



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

<https://www.inderscience.com/ijict>

Intelligent physical health testing method integrating Fast-OpenPose and BlazePose

Wei Xu

DOI: [10.1504/IJICT.2026.10076788](https://doi.org/10.1504/IJICT.2026.10076788)

Article History:

Received:	14 October 2025
Last revised:	16 December 2025
Accepted:	16 December 2025
Published online:	19 March 2026

Intelligent physical health testing method integrating Fast-OpenPose and BlazePose

Wei Xu

Department of Sports and Military Education,
Jingdezhen Ceramic University,
Jingdezhen, 333403, China
Email: xu156462188@163.com

Abstract: Traditional manual methods are often subjective and inefficient, falling short of the demands for large-scale and high-efficiency testing. To address this, an intelligent and automated fitness testing platform has been developed. The study integrates an improved Fast-OpenPose model with a support vector machine for sit-up detection, and combines BlazePose with an enhanced DeepLabV3+ for pull-up evaluation. Finally, a comprehensive platform incorporating both Fast-OpenPose and BlazePose is established. Results indicated that the improved Fast-OpenPose model achieved an average accuracy of 71.2%, an 8.9% increase over the original OpenPose, with a processing speed of 20 FPS. After 2,500 iterations, the refined DeepLabV3+ model attained 78.5% mIoU and 81.7% FWIoU. The platform showed minimal deviation between automated and manual counts for sit-ups and pull-ups, achieving an overall accuracy of 99.82%. This study offers a more intelligent, efficient, and accurate solution for physical fitness assessment, advancing the field toward intelligent automation.

Keywords: OpenPose; BlazePose; physical health test; PHT; automation; DeepLabV3+; support vector machine; SVM.

Reference to this paper should be made as follows: Xu, W. (2026) 'Intelligent physical health testing method integrating Fast-OpenPose and BlazePose', *Int. J. Information and Communication Technology*, Vol. 27, No. 25, pp.20–41.

Biographical notes: Wei Xu is an Associate Professor in the Department of Physical Education and Military Training at Jingdezhen Ceramic University. He obtained his Doctor of Education (EdD) from the Philippine College of Health Sciences in 2023. His research interests include physical education and sports management.

1 Introduction

As public health awareness continues to grow, physical health testing (PHT) has received widespread attention as an important means of evaluating individual physical function and health status (Arya and Maji, 2024; Woo and Jeong, 2025; Liu and Shi, 2024). Traditional PHT methods, such as sit-ups and pull-up tests, typically rely on manual counting and judgment. This not only consumes a significant amount of manpower and material resources, but also presents several problems, including strong subjectivity, low

efficiency, and a high likelihood of counting errors, making it difficult to meet the large-scale and high-efficiency PHT requirements (Surkov et al., 2025; Wang and Dong, 2024; Uo et al., 2025). Therefore, developing an intelligent and automated PHT platform has important practical significance (Yang and Kim, 2024). Advances in computer vision and artificial intelligence (AI) technology have provided new ideas for solving this problem. Human pose estimation (HPE) and semantic segmentation, as two key technologies in computer vision, have shown great potential in human action recognition (HAR), object detection, and other areas (Hassanin et al., 2025; Xu et al., 2025). Many scholars have also conducted research on this issue.

To save on the cost of human gait assessment, Hulleck et al. integrated the Seq2seq autoencoder deep learning model into the low-power BlazePose algorithm and proposed a smartphone-based pose estimation algorithm. This method reduced the MAE of the left ankle joint angle from 13.4° in fast gait to 5.3° , from 16.3° in normal gait to 7.5° , and from 15.5° in slow gait to 7.5° (Hulleck et al., 2024). To effectively identify the unique features of traditional dance, Yuan et al. proposed an AI-based Korean traditional dance action recognition model. This model applied the BlazePose model to generate real-time bone keypoints. These key points were integrated with the enhanced metadata set and processed through customised methods to achieve real-time recognition of dance movements. This method could effectively identify five basic traditional Korean dance movements and had great potential in analysing complex movement patterns (Yuan and Zhou, 2025). In response to the issues of high matching error and low recognition rate in traditional skeleton point-based human motion matching methods, Su et al. constructed a skeleton point coordinate motion matching model built on an OpenPose model. This model used a feedforward network for 2D confidence mapping and introduced a greedy relaxation algorithm to optimise the accuracy of multi-body bone point association matching. The loss rate of this method was 12.4%, and the Rank-n recognition rate could reach 95.2%, indicating higher accuracy. Its application in physical education could help improve the completion of course tasks (Su, 2025). To evaluate the effectiveness of motion detection algorithms, Lopes et al. evaluated the extraction performance of OpenPose (2D), YOLO Pose (2D), and BlazePose (3D) on human skeletons. It was found that using OpenPose skeleton extraction (2D) achieved the highest F1-score of 0.745, followed by calculating the angles of each joint. This indicated that OpenPose skeleton representation could be a feasible solution for computing human motion in serious games (Lopes et al., 2024). To efficiently recognise human movements, Jayamohan and Yuvaraj (2025) developed a new method for recognising human movements through image semantic segmentation. This method achieved a precision of 98.52% on UCF101 and 84.25% on HMDB51. This model achieved advanced results in HAR tasks, surpassing existing methods in terms of accuracy (Jayamohan and Yuvaraj, 2025).

In summary, OpenPose and BlazePose, as advanced HPE models, can precisely detect the position information of key points in the human body, providing strong support for action recognition and analysis. Semantic segmentation technology can accurately divide different regions in an image, which helps extract key targets from complex backgrounds. However, there are still three core bottlenecks in the large-scale application of existing related technologies in PHT: firstly, in terms of computational efficiency, the traditional OpenPose parameter quantity is difficult to support multiple concurrent testing scenarios. Although lightweight models such as MediaPipe Pose have improved speed, their accuracy in detecting core joint points for specific body measurement movements, such

as sit-ups and pull-ups, is insufficient. Secondly, in terms of specific action recognition accuracy, most of the existing research focuses on gait, dance, and other actions. There is no special feature extraction mechanism designed to standardise the judgment of anthropometric actions (such as sit-up angles, pull-up bar standards), which can easily lead to counting deviations. Thirdly, at the system integration level, most research only focuses on the verification of a single algorithm, and does not realise the full-process integration of “attitude detection, action discrimination, counting, and statistics platform deployment.” It lacks optimised solutions for edge devices and is difficult to adapt to large-scale testing scenarios such as campuses. In response to this issue, this study innovatively integrates the advantages of Fast-OpenPose and BlazePose HPE models and combines semantic segmentation technology to build an intelligent physical health testing platform (IPHTP). In terms of sit-up tests for physical health, this study will deeply integrate the improved Fast-OpenPose algorithm with the powerful support vector machine (SVM). An efficient and reliable automatic counting method is innovatively proposed, effectively avoiding the subjectivity and errors of manual counting. In terms of pull-up testing, a method combining semantic segmentation and HPE has been proposed. The paper aims to offer a more intelligent and accurate solution for PHT by integrating multiple advanced technologies, and promote the intelligent development of PHT.

2 Methods and materials

2.1 *Design of sit-up testing method based on improved Fast-OpenPose*

The current traditional sit-up test mode has obvious drawbacks. During the testing process, not only does it require the deployment of numerous manpower inputs, but it is also susceptible to subjective factors from the counter, resulting in low overall counting efficiency and greatly reduced accuracy, making it difficult to meet testing requirements (Bhosle and Musande, 2023; Kumar and Worsley, 2023; Giles et al., 2020). To promote the intelligent direction of PHT, this study deeply integrates the improved Fast-OpenPose with the powerful SVM, and innovatively proposes an efficient and reliable automatic counting method. When processing test videos, this method first converts the video into a single frame image, uses the OpenPose model to accurately extract human features, and predicts key point position information. Subsequently, the SVM classification algorithm is used to classify and process the key point data, to determine the regularity of the action. If standing, the detection will stop, and the number of movements during that time period will be output.

The OpenPose model used in this study is a multi-person pose estimation model based on convolutional neural networks. Its principle is to achieve keypoint detection and correlation through a two-stage branch network (Wang et al., 2022; Tristan et al., 2022). Firstly, the backbone network extracts multi-scale features of the input image. Branch 1 uses part affinity fields (PAFs) to encode spatial relationships between human body parts. Branch 2 predicts the key point heatmap. By using non-maximum suppression to locate key point coordinates, combined with the vector field information of PAF, a greedy algorithm is employed to match different body parts, ultimately constructing a complete human skeletal structure (Worsey et al., 2021; Dai and Lu, 2021; Ren, 2023). The expressions for the two branches are shown in equation (1).

