.

IOP Conf. Series: Materials Science and Engineering **677** (2019) 032057 doi:10.1088/1757-899X/677/3/032057

In this paper, OpenPose is used to extract the sitting posture characteristics of the students from the classroom monitor. The extracted feature maps marked with the student bone nodes are used as the training set. Then use the Keras deep learning framework to build a deep learning network, and use the dataset to train the network, so that the model can be used to judge the sitting posture form the real-time monitoring screen.

2. Constructing Datasets

One tool for extracting human posture is OpenPose. The OpenPose Human Body Attitude Recognition Project is an open source library developed by Carnegie Mellon University (CMU) based on convolutional neural networks and supervised learning and developed in the framework of caffe [14].

Attitude estimation such as human motion, facial expression, and finger movement can be achieved. Suitable for single and multi-person, with excellent robustness [15]. OpenPose provides a bottom-up approach to real-time estimation of multi-person gestures without the need for any character detectors.

Loading the OpenPose pre-trained model, the algorithm will extract 18 body joint and 17 lines connecting joints. Figure 1 and Figure 2 shows a partial datasets after extracting joint point information features using OpenPose.

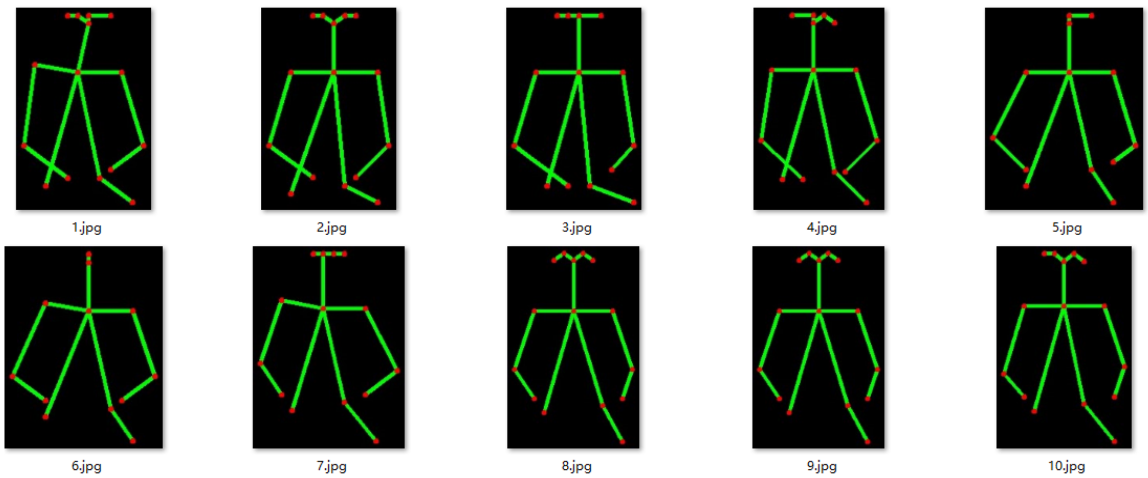

Figure 1. Dataset of Correct Sitting Posture.

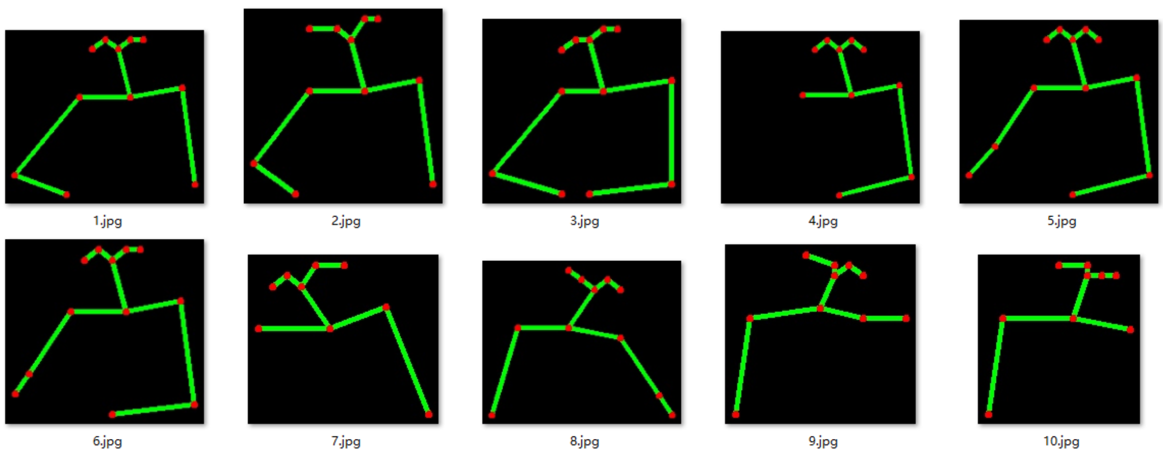

Figure 2. Dataset of Incorrect Sitting Posture.

