Journal of Engineering Mechanics and Machinery (2025)
Clausius Scientific Press, Canada
DOI: 10.23977/jemm.2025.100203
ISSN 2371-9133 Vol. 10 Num. 2

Development of A Sitting Posture Health Detection System Based on the Centernet Model

Laihong Zhu[1,a,*], Yihe Zhu[2,b]

[1]*Haojing College of Shanxi University of Science and Technology, Xi'an, Shaanxi, 712046, China*
[2]*School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi, 710129, China*
[a]*kdhjkjc@sust.edu.cn*, [b]*zhuyihe2007@nwpu.edu.cn*

Keywords: Sitting Posture Health; Centernet Model; Fully Connected Neural Network Model

Abstract: At present, health problems caused by poor sitting posture have attracted much attention, especially affecting specific groups such as students and office workers not only causing physical health problems such as spinal curvature and cervical pain, but also triggering psychological problems such as anxiety and fatigue. Compared with traditional sitting posture detection methods, this paper proposes non-contact sitting posture detection system based on machine vision, which can capture sitting posture information in real time, accurately and conveniently, and helps to improve sitting posture habits and prevent health. Based on the lightweight human pose estimation model MoveNet under the CenterNet model, this model classifies the pose information (the coordinates of the 17 key points of the body) output by MoveNet to judge which sitting posture state the person in the picture is in. The application of this technology can help people correct bad sitting habits in time, reducing the occurrence of problems such as myopia, spinal diseases, and muscle stiffness and fatigue, and improving physical health. This system verifies the feasibility, stability, and accuracy of sitting posture detection system based on the CenterNet model. The test results show that this system can recognize the user's sitting posture state in real time and accurately, and give corresponding and suggestions, which improves the user's physical and mental health problems, and provides a solution for the future sitting posture health monitoring field.

1. Introduction

Since the 17th century, researchers have been studying the posture of the human body mainly through methods such as naked eye observation, manual measurement, and manual questionnaire surveys. In developed countries in Europe and America, due to the generally high living standards and high degrees industrialization and informatization, health problems related to poor sitting posture are relatively common. Therefore, these countries started research on sitting posture relatively early and accumulated some results. According to research of the World Health Organization on different age groups in many countries around the world, millions of cases of occupational skeletal diseases (WMSDs) occur every year due to-term poor sitting posture [1-4].

Unhealthy sitting posture has become one of the important factors affecting people's physical and mental health. This bad habit first poses a threat to vision and exacerbate the risk of myopia [5]. In addition, uneven lumbar pressure can also lead to scoliosis, further affecting the normal function of the respiratory and digestive systems. taining poor sitting posture for a long time can also lead to thoracic deformation [6], especially in childhood, this sitting posture can have a bad effect on the development of back muscles causing muscle damage [7-9], spasms and other problems, and long-term accumulation may lead to symptoms such as dizziness and blurred vision. Over time, this poor sitting posture will affect the health of the cervical spine and increase the risk of intervertebral disc lesions and cervical diseases in adulthood [10]. In daily learning and work, students and office workers face for a long time and are prone to form unscientific sitting postures, and many people maintain this poor sitting posture for a long time because they feel more comfortable. In fact, kind of sitting posture unconsciously changes the way muscles exert force, making the body unevenly stressed, and long-term will inevitably affect bone growth and posture. Therefore, we should importance to the correctness of sitting posture and protect our physical and mental health. With the rapid development of computer and other technologies, the methods used by researchers in the study of sitting posture also achieved a qualitative leap. As an important branch of the computer field, the human motion posture detection and analysis dominated by machine vision technology has always been a research hotspot in field of pattern recognition and computer vision [11]. The essence of using machine vision technology for human motion posture detection and analysis is: through a certain amount of training and learning, the computer detect, track and classify the moving targets in the video image frames, so that the computer can recognize and identify the targets in the external environment, just like humans. Until today, vision technology has great application space in intelligent monitoring field, human-computer interaction field, motion analysis field, and virtual reality field, etc. Therefore, machine vision technology is of quite research significance. Through the study of lightweight posture estimation models based on computer vision, such as CenterNet algorithm model, it can be realized to monitor and recognize human posture through mobile software The application of this technology can help people correct bad sitting habits in time, thereby reducing the occurrence of myopia, spinal diseases, muscle stiffness and fatigue, and improving physical health.