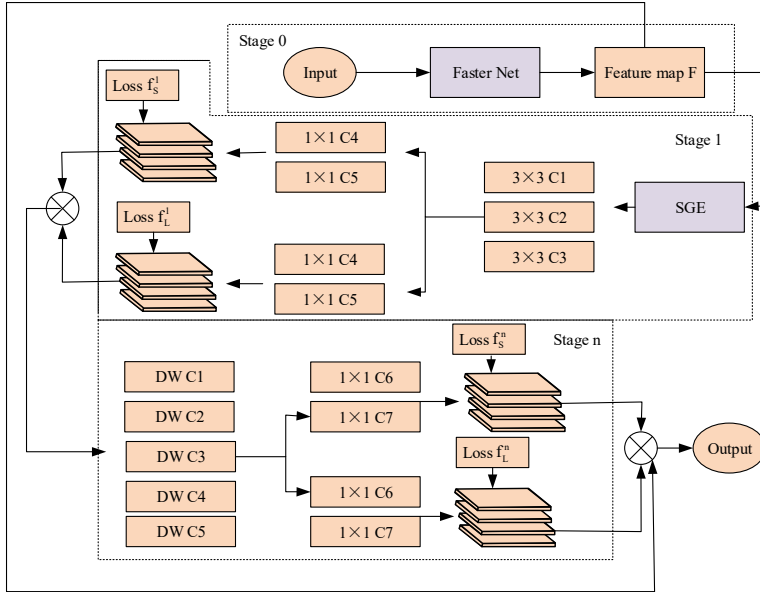
$$\begin{cases} S^n = \rho^n(F, S^{n-1}, L^{n-1}), & \forall n \geq 2 \\ L^n = \phi^n(F, S^{n-1}, L^{n-1}), & \forall n \geq 2 \end{cases} \quad (1)$$

In equation (1), S^n and L^n are the outputs of branches one and two. ρ^n and ϕ^n are the intermediate results of the n^{th} iteration of the network. After each stage of model training, it is necessary to calculate the confidence map of key points and the loss function values of PAFs, store them properly, and then accumulate the losses of each stage to calculate the gradient. As the training stages increase, the model is more likely to converge, and the accuracy of the generated prediction graph will also significantly improve. The expressions for the two loss functions are shown in equation (2).

$$\begin{cases} f_s^n = \sum_{j=1}^J \sum_P W(P) \cdot \|S_j^n(P) - S_j^*(P)\|_2^2 \\ f_L^n = \sum_{c=1}^C \sum_P W(P) \cdot \|L_c^n(P) - L_c^*(P)\|_2^2 \end{cases} \quad (2)$$

In equation (2), P is the key point position area. f_L^n and f_s^n are the loss functions of PAFs and CPMs. L_c^* and S_j^* are true values. $W(P)$ is the weight of the key point in the P region. S_j^n and L_c^n are predicted values for keypoint confidence and affinity.

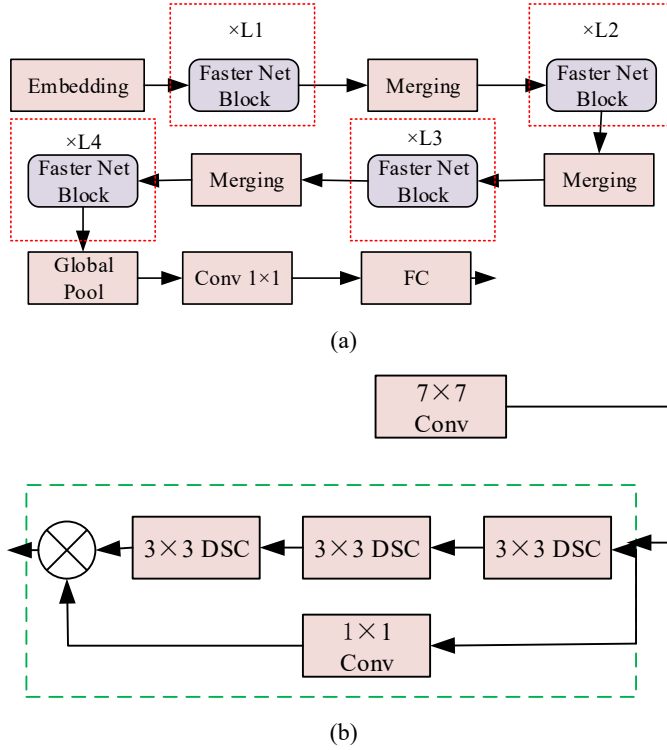
Figure 1 Improved OpenPose structure (see online version for colours)



However, traditional OpenPose models have exposed many shortcomings in practical applications, such as high complexity and an extremely large number of model parameters, resulting in a time-consuming detection process that is hard to meet the needs of real-time and efficiency (Li et al., 2021; Wang and Shi, 2023; Buttfield and Ball,

2020). To address these issues, this study combines FasterNet with the spatial group-wise enhancement (SGE) mechanism to carry out targeted improvements to the OpenPose model. The improved OpenPose framework is displayed in Figure 1.

Figure 2 FasterNet architecture diagram, (a) FasterNet network (b) improved convolutional block structure (see online version for colours)

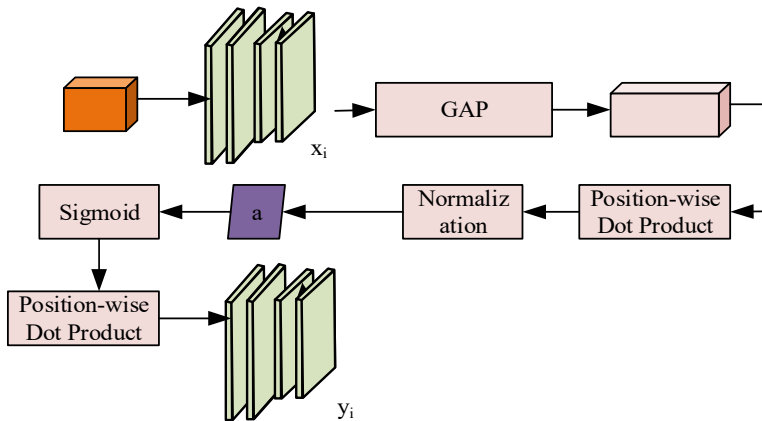


In Figure 1, this study implements four systematic improvements in model architecture and key module design. Firstly, the architecture upgrade is carried out for the backbone network, replacing the traditional VGG-19 network with a lightweight FasterNet architecture, which significantly reduces computational complexity through grouping convolution and channel shuffling mechanisms while maintaining feature extraction capabilities. Secondly, in terms of optimising the prediction branch structure, the symmetric dual branch structure of stages 1 to 6 has been reconstructed into a hybrid architecture consisting of a single branch shared convolution in the first half and independent optimisation of dual branches in the second half. This design enables efficient reuse of low-level features and reduces the amount of redundant calculations. The third step is to improve the convolution operation of the common branch in stages 2 to 6 by using depthwise separable convolution instead of standard convolution, and reconstructing the residual connection path to enhance gradient propagation efficiency. Finally, after the feature extraction layer in stage 0, the SGE module is introduced to dynamically allocate spatial weights to enhance the feature response of key areas of the human body (such as joints and contours), thereby improving the spatial semantic modelling ability for pose estimation tasks (Ellens et al., 2022). The conventional

VGG-19 backbone network suffers from excessive parameter size and high computational complexity, making it inadequate for real-time requirements in embedded systems. As a lightweight architecture, FasterNet uses grouped convolution and channel transformation mechanisms to significantly reduce parameter counting and computational overhead while maintaining strong feature extraction capabilities. Therefore, it is more suitable for real-time applications in physical health monitoring. The FasterNet structure and improved convolutional block structure used for optimising the backbone network are shown in Figure 2.

In Figure 2(a), the overall FasterNet architecture consists of four stages. Before starting each stage, targeted embedding or merging layers will be setup to prepare for subsequent processing. Each stage is mainly built by multiple carefully designed FasterNet blocks. At the end of the architecture, the classification task is completed sequentially through the global pooling layer and the fully connected layer. The inverse residual structure that combines partial convolution and point convolution is used within the block to divide the input feature map (Fmap) into two parts for processing and then add them up to output, thereby improving performance. In Figure 2(b), in terms of convolutional block optimisation, there are many 7×7 convolutions and high computational costs in the prediction stage. It is replaced with convolutional blocks that can capture long-range spatial dependencies and reduce complexity. The stacking depth can separate the convolutional layers and add residual connections to expand the receptive field, while reducing the quantity of parameters and computational complexity.