3. Proposed algorithm

Since the change of body posture will result in changes in positions of body joints, only the graph information within the range between the minimum and maximum coordinates of x and y will be preserved when the dataset is constructed. The background is set black and the redundant data is diminished.

The whole dataset is processed before being applied to the training process. The procedure is listed below.

(a) Cut all images to a size of (60, 60) pixels, of which the edge parts are filled with black pixels.

(b) Label the images of correct sitting postures with “0” and those of incorrect sitting postures with “1”.

(c) Apply the data enhancement to expand the dataset. Due to the limited experimental conditions, the dataset constructed is not big enough. So, the data enhancement is necessary for expanding the dataset used for training.

(d) Normalize the image so that pixels are within the range of 0 to 1.

CNN model is composed of 19 layers, the input layer is for inputting the processed images with (64, 64) pixels. The output layer is the Classification layer, using the SoftMax Classifier. The Activation layer applies ReLU function, and the Pooling layer has a stride of (2, 2).

The SGD + momentum Optimizer is used for training. The categorical_crossentropy is adopted as the loss function, so, it is necessary to vectorize the classification labels with One-Hot Encoding based on the number of classifications. There are only two categories so the labels are turned to 2-dimensional.

The range of random rotation of images during the data enhancement is set to 20. The range of horizontal shift of images is set to 0.2 (the ratio of width before and after the shift) and images can do horizontal reversal randomly.The dataset is categorized as training set, verification set and testing set, which takes up 80%, 10% and 10% respectively. The constructed dataset is then used to train the model to optimize the parameters and is stored in TensorBoard daily record.

4. Result Analysis

The number of epochs is chosen to be 100. Figure 3 shows the value of the accuracy of each iteration recorded with the number of iterations. It can be seen from the figure that the accuracy of the training is very low at the beginning, and the accuracy increases rapidly as the number of iterations increases. When iterating to about 100 times, the accuracy reached a maximum, although there was some fluctuation, but it remained basically above 90%.

Figure 3. Curve of training accuracy.

The loss curve shown in Figure 4 is recorded as the loss value with each iteration. In this paper, the cross entropy loss function is selected as the index for judging the network performance. It can be seen from the figure that when the network is just beginning to train, the loss value is large, and as the number of iterations increases, the loss value decreases rapidly. At the iteration of 100 times, the loss value is minimized and remains stable.

Figure 4. Curve of training loss.

The real-time video image is acquired by using the deep learning network model completed by training, and the pose characteristics of each frame are extracted, and then the correctness of the sitting posture is judged. The recognition results are printed in the video interface, as shown in Figures 5 and 6, each of picture is a screenshot of the video display interface.

Figure 5 shows the recognition results for a set of correct sitting postures, including four images of different sitting postures. From the recognition results, OpenPose can effectively extract the joint information of the students regardless of the table occlusion, especially the joint point information of the upper body that can be used for sitting posture recognition.

Figure 6 shows the recognition results for a set of incorrect sitting postures. As can be seen from the figure, when the student's cervical spine or body is tilted, it is considered to be the wrong sitting posture. Although some joint points are missing, this does not affect the correct recognition results.

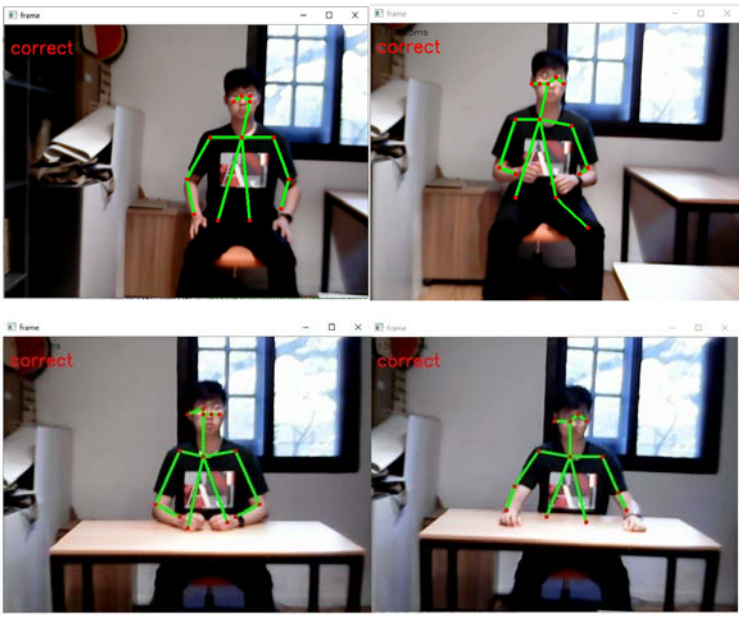

Figure 5. Correct sitting posture recognition result.