2. Key Technology Analysis

2.1 CenterNet Model

The design of the CenterNet model is exquisite and efficient, with its core including a resnet50 feature extractor, a deconvolution module Deconv, and three branch convolutional networks. Firstly, resnet50, as the feature extractor, responsible for capturing abundant semantic information from the input image. Subsequently, this feature information is fed into the deconvolution module Deconv, which consists of three deconvolution groups, each a 3x3 convolutional layer and a deconvolution layer. In the deconvolution module, each deconvolution group plays a role in gradually doubling the size of the feature map, that the network can capture more fine-grained spatial information [12]. It is worth noting that to enhance the model's fitting ability, some advanced implementations replace the 3x3 convolutional before the deconvolution with DCNv2 (Deformable Convolutional Networks Version 2). DCNv2 allows the convolution kernel to undergo spatial deformation, better to the geometric transformation of the target and enhancing the model's adaptability to complex scenes. After processing by the deconvolution module, the feature map is sent into three branch networks. These three branches are responsible for predicting the heatmap, the width and height of the target, and the coordinates of the target's center point, respectively. By integrating information, CenterNet can achieve accurate detection and localization of targets in images. The overall structure of the

CenterNet network is shown in Figure 1.

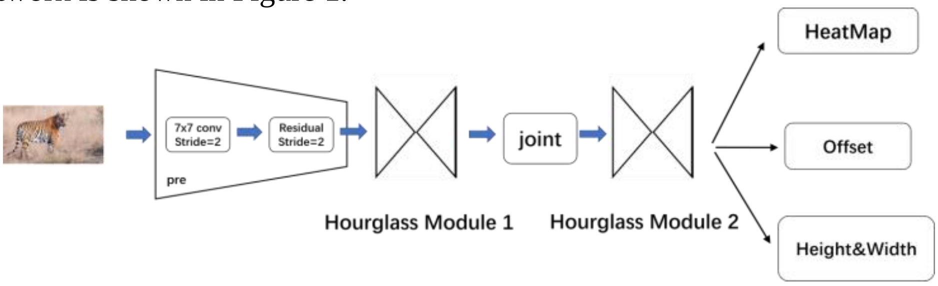

Figure 1 The overall structure diagram of CenterNet network

2.2 MoveNet Model

The MoveNet model, meticulously crafted by Google in 2021, is an efficient lightweight human pose estimation model. It borrows the core idea from CenterNet, adopting a bottom-up detection strategy that accurately captures the positions of 17 key points on human body, including the nose, both eyes, both ears, both shoulders, both elbows, both wrists, both hips, both knees, and both ankles. Once, the MoveNet model can be easily deployed to provide users with a smooth and precise experience. The MoveNet model excels in both speed and accuracy, capable of operating at a of over 50 frames per second (fps) on laptops and mobile devices, thus achieving real-time detection of human poses. This speed is sufficient for the majority of applications whether in daily life or in professional fields. To cater to different scenarios, the MoveNet model provides two variants: Lightning and Thunder. The Lightning version is particularly suitable for applications with latency requirements, such as real-time interaction and gaming; while the Thunder version, which prioritizes accuracy, is ideal for scenarios that demand precise pose analysis. In practical applications, Thunder model typically scores slightly higher in key point detection than the Lightning model, demonstrating its superior performance.

2.3 The fully connected neural network (DNN) model

The fully connected neural network, which should be called fully connected feedforward neural network fully connected (Fully Connect) refers to the fact that each neuron in the next layer is connected to each neuron in the previous layer; while feedforward (Feedforward) refers the fact that the signal flow is unidirectional after the input signal enters the network, and there is no feedback in between. Inputs —> multiplied by weights and then added the bias (the ones on the line are weights, and the ones in the green box are biases) —> activation function —> input to the next layer. Its structural diagram shown in Figure 2.

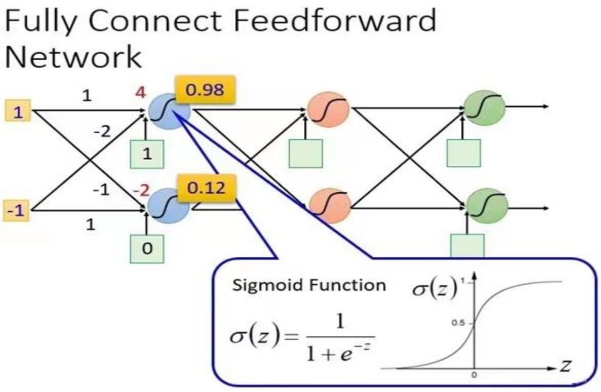

Figure 2 The fully connected neural network structural diagram

Chain rule is almost like, when we go backwards, we can layer by layer compute the gradient of the weights in front with respect to the final loss function. The formula shows below:

Case 1 $y = g(x) \quad z = h(y)$

$$\Delta x \rightarrow \Delta y \rightarrow \Delta z \qquad \frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