Figure 3 SGE attention module structure (see online version for colours)



Finally, in the process of network lightweighting improvement, issues such as limited performance, insufficient expression of key areas, and noise interference urgently need to be addressed. Therefore, this study introduces the SGE module. In HPE tasks, SGE demonstrated outstanding performance. In sit-up testing, the lumbar region, hip joint, and knee joint serve as key detection points for core motion identification, which are prone to background interference, causing detection deviations. The dynamic weight adjustment mechanism of the SGE module operates through four steps: First, input feature maps are divided into eight subgroups by channel, with three subgroups specifically matching feature channels for lumbar, hip, and knee joints. Second, global average pooling (GAP) is applied to extract spatial semantic statistics from these subgroups, generating a

semantic similarity matrix between groups. Third, a spatial weight distribution is created based on this similarity matrix, assigning high weights (0.8–1.0) to core joint regions and low weights (0.2–0.5) to non-core areas like the head and ankle joints. Finally, an attention mask is generated using the Sigmoid activation function to enhance core region features while suppressing background and non-core areas, ultimately improving core joint detection accuracy. The structure of the SGE attention module is exhibited in Figure 3.

In Figure 3, the SGE module first divides the input Fmap into multiple subgroups based on channel dimensions, with each subgroup independently carrying specific semantic feature information. Subsequently, the intra-group spatial semantic statistics are extracted through GAP, and the dot product operation is used to calculate the inter-group semantic similarity matrix. It is transformed into a spatial weight distribution through standardised mapping and finally generates an attention mask through the Sigmoid activation function, achieving spatial selective enhancement of key semantic features and dynamic suppression of irrelevant regions. The descriptive formula for statistical vector g is shown in equation (3).

$$g = F_{gp}(x) = \frac{1}{m} \sum_{i=1}^m x_i \quad (3)$$

In equation (3), $F_{gp}(x)$ is the spatial average function, m is the number of pixels, and x_i is the group spatial local eigenvector.

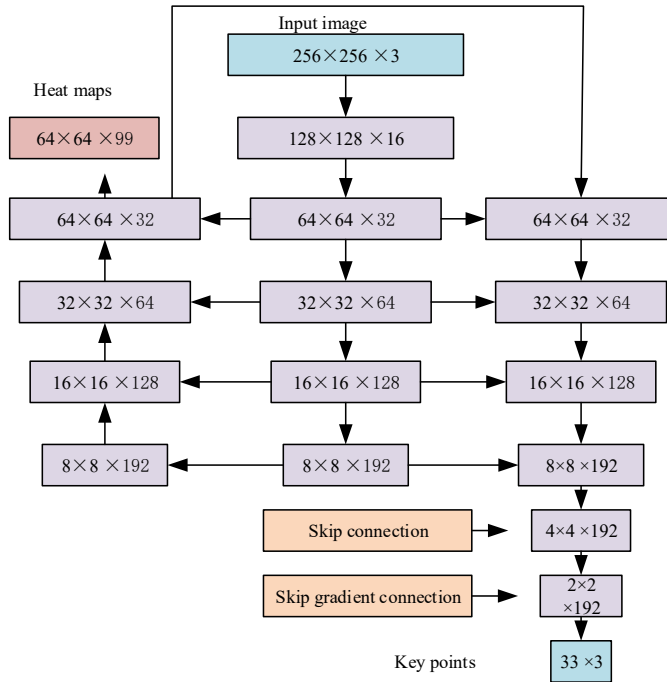
2.2 Design of pull-up testing method based on BlazePos

After completing the innovative design of the sit-up test method in the PHT project and ensuring its scientific validity and feasibility, this study innovatively proposes a new testing method to continuously optimise the physical testing process and further improve overall efficiency. This method combines BlazePose with the improved DeepLabV3+'s pull-up testing method. When processing video data, this method first breaks down the video into single-frame images and then uses the improved lightweight DeepLabV3+ to accurately extract the boundary information of the horizontal bars in the first frame image. Subsequently, it uses the BlazePose model to detect and locate the coordinates of key points on the human body. The behaviour discrimination module develops specific algorithms built on the coordinates of these key points to classify human postures. In the pull-up test, BlazePose and the improved DeepLabV3+ adopt a 'segmentation-first, detection-second' collaborative approach. First, the improved DeepLabV3+ performs semantic segmentation on the initial frame to identify the pull-up bar region, outputting a bar mask image and extracting the upper edge coordinate threshold of the bar. Subsequently, BlazePose conducts human keypoint detection on subsequent video frames, focusing on extracting upper limb keypoints such as shoulders, elbows, and wrists. The keypoint coordinates from BlazePose and the bar position information from DeepLabV3+ are fed into the action discrimination module, where action classification is achieved by comparing the relative positions of the shoulder keypoints. The bar region features output by DeepLabV3+ are transmitted to the action discrimination module as parameterised spatial references, while the human keypoint features from BlazePose are transmitted in the form of coordinate matrices. The two sets of features are fused through a 'spatial reference-keypoint matching' algorithm, which calibrates the relative

coordinates of human keypoints based on the bar position to eliminate positional deviations caused by image capture angles.

The BlazePose used in this method is an efficient HPE network. It is designed for real-time and mobile deployment, using a lightweight architecture that can run quickly on resource-constrained devices. This network combines heat maps with keypoint offset vectors to accurately locate the coordinates of multiple keypoints on the human body. It not only performs well in conventional scenarios but also adapts to complex backgrounds and different human postures, and is widely used in fields such as fitness guidance and motion analysis. The BlazePose network structure is shown in Figure 4.

Figure 4 BlazePose network structure (see online version for colours)



In Figure 4, BlazePose adopts an encoder-decoder structure, where the encoder extracts human features, and the decoder generates keypoint heatmaps and offset vectors. The two are fused to accurately locate keypoints, achieving efficient real-time HPE. In the proposed pull-up testing method, the traditional DeepLabV3+ model has high computational complexity and is difficult to deploy on embedded platforms with limited computing power. The feature extraction part adopts ShuffleNetV2 for lightweight transformation, effectively reducing the computational load. Meanwhile, in the decoding module, a channel attention mechanism is introduced to lift the model's capacity to capture key features. The improved model is applied to the segmentation task of horizontal bars in pull-up scenes to achieve efficient and accurate segmentation. Figure 5 shows the improved DeepLabV3+.

In Figure 5, the improved model follows the encoding and decoding structure. The encoder part uses the lightweight and efficient ShuffleNetV2 as the backbone network. By optimising the size of the convolution kernel in a targeted manner, unnecessary

computation is reduced, significantly improving the efficiency of feature extraction and enabling the model to capture key information faster. The hollow space pyramid pooling module cleverly utilises different hollow convolution rates to achieve full fusion of multi-scale features, followed by dimensionality reduction and upsampling, and effective fusion with underlying features to enrich semantic information. The ECA module is introduced into the decoding structure to ensure accuracy while making the segmentation boundaries clearer. Finally, the model first uses 3×3 convolution to deeply extract and integrate features, and then restores image resolution through bilinear $4\times$ upsampling, accurately outputting clear and precise horizontal bar segmentation results. The ShuffleNetV2 and ECA structures used in the improvement of DeepLabV3+ are displayed in Figure 6.