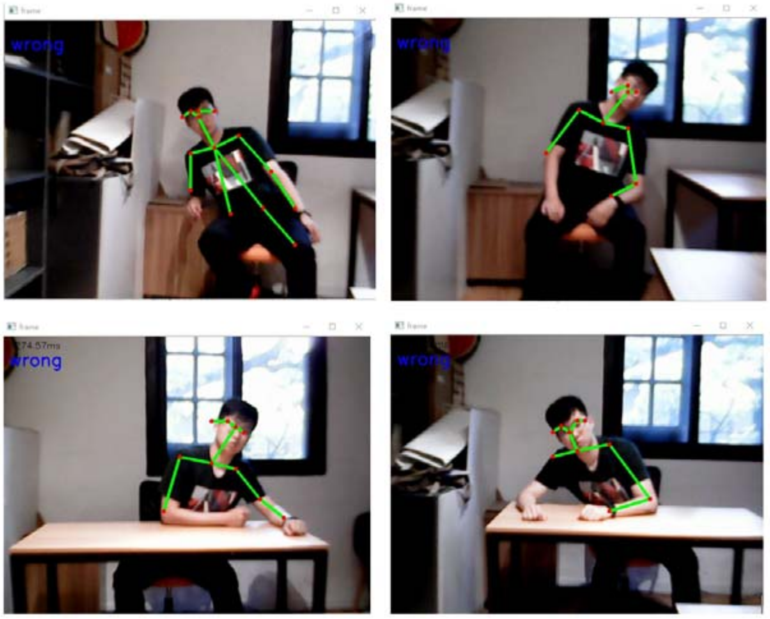

Figure 6. Wrong sitting posture recognition result.

5. Conclusion

This paper developed a computer vision-based classroom student sitting recognition system for finding and correcting students' bad posture. The algorithm uses OpenCV to extract the video information captured by the monitor. Then OpenPose is used to extract the posture feature of students. A convolutional neural network was built using Keras deep learning framework to train data sets. Finally, the trained network model is used to identify the student's sitting posture. Experiments show that the accuracy of the test set and the verification set is up to 90% through 100 epoch training. After actual tests, it can effectively identify the sitting posture of students in the classroom, help young people develop good sitting habits and promote their healthy growth.

References

[1] Xie H L. Analysis of prevalence rate and related factors of adolescent myopia [J]. Chinese Medical Journal. 2010.
[2] Toshev A, Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks [J]. 2013.
[3] Wei S E, Ramakrishna V, Kanade T, et al. Convolutional Pose Machines [C]// CVPR. IEEE, 2016.
[4] Newell A, Yang K, Deng J. Stacked Hourglass Networks for Human Pose Estimation [J]. 2016.
[5] Yang W, Li S, Ouyang W, et al. Learning Feature Pyramids for Human Pose Estimation [J]. 2017.
[6] Ke L, Chang M C, Qi H, et al. Multi-Scale Structure-Aware Network for Human Pose Estimation [J]. 2018.
[7] Fang H, Xie S, Tai Y, et al. RMPE: Regional Multi-person Pose Estimation [J]. 2016.
[8] Cao Z, Simon T, Wei S E, et al. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [J]. 2016.
[9] Xing Z, Bei S, Wu-Sheng L, et al. Sitting Posture Detection System Based on Depth Sensor [J]. Computer Science, 2018.
[10] Li S S. Design and Implementation of Sitting Position Recognition Software Based on Kinect Sensor [D]. University of Electronic Science and Technology, 2018.
[11] Liu S H. Design and Implementation of Sitting Position Monitoring and Reminding System Based on Flexible Array Pressure Sensor [D]. Anhui University, 2018.
[12] Zhu W P, Yin S S, Liu G L. Cervical spondylosis prevention system based on pressure sensor and camera [J]. Computer system and application, 2017.

[13] Liu X. Design and Implementation of Sitting Position Detection System Based on Inertial Sensor [D]. Harbin University of Science and Technology, 2017.
[14] Cao Z, Simon T, Wei S E, et al. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [J]. 2016.
[15] Qiao S, Wang Y, Jian L. Real-time human gesture grading based on OpenPose [C]// 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2018.