Case 2

$$x = g(s) \qquad y = h(s) \qquad z = k(x, y)$$

$$\begin{tikzcd} & \Delta x \arrow[dr] & \\ \Delta s \arrow[ur] \arrow[dr] & & \Delta z \\ & \Delta y \arrow[ur] & \end{tikzcd} \qquad \frac{dz}{ds} = \frac{\partial z}{\partial x}\frac{dx}{ds} + \frac{\partial z}{\partial y}\frac{dy}{ds}$$

The loss function is defined on a single training example, such as in a classification problem, it is the difference between the predicted and the actual class. Total loss function is defined over the entire training set, the sum of all the errors. It is also the value that backpropagation needs to minimize. The formula shows below:

$$L(\theta) = \sum_{n=1}^{N} C^n(\theta) \Rightarrow \frac{\partial L(\theta)}{\partial w} = \sum_{n=1}^{N} \frac{\partial C^n(\theta)}{\partial w}$$

3. System Design

3.1 Data Collection and Preprocessing

Data collection and preprocessing are crucial steps in building a deep learning model directly affecting the training effectiveness and real-time detection accuracy.

(1) Data Collection

Data collection is the primary task in building a deep learning model, aiming to obtain large, diverse, and accurately annotated dataset. For tasks such as object detection and human pose estimation, image or video data containing target objects and human poses need to be collected. These come from public datasets, internet resources, and actual shooting scenarios. When collecting data, to ensure the diversity and balance of the dataset, images are collected from different groups of people different scenes, different angles, and different lighting conditions to improve the model's generalization ability [13]. The number of samples of each category in the dataset should be relatively balanced to avoid the's bias towards a certain type of sample.

(2) Data Preprocessing

Data preprocessing is an important step before training a deep learning model, which aims to improve quality and reduce the difficulty of model training by cleaning and formatting the raw data. Data cleaning: By manually screening, noise, duplicates, or invalid samples in the dataset are to ensure data accuracy and validity. Data formatting: Image data is converted into tensor format that the model can process, and annotation information is converted into labels, vectors, and forms that the model can understand. The raw data is converted into the format needed for model training.

3.2 Selection of Models and Training Methods

For tasks such as object detection and human pose estimation, system will select appropriate models and training methods. For the MoveNet model, this system will utilize its excellent pose estimation capabilities and adjust the network structure and parameters to adapt to the specific of the scenario. For the design of fully connected neural networks (DNNs), this graduation project will construct a suitable network structure according to the characteristics of the task. During the training process, a suitable dataset partitioning strategy, such as cross-validation, will be adopted to ensure the stability and reliability of the model.

The design of the MoveNet aims to optimize the performance of human pose estimation. The network architecture of MoveNet will be studied in depth, and the network depth and width, as well as the size and number convolutional kernels, will be adjusted according to the specific application scenario. At the same time, different training strategies will be explored to further improve the accuracy and efficiency of the MoveNet model.

The training of the MoveNet model mainly includes the following steps: data preparation, model training, feature extraction, sequence modeling, and classification recognition. Data preparation: Collect and a dataset of human poses, including RGB images or video data of single or multiple people. The data preprocessing phase involves operations such as cropping, scaling, and normalization to enhance the model's feature learning capabilities. The workflow for the pose estimation model consists of the following key steps:

(1) Model Initialization: The model parameters are initialized, and hyperparameters like the learning rate and optimizer are configured.

(2) Feature Extraction: The input video or image is divided into multiple temporal segments or spatial regions. A Convolutional Neural Network (CNN) is then employed to extract a set of feature vectors from each of these segments or regions.

(3) Sequence Modeling: Algorithms such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory networks (LSTMs) are utilized to model the sequences of extracted feature vectors. This step captures the temporal dynamics of human poses and understands the correlations and dependencies across different time steps.

(4) Classification and Recognition: A classifier, typically a fully connected neural network, is used to categorize the output from the sequence modeling. This final stage recognizes the human pose, outputting either keypoint coordinates or a specific pose category.

For deep learning inference tasks, this system will design an efficient fully connected neural network (DNN. The number of network layers, the number of neurons, and the connection method will be determined according to the task requirements, and the appropriate activation function and loss function will be selected. In addition, optimization strategies for the network, such as gradient descent algorithms and batch sizes, will also be carried out to improve the training speed and performance of the DNN.

In summary, to ensure the effectiveness and reliability of the model, this system will perform a reasonable division of the dataset, including training set, validation set, test set. During the model training process, appropriate loss functions and optimization algorithms will be adopted to update the model parameters through iterative training until the predefined performance indicators are reached. At same time, the training process will be monitored and evaluated to ensure the stability and convergence of the model.