Figure 5 Improved DeepLabV3+ structure (see online version for colours)

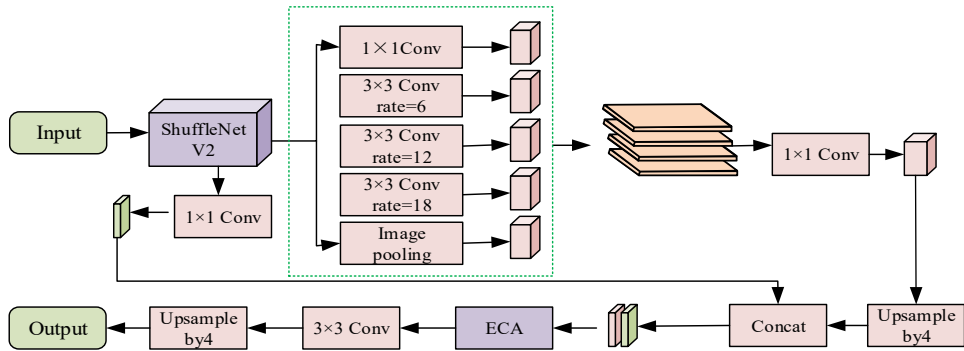
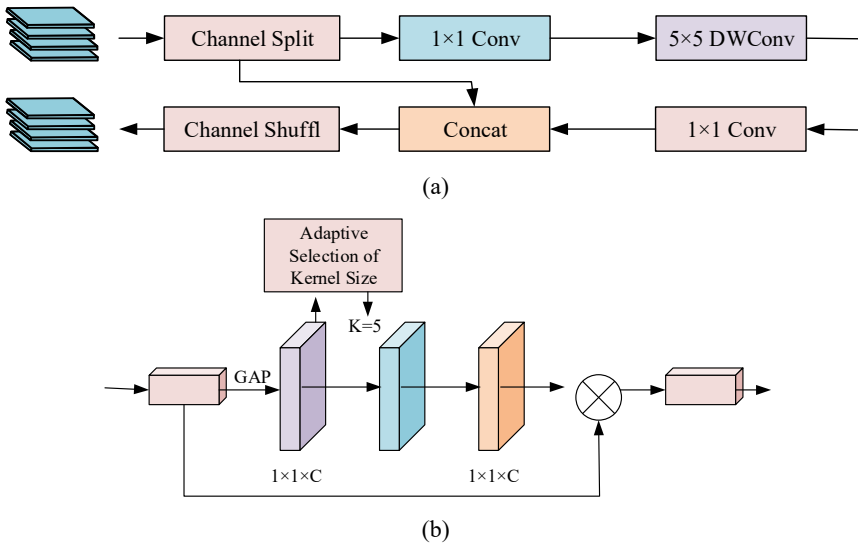


Figure 6 Network structure of ShuffleNetV2 and ECA, (a) ShuffleNetV2 basic unit (b) ECA network structure (see online version for colours)



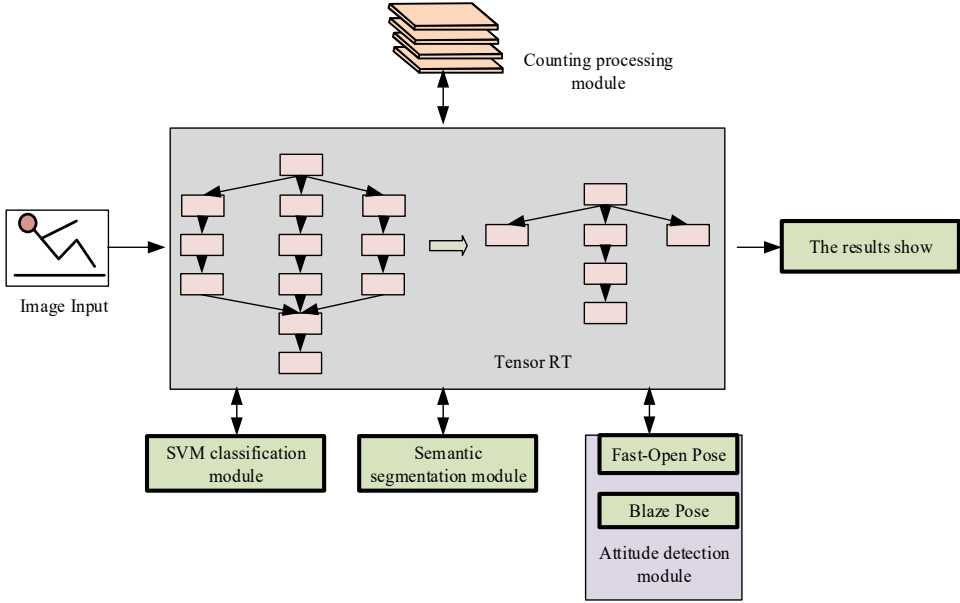
In Figure 6(a), the distribution of computational complexity in ShuffleNet is observed, and it is found that a large number of computational tasks in the model are concentrated

in the 1×1 convolution part. The proportion of computational complexity undertaken by depthwise separable convolution is relatively small. Based on this discovery, this study carries out innovative attempts to expand the convolution kernels of depthwise separable convolutions without significantly increasing the computational weight of deep convolution, to improve image segmentation performance. In Figure 6(b), ECA has an efficient process for processing Fmaps. This module takes an Fmap with dimensions of $H \times W \times C$ as input, where H and W are the height and width of the Fmap, and C is the number of channels. Firstly, the GAP operation is applied to compress the spatial dimensions (i.e., H and W dimensions) of the Fmap, obtaining the global statistical information of the channel dimension. Subsequently, 1D fast convolution is utilised to efficiently capture local cross-channel interaction information in the channel dimension, and then channel features are further learned through 1×1 convolution. Finally, the learned features are multiplied channel by channel with the original Fmap to output a brand new Fmap with channel attention.

2.3 Building of IPHTP integrating Fast-OpenPose and BlazePose

After completing the design of sit-up and pull-up test methods in the PHT project, this study deploys the two methods to the edge computing platform Jetson Nano and builds an IPHTP that integrates Fast-OpenPose and BlazePose. The intelligent motion analysis platform is based on the powerful and compact Jetson Nano as the core body, and relies on multiple functional modules to work together. The SVM classification module, with its powerful classification ability, is responsible for accurately classifying the features of sit-up movements. The semantic segmentation module can efficiently obtain the position information of the pull-up horizontal bar. The posture detection module integrates two advanced network models, Fast-OpenPose and BlazePose, which can perform detailed detection of key points in the human body. The counting processing module focuses on the accurate counting of pull-ups and sit-ups. In addition, the TensorRT acceleration module can optimise and accelerate the models involved in the above modules, significantly improving the overall platform's operational efficiency and response speed, ensuring stable and efficient implementation of various functions. The IPHTP architecture is shown in Figure 7.

In Figure 7, the platform adopts a top-down hierarchical architecture consisting of a perception layer, processing layer, decision layer, and application layer. The perception layer collects video and environmental data through high-definition cameras and environmental sensors, transmits it to the processing layer via RTSP protocol, and receives instructions from the processing layer to start the device. The processing layer integrates image pre-processing, pose estimation, segmentation, and classification modules, and calls corresponding algorithms according to the decision layer task instructions to output feature data. The decision layer schedules the process through counting processing, TensorRT acceleration, and data verification modules, and feeds back the results to the application layer. The application layer provides user interaction, data storage, and report generation functions to respond to user testing requests. In addition, the data exchange format is clarified: the video is encoded using H.264, the key point coordinates are in JSON format, the horizontal bar mask is in PNG format, and the counting results are in XML format to ensure efficient and standardised data transmission between layers.

Figure 7 IPHTP architecture (see online version for colours)

3 Results

3.1 Dataset processing and model training

To validate the effectiveness of the designed PHT method, it is necessary to first test the performance of the improved Fast-OpenPose model and the improved DeepLabV3+. For the testing of the improved Fast-OpenPose, the MPII multiplayer dataset is chosen for this experiment. The MPII dataset is mainly used for HPE research and contains approximately 25,000 images covering 410 human activities. Each image is annotated with the positions of key points on the human body, such as the head, shoulders, elbows, etc., for a total of 16 key points. The dataset also provides information on human activity categories, perspectives, occlusions, etc., providing a comprehensive and challenging benchmark for evaluating the performance of HPE algorithms. For the MPII dataset, scene data that are irrelevant to the target task are analysed and filtered out in depth to enhance the targeting and effectiveness of the dataset. At the same time, to enrich the diversity of data, a dataset of 2,000 sit-ups is carefully created and added. For the convenience of subsequent processing, in the data pre-processing stage, this study uniformly adjusts the input size of all data to $368 * 368$ pixels to achieve standardisation of data format. When testing the improved DeepLabV3+ network model, work is conducted on a self-made dataset. Self-built sit-up and pull-up dataset contains 2,000 images (50 participants aged 12–15, half male and half female) collected in different indoor and outdoor lighting environments, labelled with 16 key points and motion states of the human body, and enhanced by flipping, brightness adjustment, and other methods. The pull-up dataset consists of 800 original images and 1,600 expanded images (40 males aged 12–15), collected in indoor gyms. The horizontal bar area and key points of the

human body are annotated, and the data are expanded through cropping, contrast adjustment, and other methods to clarify the annotation accuracy and anti-interference processing details of the two datasets. Two self-built datasets are divided into training set, testing set, and validation set in a ratio of 6:1:1. The MPII dataset and the self-made dataset are divided into 3:1:1 and 6:1:1, respectively, because the sample size of the self-made dataset is relatively small, so the proportion of the training set is increased to prevent overfitting. The experimental operating environment and parameter configuration are listed in Table 1.

Table 1 Simulation environment and parameter settings

/	Category	Set items
JRE	Operating system	Win 7 operating system
	Developer components	JDK 7, Eclipse
	CPU	Intel (R) Core (TM) i9-10900X
	GPU	Nvidia RTX 3090 GPU
	Platform software environment	Python3.8, PyTorch1.10.0
	Programming language	Java
Parameter setting	Batch size	4
	Learning rate initial value	0.0002
	Optimiser	Adam
	Learning rate decay	100

The testing metrics selected for the improved Fast-OpenPose model include mean average precision (mAP), AP50, AP75, parameter count, and frames per second (FPS). Among them, mAP, AP50, and AP75 are used to test the matching effect of human keypoints, and the parameter quantity and FPS are the evaluation criteria for model complexity and running speed. The traditional open pose model, high-resolution network (HRNet), simple baseline for human pose estimation (SimpleBaseline), and cascaded pyramid network (CPN) are selected as comparisons for the improved Fast-OpenPose model. The improved DeepLabV3+ selects testing metrics such as mean intersection over union (mIoU) and frequency weighted IoU (FWIoU). Full convolutional networks (FCN), pyramid scene parsing network (PSPNet), and object contextual representation network (OCRNet) are comparative models. Firstly, model training is conducted on both the MPII dataset and the self-built dataset, as shown in Figure 8.

Figures 8(a) and 8(b) show the fitness iteration curves of various models in the training set. In Figure 8(a), the improved Fast-OpenPose model has the best iterative performance in the training set. In the training set, the improved model only needs 229 iterations to reach a stable state, with a fitness value of 0.16. In Figure 8(b), the improved DeepLabV3+ can reach a stable state after 181 iterations, and the fitness value in the stable state is 0.14. The improved Fast-OpenPose and DeepLabV3+ are effective and can be applied to intelligent testing of physical health. The loss changes of the improved Fast-OpenPose and improved DeepLabV3+ on the MPII dataset and self-made dataset are shown in Figure 9.

Figure 9(b) shows the change in loss of the improved DeepLabV3+. In the improved Fast-OpenPose of Figure 9(a), both improved networks exhibit good convergence characteristics on both datasets. Specifically, the improved networks trained on both

datasets converge after 150 iterations. When improving Fast-OpenPose to 150 iterations, the loss function value decreases to around 0.8. When improving DeepLabV3+ to 150 iterations, the corresponding loss function value is about 1.0. Therefore, taking into account all factors, the maximum iteration of the experiment is set to 150.

Figure 8 Comparison of training effects of different models, (a) improve fast OpenPose (b) improve DeepLabV3+ (see online version for colours)

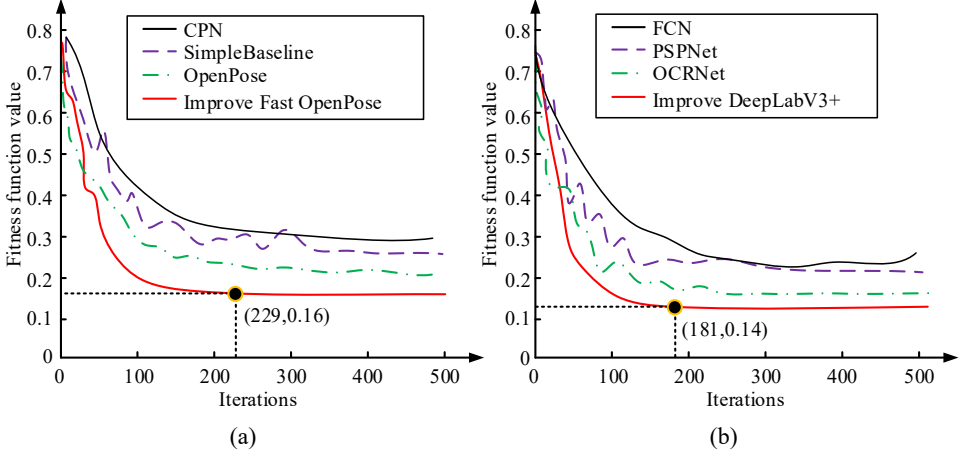
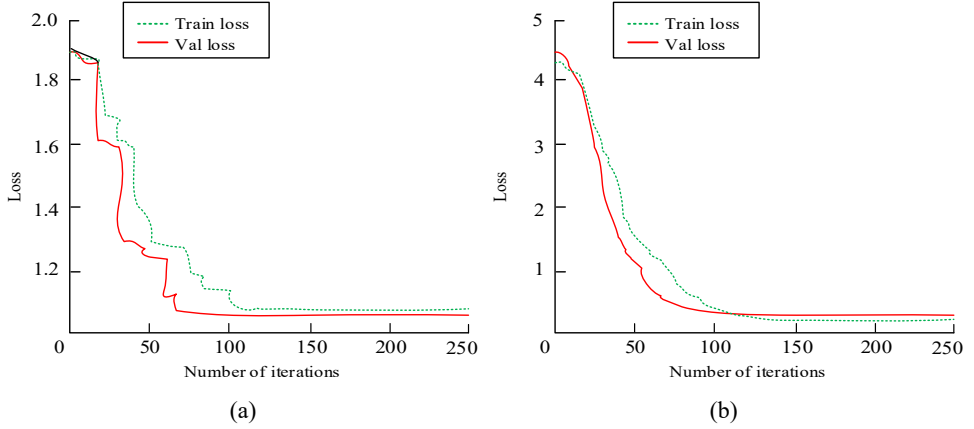


Figure 9 Changes in loss of improved model, (a) improve fast opening (b) improve DeepLabV3+ (see online version for colours)



3.2 Performance testing of improved models

After completing the model training, this study compares the improved Fast-OpenPose with traditional OpenPose, HRNet, SimpleBaseline, and CPN models. The key point matching effect is evaluated through mAP, AP50, and AP75, and the parameter quantity and FPS are used to measure the model complexity and running speed. The comprehensive performance of the improved model is verified, as exhibited in Table 2.

Table 2 Performance evaluation of improved Fast-OpenPose model

<i>Model name</i>	<i>mAP (%)</i>	<i>AP50 (%)</i>	<i>AP75 (%)</i>	<i>Number of parameters (M)</i>	<i>Frames per second (FPS)</i>	<i>Model size (MB)</i>
Traditional OpenPose	62.3	81.5	68.7	42.5	12	168.2
HRNet	70.1	88.3	76.9	65.8	8	261.5
SimpleBaseline	68.5	86.7	74.2	34.2	15	135.8
CPN	69.7	87.9	76.1	58.3	10	231.1
MediaPipe Pose	69.2	87.5	75.3	12.8	22	50.3
MoveNet	70.5	88.6	77.2	4.2	30	16.5
PoseNet	67.8	85.2	73.1	8.5	25	33.7
<i>Improve Fast-OpenPose</i>	<i>71.2</i>	<i>89.1</i>	<i>77.8</i>	<i>28.6</i>	<i>20</i>	<i>113.2</i>

Note: FPS test conditions are input image resolution of 368x368, hardware environment of Nvidia RTX 3090 GPU, software environment of Python3.8+PyTorch1.10.0, and test data are the average of 1,000 consecutive single frame image detection.

In Table 2, the improved Fast-OpenPose model demonstrates significant advantages. In terms of keypoint matching performance, mAP reaches 71.2%, which is 8.9% higher than traditional OpenPose, and 1.1%, 2.7%, and 1.5% higher than HRNet, SimpleBaseline, and CPN. AP50 is 89.1%, which is 7.6% higher than traditional OpenPose, and also has a 1%–2% improvement compared to other models. AP75 achieves 77.8%, which is also higher than other models, demonstrating its excellent performance under strict matching criteria and more accurate prediction of keypoint positions. In terms of model complexity, the improved Fast-OpenPose has 28.6 M parameters, which is significantly reduced compared to HRNet's 65.8 M and CPN's 58.3 M, and only slightly higher than SimpleBaseline. This indicates that optimising the model structure effectively reduces the parameters while ensuring accuracy. In terms of running speed, FPS reaches 20 frames/s, far exceeding other models, which means that the improved Fast-OpenPose can process image or video frames more quickly, meeting the needs of real-time applications. While MoveNet outperforms in parameter count and FPS, its mAP is 0.7% lower than the improved model, and its detection stability for core joints (hip and knee) in sit-up movements is inadequate. Both MediaPipe Pose and PoseNet show lower mAP values than the research model, demonstrating that the enhanced Fast-OpenPose achieves better performance in motion detection for physical fitness testing while maintaining lightweight characteristics. Overall, the improved Fast-OpenPose has achieved a good balance in accuracy, complexity, and running speed, with better overall performance and better suitability for applications that require high real-time performance. The performance of the improved Fast-OpenPose model is evaluated when the input resolution is reduced from 368×368 to 256×256 . The evaluation metrics include mAP, AP50, AP75, parameter count, and FPS, with the results presented in Table 3.