3.3 Real-time detection software design

3.3.1 Model integration

The system function design can be seen in two parts in Figure 3 , the first part is MoveNet to obtain the key point coordinates of the human body, and the second part is a classifier that uses these coordinates to determine whether the sitting posture is. MoveNet is launched by Google, and Google has a special tool Tensorflow Lite for making mobile AI applications. The classifier of this project enables AI to learn to distinguish between standard sitting posture, neck extension, and crossing legs.

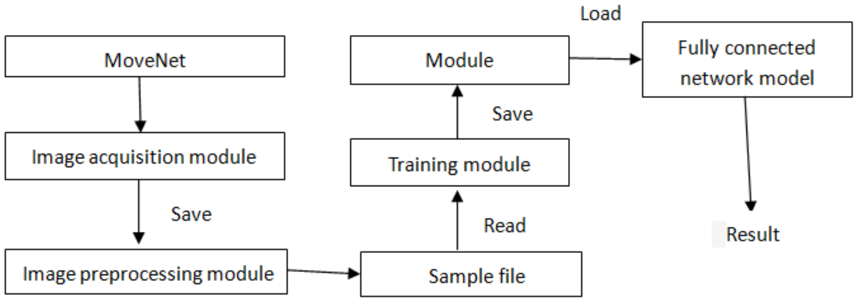

Figure 3 System function design diagram

3.3.2 Software Interface Design

The system will design a clear layout and operation process, allowing users to easily perform operations such as selection, parameter setting, and real-time detection. Meanwhile, this graduation project will provide friendly prompts and feedback information to help users better understand and use the software.

The interface will adopt a clear block design, dividing different functional areas explicitly, allowing users to quickly locate the required operations. The main functional areas include the video display area, result display area and parameter setting area, etc. The video display area will display the user's video stream in real-time, allowing users to observe their posture status. The result display area show the user's posture evaluation results in real-time; the parameter setting area allows users to adjust the detection parameters according to their personal needs. The operation process will be as much as possible to allow users to get started easily [14]. After the user opens the software, they will directly enter the main interface. During the detection process, the video display area display the user's video stream in real-time, and the result display area will continuously update the user's posture evaluation results. Users can also adjust the detection parameters through parameter setting area at any time to meet different scenarios.

3.3.3 Real-time Sitting Posture Detection and Feedback

Real-time sitting posture detection and feedback is one of the core functions the software system. The MoveNet model that has been integrated will be used to estimate the human pose of the input image or video in real time, and the user's sitting status will be judged by analyzing the position of key points. Once an unhealthy sitting posture is detected, the system will immediately provide feedback and suggestions to help the user adjust the sitting posture maintain a good posture.

4. System Testing

4.1 Test Objectives

With the thriving development of the Internet industry, the core position of software in the product R&D process has become increasingly prominent. It runs through every stage of product development and is responsible for ensuring software quality, optimizing user experience, and enhancing R&D efficiency. Its fundamental purpose is to create high-quality software products that bring users a smooth and satisfactory experience, thus laying a market foundation for the product. Given modern Internet products especially those involving health and privacy, such as body monitoring products, data security and privacy protection are crucial. Therefore, it has become imperative to protect user privacy through data encryption and strict measures. Any security vulnerabilities could lead to unimaginable consequences, damage user trust, and even threaten the reputation and interests of the enterprise. During the product R&D process the software testing team continuously tracks requirements, verifies product quality, and provides timely defect reports to promote continuous technological progress. By building a comprehensive testing system, the quality of the can be more effectively controlled. For instance, conducting compatibility tests across terminal devices can significantly reduce the cost and risk of subsequent version development. The value of software testing is not only reflected reducing development costs but also in effectively reducing business risks caused by software defects. A product with frequent issues finds it difficult to gain trust and recognition from partners, which can lead to significant losses. Therefore, software testing is an indispensable link before product launch. By intervening in testing early, it can be ensured that software quality is comprehensively protected. After meticulous and necessary adjustment and optimization, the product can be brought to the market, which not only improves user experience but also enhances user confidence in the product. Thereby, the product can avoid negative user feedback or complaints upon release and earn widespread recognition in the market.

4.2 Test Content

The test content is the core part of the system test, which covers multiple aspects to ensure the quality of the system The test content mainly includes functional testing, performance testing, and compatibility testing of Android applications. Some of the test content is as follows:

(1) Functional Testing: The testers verify that the Android application can correctly capture image data and transmit it to the backend. The testers confirm that the backend properly receives requests and returns inference results. The testers ensure that the Android application can parse the backend response and display it to the user.