As shown in Table 3, after reducing the input resolution from 368×368 to 256×256 , the model's FPS increases from 20 to 32, meeting the requirements for ultra-high real-time scenarios (>30 fps). The number of parameters remains unchanged (28.6 M), as only the input size is adjusted without altering the network structure. In terms of accuracy, the mAP decreases by 2.7%, while AP50 and AP75 decrease by 2.8% and

3.3%, respectively, but still remain higher than traditional OpenPose (mAP 62.3%). This indicates that reducing the resolution can significantly improve speed, with the accuracy loss being within an acceptable range, allowing for flexible selection of the input size based on the real-time requirements of the testing scenario.

Table 3 Input resolution effects on accuracy-speed trade-off of improved Fast-OpenPose

<i>Input resolution</i>	<i>mAP (%)</i>	<i>AP50 (%)</i>	<i>AP75 (%)</i>	<i>Parameter (M)</i>	<i>FPS</i>
368 × 368 (original experiment)	71.2	89.1	77.8	28.6	20
256 × 256	68.5	86.3	74.5	28.6	32

To evaluate the impact of various parts of the improved Fast-OpenPose on model performance, four experimental groups are setup. One is the basic model group, which only uses the original OpenPose network structure; the second is the backbone network optimisation group, which only replaces the backbone network from VGG-19 to FasterNet; the third is the prediction branch optimisation group, which optimises only the prediction branch structure based on the basic model; the fourth is the complete improvement group, which is the improved Fast-OpenPose. The improved Fast-OpenPose includes all four improvements, namely optimising the backbone network, predicting branch structures, replacing common branch convolution types, redesigning residual connections, and introducing SGE modules after stage 0. Figure 10 shows the results of the ablation experiment.

Figure 10 Experimental results of ablation using the improved Fast-OpenPose model, (a) mAP (b) parameter quantity mAP (c) FPS (see online version for colours)

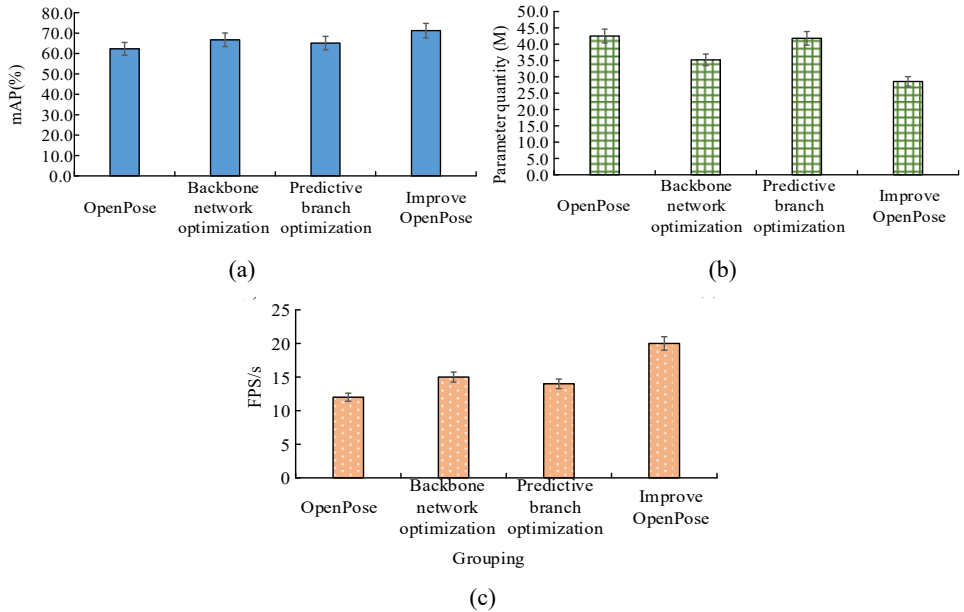


Figure 10 shows that all the improved parts have an enhancing effect on the performance. The mAP of the backbone network optimisation group increases by 4.4%, the number of parameters decreases by 7.3 M, and the FPS increases by 3 frames/s, indicating that

FasterNet can improve accuracy, reduce complexity, and accelerate running speed. The mAP of the prediction branch optimisation group increases by 2.8%, with small changes in parameter quantity and FPS, indicating that this optimisation has a certain improvement in accuracy and has little impact on model complexity and speed. The comprehensive improvement group has achieved a mAP of 71.2%, which is 8.9% higher than the basic model. The parameter count has been significantly reduced to 28.6 M, and the FPS has been increased to 20 frames/s. This proves that the synergistic effect of various improvements significantly improves the performance. The interaction results between components of the improved Fast-OpenPose model are shown in Table 4.

Table 4 Results of improved Fast-OpenPose model component interactions

<i>Experimental group</i>	<i>Included improved components</i>	<i>mAP (%)</i>	<i>FPS</i>	<i>Collaborative gain</i>
1	FasterNet + predictive branch optimisation	68.5	17	1.3%
2	FasterNet + SGE	70.2	18	2.5%
3	Prediction branch optimisation + SGE	69.3	15	1.7%
4	Convolution type replacement + SGE	69.8	19	2.1%
5 (complete improvement group)	All 4 improved components	71.2	20	3.2%

As shown in Table 4, the synergy between backbone network optimisation and SGE module integration demonstrates the most significant impact. This stems from FasterNet's lightweight feature extraction, which provides a more efficient feature foundation for SGE's spatial attention computation. Predicted branch optimisation shows minimal redundancy with other components, as both include feature reuse mechanisms, although the redundancy is still low (only 1%) and therefore does not affect overall performance. All component combinations feature no redundant designs, achieving a synergistic improvement in both accuracy and speed. The correlation analysis results of the SGE module and specific human body part features are shown in Table 5.

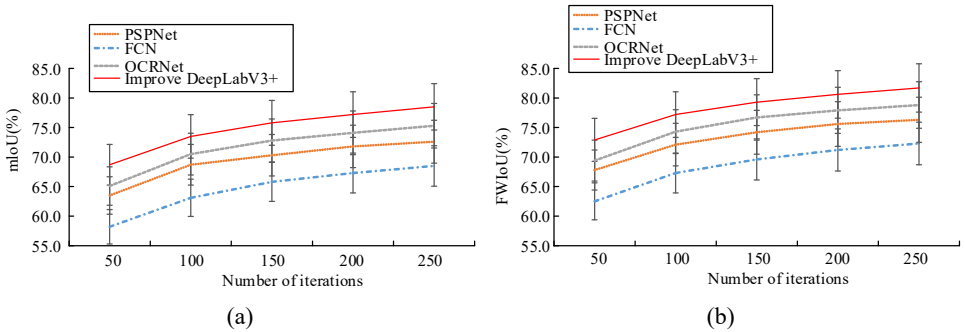
Table 5 Correlation analysis results of SGE module with specific human body part features

<i>Body parts</i>	<i>Detection accuracy rate without SGE (%)</i>	<i>Detection accuracy with SGE (%)</i>	<i>Improvement in accuracy rate (%)</i>
Head	89.2	90.1	0.9
Loin	78.5	86.3	7.8
Hip joint	76.3	85.7	9.4
Knee-joint	77.1	84.9	7.8
Ankle	82.4	83.2	0.8

As shown in Table 5, the SGE module demonstrates the most significant improvement in detection accuracy for core assessment areas (waist, hip, and knee joints) during sit-ups, with an average increase of 8.3%. The enhancement for non-core areas is relatively minor, indicating its ability to precisely capture key motion-related features and effectively improve the precision and specificity of posture estimation. On the self-made dataset test set, this study compares the improved DeepLabV3+ with FCN, PSPNet, and

OCRNet, using mIoU and FWIoU as test metrics to observe the performance of different models at different iterations, as shown in Figure 11.

Figure 11 Performance testing of improved DeepLabV3+ model, (a) mIoU (b) FWIoU (see online version for colours)



Overall, as the iterations increase, the mIoU and FWIoU of each model show an upward trend. In the early stages of iteration, the performance of each model improves rapidly, but the growth rate gradually slows down in the later stages. Under the same iteration number, the improved DeepLabV3+ model has higher mIoU and FWIoU than FCN, PSPNet, and OCRNet. In Figure 11(a), after 2,500 iterations, the improved DeepLabV3+ achieves a mIoU of 78.5%, which is 10%, 5.9%, and 3.2% higher than FCN, PSPNet, and OCRNet. In Figure 11(b), after 2,500 iterations, the FWIoU of the improved DeepLabV3+ reaches 81.7%, which is 9.4%, 5.4%, and 2.9% higher than FCN, PSPNet, and OCRNet. This indicates that the improved DeepLabV3+ performs better in the task of segmentation of pull-up horizontal bar regions, and can more accurately segment the target region, verifying the effectiveness of the improvement measures.

3.3 Testing and application analysis of IPHTP

To verify the practical application effect of the IPHTP, this study conducts testing and analysis on the platform. This study first deploys the designed sit-up and pull-up testing method to the proposed IPHTP and conducts performance testing using the LoadRunner tool. The minimum response time (MinRT), maximum response time (MaxRT), CPU and memory usage (CPU-MU) of the platform under different concurrent user numbers are listed in Table 6.

Table 6 Results of sports venue management platform

Number of concurrent users	Minimum response time/s	Maximum response time/s	CPU and memory footprint/%	Stability (continuous operation time/h)	Robustness (error rate under interference/%)	Power consumption (W)
100	0.592	0.949	7.5	72	0.8	18.2
300	0.684	1.132	27.3	72	1.2	25.6
500	0.903	1.543	47.2	72	2.1	32.8

In Table 6, overall, as the number of concurrent users increases, the platform's MinRT, MaxRT, and CPU-MU all continue to increase. When the concurrency reaches 500, the platform MinRT is 0.903 s, MaxRT is 1.54 s within 1 s, and CPU-MU is below 50% within 2 s. The proposed IPHTP can provide stable and reliable services for PHT of college students. The platform can run stably for 72 hours without any faults under different concurrent user loads, with a robust error rate below 2.1%, indicating strong interference resistance in practical application scenarios. Power consumption increases reasonably with the number of concurrent users, reaching only 32.8W at 500 concurrent users, meeting the requirements of embedded deployment. The platform excels in stability, robustness, and power consumption, fulfilling the needs of large-scale physical fitness testing. Then, this study introduces the platform into a certain school and selects two classes in the second year of junior high school as test subjects. Each class has 30 students, who will undergo sit-ups and pull-up tests. This test utilises the built IPHTP to automatically record and analyse students' action data, while comparing it with manual counting results to evaluate the accuracy and efficiency of the platform, as shown in Table 7.

Table 7 Analysis of IPHTP application results

Test item	Class	Total number of automatic counts	Total number of manually counted items	Counting error	Accuracy (%)	t value	p value	p value
Sit-up	1	1,205	1,210	5	99.59	1.236	0.228	$p > 0.05$
	2	1,180	1,185	5	99.58	1.184	0.246	$p > 0.05$
Pull-up	1	450	448	2	99.55	0.972	0.339	$p > 0.05$
	2	435	433	2	99.54	1.025	0.312	$p > 0.05$
Overall	/	3270	3276	6	99.82	1.103	0.275	$p > 0.05$

In Table 7, the results of automatic counting and manual counting are highly consistent, with minimal counting errors. In the sit up test, the error between the total count of automatic counting and manual counting in both classes did not exceed 5, and the accuracy rate exceeded 99.5%. In the pull up test, the error does not exceed 2, and the accuracy remains above 99.5%. Overall, the platform has achieved an accuracy rate of 99.82%, demonstrating extremely high accuracy and reliability. Paired sample t-test shows that there is no significant difference between automatic counting and manual counting of all test items ($p > 0.05$). This proves that the improvement in the accuracy of the platform is statistically significant and is not caused by random errors, further verifying the reliability and effectiveness of the platform. This indicates that the platform can effectively replace manual counting in practical applications, improve the efficiency of PHT, and has broad application prospects.

4 Discussion

The paper aimed to solve the problems of strong subjectivity and low efficiency in traditional PHT. By integrating Fast-OpenPose and BlazePose and combining semantic segmentation technology to build an intelligent testing platform, certain achievements have been made. In terms of HPE, GraphMLP and others have conducted 3DHPE

research using the GraphMLP class architecture (Li et al., 2025). This study was different in its focus on the scene, as it used an improved Fast-OpenPose for 2DHAR research. The improved model performed outstandingly in keypoint matching, with an mAP of 71.2%, an 8.9% increase compared to traditional OpenPose, and an FPS of 20 frames/s. This indicated that the improvement strategy effectively enhanced the accuracy and running speed, and could better meet the real-time and accuracy requirements of PHT. Semantic segmentation played a crucial role in action recognition assistance. SHARP (Mucha et al., 2025) used arm segmentation to assist in 3D hand pose estimation and action recognition. This study utilised an improved DeepLabV3+ to segment pull-up horizontal bars and combined it with BlazePose to detect human keypoints for action classification. The improved DeepLabV3+ achieved 78.5% mIoU and 81.7% FWIoU in the horizontal bar segmentation task after 2,500 iterations, with accurate segmentation results and a reliable basis for action discrimination. Unlike the research on motion classification based on wearable sensor data (Wu et al., 2023) and deep learning for martial arts athlete action recognition (Zhuang et al., 2023), this study was based on computer vision technology and did not require additional equipment. The research method had the advantages of easy operation and low cost, and was more suitable for large-scale PHT scenarios. It should be acknowledged that computer vision methods do pose higher requirements for the testing environment. The core challenges and corresponding solutions are as follows: firstly, the camera angle problem. Side view can easily cause joint occlusion, while the front view can easily ignore the determination of the vertical displacement of movements. The solution is to adopt a dual camera deployment of ‘front + side’ and calibrate the 3D posture of the human body through coordinate fusion. The second issue is lighting interference. Strong light or backlighting can reduce the accuracy of keypoint detection. The platform integrates an adaptive lighting balance algorithm, which adjusts the brightness and contrast of the image through Gamma correction, and can maintain stability within the range of 500–10,000 lux lighting intensity. The third issue is occlusion, as there may be limb overlap during multi-person testing. An improved non-maximum suppression algorithm is used to optimise keypoint matching, and a test area isolation line is set to limit single-person single-area testing. Actual testing has shown that this can control the error rate caused by environmental interference within 2.1%.

However, this study also has limitations. The limited size of the dataset may affect the model’s generalisation ability, and its adaptability in different scenarios and populations needs further verification. Future research can expand the dataset to include more diverse testing scenarios and population data, while exploring the intelligent implementation of more PHT projects, continuously improving and optimising the intelligent testing platform, and promoting PHT to fully move towards the intelligent era.

5 Conclusions

This study focused on the shortcomings of traditional PHT methods, innovatively integrated the advantages of Fast-OpenPose and BlazePose HPE models, and combined semantic segmentation technology to successfully build IPHTP. Automatic counting of sit-ups was achieved through improved Fast-OpenPose and SVM, combined with semantic segmentation and BlazePose, to complete pull-up testing. The experimental results showed that the proposed sit-up testing method, which combines improved Fast-

OpenPose and SVM, improved mAP by 8.9% and FPS by 20 compared to traditional OpenPose through optimisation of the FasterNet and SGE. A pull-up testing method was designed that combines BlazePose and improved DeepLabV3+ to achieve precise collaboration between horizontal bar segmentation and human pose detection, with a segmentation mIoU of 78.5%. A layered architecture intelligent testing platform was constructed, which, combined with TensorRT acceleration technology, achieved real-time and stability for large-scale testing. The improved Fast-OpenPose model, with a 79.3% reduction in parameter count, showed an 8.9% increase in mAP compared to traditional OpenPose, outperforming the latest lightweight models such as MediaPipe Pose. The overall accuracy of the platform was 99.82%, which was 3%-5% higher than the accuracy of existing similar testing systems. However, there is a problem of limited dataset size in the research. In the future, it is necessary to first expand the size of the dataset and include data from testers of different ages, different body types, and complex environments (such as strong light, backlight, and multiple people overlapping) to improve the generalisation ability. Secondly, the coverage of test items should be expanded to make other physical fitness test items, such as the standing long jump and seated forward bend more intelligent. It can also optimise platform edge deployment solutions, develop mobile applications, and improve testing convenience. Finally, detection algorithms for extreme body postures (such as incomplete sit-ups and upward body swings) can be studied to further enhance the robustness of the model.

Declarations

The authors declare no conflicts of interest.