(2) Performance Testing: The testers evaluate the inference speed of the backend to ensure results are returned within a reasonable time. They assess the responsiveness and smoothness of the Android application to guarantee a satisfactory user experience. They monitor resource utilization during operation to confirm that the system does not consume excessive resources.

(3) Compatibility Testing: The testers test the compatibility and stability of the Android application across different models and configurations of Android devices.

5. Conclusion

Through systematic design and implementation, this study successfully constructed an efficient and stable sitting posture detection system. During the research process, background analysis, technology comparison to clear requirement definition, the theoretical construction and engineering implementation of the system were gradually completed. In terms of model selection, this paper explores in depth theNet, MoveNet, and fully connected neural network (DNN) and other architectures, significantly improving the detection accuracy and reasoning efficiency through experimental comparison and parameter optimization, laying a solid for the core functions of the system. Based on the previous functional and non-functional requirement analysis, the system design adopts a modular architecture, achieving a complete process from data collection, processing, model reasoning to real-time feedback, taking into account real-time, accuracy and user experience. System testing shows that the system exhibits good stability, compatibility and robustness in variety of real-world scenarios, verifying its feasibility in actual applications. This research not only achieves key technological breakthroughs in sitting posture detection systems but also provides a solution that can referred to for the development of related health monitoring tools, and has certain theoretical significance and practical value.

Acknowledgements

This work was supported by Project of Haojing College of Shanxi University of Science and Technology , 2024 school-level scientific research innovation team project (Project name: Artificial intelligence health detection, Project number: HJTD09).

References

[1] Gebreyesus Tsiwaye, Nigussie Kalkidan, et al. The prevalence and risk factors of work-related musculoskeletal disorders among adults in Ethiopia: a study protocol for extending a systematic review with meta-analysis of observational studies [J]. Systematic reviews, 2020, 9(1): 103-117.

[2] Wu Lei, Wang Shuai, Liao Haoran, Wu Qingming, Zhu Changcai, Guo Zhenzhong. Quantitative analysis method for risk factors of occupational musculoskeletal disorders. Chinese Journal of Industrial Hygiene and Occupational Diseases. 2018,36(04):318-320.

[3] Egern Matthias, Aasa Ulrika, Äng Björn O, et al. Musculoskeletal disorders and their associations with health-and-related factors: a cross-sectional comparison between Swedish air force personnel and army soldiers. [J]. BMC musculoskeletal disorders, 200, 21(1): 161-175.

[4] Oakman Jodi, Kinsman Natasha. Barriers to more effective of work-related musculoskeletal and mental health disorders. [J]. Applied ergonomics, 2019, 75: 6674.

[5] Sun Qian, Hu Bo, Chen Shengxiong, Huang Lin, Hu Hao, Liao Yuanpeng. Analysis of risk factors for occupational musculoskeletal diseases among office clerks [J]. Occupational Health and Emergency Rescue. 2020,38(03):233-237.

[6] Shi Ping, Li Wei, Wang Weimin, Yu Hongliu, Liu. A Portable Decubitus Ulcer Prevention System Based on Pressure Distribution Detection [J]. Biomedical Engineering & Clinical, 2020, 24(3): 233-238.

[7] Du Yinju, Du Guiying, Zhang Furen, et al. Analysis of the situation cervical spondylosis in primary and secondary school students in Liaocheng City [J]. Chinese Journal of School Health. 2006(4): 339-340.

[8] Wang Chunyang. Analysis of the Current Situation of Posture Survey among Adolescents in [J]. Sports Teacher and Friend, 2018, 41(01): 44-46.

[9] Wangm, Chen Hl. Chaotic multi-swarm whale optimizer boosted support vector machine formed medical diagnosis [J]. Applied Soft Computing, 2020, 8.

[10] Ma Zhaolong, Qiu Yong, Wang Bin, Zhu Zezhang, et al. Spinal cord and vertebral malations in patients with congenital scoliosis [J]. Chinese Journal of Spine and Spinal Cord, 2007, 17(8):588-592

[11] Jia Yunde. Machine Vision [M]. Beijing: Science Press, 2004: 1-4.

[12] Min Weidong, Cui Hao, Han Qing, et al. A Scene Recognition and Semantic Analysis Approach to Unhealthyitting Posture Detection during Screen Reading [J]. Sensors, 2018, 18(9): 3119-3129.

[13] Zhou Yu. Design and Implementation of a Posture Monitoring System Based on Machine Learning [D]. Zhejiang University, 2018.

[14] Ma Yehan. Design and Implementation of an Office Health Analysis System Based on Kinect [D]. Harbin Institute of Technology, 2015.