References

- Arya, V. and Maji, S. (2024) 'Enhancing human pose estimation: a data-driven approach with MediaPipe BlazePose and feature engineering analysis', *2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT)*, Vol. 1, No. 1, pp.1–6, DOI: 10.1109/IC2SDT62152.2024.10696215.
- Bhosle, K. and Musande, V. (2023) 'Evaluation of deep learning CNN model for recognition of Devanagari digit', *Artificial Intelligence and Applications*, Vol. 1, No. 2, pp.114–121, <https://doi.org/10.47852/bonviewAIA3202441>.
- Buttfield, A. and Ball, K. (2020) 'The practical application of a method of analysing the variability of within-step accelerations collected via athlete tracking devices', *J Sport. Sci.*, February, Vol. 38, No. 3, pp.343–350, DOI: 10.1080/02640414.2019.1699987.
- Dai, C. and Lu, Y. (2021) 'Improved biological image tracking algorithm of athlete's cervical spine health', *Rev. Bras. Med. Esporte.*, September, Vol. 27, No. 3, pp.274–277, DOI: 10.1590/1517-8692202127032021_0129.
- Ellens, S., Hodges, D. and McCullagh, S.L. (2022) 'Interchangeability of player movement variables from different athlete tracking systems in professional soccer', *Sci. Med. Football*, January, Vol. 6, No. 1, pp.1–6, DOI: 10.1080/24733938.2021.1879393.
- Giles, B., Kovalchik, S. and Reid, M. (2020) 'A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis', *J. Sport. Sci.*, October, Vol. 38, No. 1, pp.106–113, DOI: 10.1080/02640414.2019.1684132.

- Hassanin, M., Khamis, A. and Bennamoun, M. (2025) 'Crossformer3D: cross spatio-temporal transformer for 3D human pose estimation', *Signal Image Video Process*, Vol. 19, No. 8, pp.1–13, DOI: 10.1007/s11760-025-04145-0.
- Hulleck, A.A., Alshehhi, A. and El Rich, M. (2024) 'BlazePose-Seq2Seq: leveraging regular RGB cameras for robust gait assessment', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 32, No. 1, pp.715–1724, DOI: 10.1109/TNSRE.2024.3391908.
- Jayamohan, M. and Yuvaraj, S. (2025) 'A novel human actions recognition and classification using semantic segmentation with deep learning techniques', *Neural Computing and Applications*, Vol. 1, No. 10, pp.1–17, DOI: 10.1007/s00521-024-10962-0.
- Kumar, V. and Worsley, M. (2023) 'Scratch for sports: athletic drills as a platform for experiencing, understanding, and developing AI-driven apps', *AAAI*, September, Vol. 37, No. 13, pp.16011–16016, DOI: 10.1609/aaai.v37i13.26901.
- Li, W., Liu, M. and Liu, H. (2025) 'GraphMLP: a graph MLP-like architecture for 3D human pose estimation', *Pattern Recogn.*, Vol. 158, No. 1, pp.110925–110936, DOI: 10.1016/j.patcog.2024.110925.
- Li, Z., Li, X. and Shi, M. (2021) 'Tracking algorithm of snowboard target in intelligent system', *J. Intell. Fuzzy Syst.*, February, Vol. 40, No. 2, pp.3117–3125, DOI: 10.3233/JIFS-189350.
- Liu, S. and Shi, C. (2024) 'A real-time fall detection system based on MoveNet and LSTM', *International Conference on the Efficiency and Performance Engineering Network*, Vol. 1, No. 170, pp.12–23, DOI: 10.1007/978-3-031-70235-8_2.
- Lopes, J.C., Van-Deste, I. and Lopes, R.P. (2024) 'Exploring human action recognition for rehabilitation game application', *2024 IEEE 12th International Conference on Serious Games and Applications for Health (SeGAH)*, pp.1–8.
- Mucha, W., Wray, M. and Kampel, M. (2025) 'SHARP: segmentation of hands and arms by range using pseudo-depth for enhanced egocentric 3D hand pose estimation and action recognition', in *Proc. International Conference on Pattern Recognition*, Springer, Cham, Vol. 15315, No. 1, pp.178–193, DOI: 10.1007/978-3-031-78354-8_12.
- Ren, H. (2023) 'Sports video athlete detection based on deep learning', *Neural Comput. Appl.*, March, Vol. 35, No. 6, pp.4201–4210, DOI: 10.1007/s00521-022-07077-9.
- Su, S. (2025) 'Research on action matching of skeletal point coordinates and sports teaching application based on open-pose', *Systems and Soft Computing*, Vol. 1, No. 13, pp.200329–200341, DOI: 10.1016/j.sasc.2025.200329.
- Surkov, E., Seredin, O. and Kopylov, A. (2025) 'Standardizing skeletal models for fall detection', *International Conference on Pattern Recognition*, Vol. 15616, No. 1, pp.345–362, DOI: 10.1007/978-3-031-87663-9_29.
- Tristan, E., Antoni, P., Rafael, D.L.T.F., Andreas, T., Sylvain, G., Mario, T., Tiia, K., Norbert, B., Osquel, B. and Reid, A. (2022) 'Application of the athlete biological passport approach to the detection of growth hormone doping', *J. Clin. Endocr. Metab.*, November, Vol. 107, No. 3, pp.649–659, DOI: 10.1210/clinem/dgab799.
- Uo, R., Chen, J. and Zhang, Y. (2025) '3D skeleton aware driver behavior recognition framework for autonomous driving system', *Neurocomputing*, Vol. 613, No. 14, pp.128743–128755, DOI: 10.1016/j.neucom.2024.128743.
- Wang, G., Zeng, X., Lai, G., Zhong, G., Ma, K. and Zhang, Y. (2022) 'Efficient subject-independent detection of anterior cruciate ligament deficiency based on marine predator algorithm and support vector machine', *IEEE J. Biomed. Health.*, October, Vol. 26, No. 10, pp.4936–4947, DOI: 10.1109/JBHI.2022.3152846.
- Wang, T. and Shi, C. (2023) 'Basketball motion video target tracking algorithm based on improved gray neural network', *Neural. Comput. Appl.*, March, Vol. 35, No. 6, pp.4267–4282, DOI: 10.1007/s00521-022-07026-6.
- Wang, Z. and Dong, S. (2024) 'The application of action recognition based on MPP-YOLOv3 algorithm in posture correction', *Informatica*, Vol. 48, No. 10, pp.1–13, DOI: 10.31449/inf.v48i10.5919.

- Woo, Y. and Jeong, H. (2025) 'Exercise assessment based on human pose estimation and relative phase for real-time remote exercise system', *IEEE Access*, Vol. 13, No. 1, pp.53203–53213, DOI: 10.1109/ACCESS.2025.3551834.
- Worsey, M.T.O., Espinosa, H.G., Shepherd, J.B. and Thiel, D.V. (2021) 'One size doesn't fit all: supervised machine learning classification in athlete-monitoring', *IEEE Sensors Letters*, March, Vol. 5, No. 3, pp.1–4, DOI: 10.1109/LSSENS.2021.3060376.
- Wu, W., Wang, R. and Hu, Y. (2023) 'Invisible experience to real-time assessment in elite tennis athlete training: sport-specific movement classification based on wearable MEMS sensor data', *PI Mech. Eng. P-J. SPO*, October, Vol. 237, No. 4, pp.271–282, DOI: 10.1177/17543371211105031.
- Xu, Z., Dai, M. and Zhang, Q. (2025) 'HRPVT: high-resolution pyramid vision transformer for medium and small-scale human pose estimation', *Neurocomputing*, Vol. 619, No. 28, pp.129154–129167, DOI: 10.1016/j.neucom.2024.129154.
- Yang, J. and Kim, R.Y.C. (2024) 'Simple single-person fall detection model using 3D pose estimation mechanisms', *IEEE Access*, Vol. 12, No. 1, pp.174640–174653, DOI: 10.1109/ACCESS.2024.3496992.
- Yuan, S. and Zhou, L. (2025) 'GTA-Net: An IoT-integrated 3D human pose estimation system for real-time adolescent sports posture correction', *Alexandra Eng. J.*, Vol. 112, No. 1, pp.585–597, DOI: 10.1016/j.aej.2024.10.099.
- Zhuang, J., Sun, J. and Yuan, G. (2023) 'Arrhythmia diagnosis of young martial arts athletes based on deep learning for smart medical care', *Neural Comput. Appl.*, July, Vol. 35, No. 20, pp.14641–14652, DOI: 10.1007/s00521-021-06159-4.