

RESEARCH ARTICLE

The Image Processing Using Soft Robot Technology in Fitness Motion Detection Under the Internet of Things

LIN YE¹ AND YINGYING ZHENG² ¹Basic Teaching Department, Jiangsu Shipping College, Nantong 226010, China²College of Physical Education and Health, Wenzhou University, Wenzhou 325035, China

Corresponding author: Yingying Zheng (zjzyy881@sina.com)

ABSTRACT In order to study the application of robotic technology in fitness motion detection, this study develops a modular high accuracy and low-latency Human Gesture Recognition (HGR) system. Firstly, an improved HGR algorithm is introduced, which compresses the HGR model through content extraction and reduces the number of parameters. Methods such as Simulated Annealing and Semi-Supervised Learning are introduced to improve content extraction, further improving the compression of the model. Secondly, a human pose recognition system is constructed based on Deep Learning (DL) and robotics, and each module is recommended. Finally, the effectiveness of the system is verified. The results show that when the accuracy is close to HRNet-32. The inference speed of the model is increased by about 2.4 times, and the model parameters are compressed by about 67%. The data demonstrate the effectiveness of the method. The robot-human posture detection technology can effectively solve human posture recognition in fitness exercises.


INDEX TERMS Human posture recognition, robot technology, simulated annealing, semi-supervised learning, deep learning, motion detection, image processing.

I. INTRODUCTION

With the rapid development of the social economy, people pay more attention to health. It has become an accepted way to improve physical fitness through physical exercise. Exercise has clear benefits [1]. It can regulate the body's metabolism to a certain extent, improve cardiopulmonary function, shape a good body, and even enhance the function of the nervous system. The Internet of Things (IoT) has become another change after the computer and the Internet in the information age. It can be applied to the fusion of network and object through communication perception technologies such as IntelliSense and recognition [2]. The IoT has been widely used in smart industries, agriculture, cities, medical care, and other fields, and the prospects are very broad. The application of IoT technology in human motion monitoring has a long history. Human posture recognition is a popular research direction both in computer vision and in the field

of multimedia. It mainly analyzes images or videos to detect critical points of the human body [3]. Human posture recognition is becoming a research hotspot for many researchers and has relatively high research value and practical significance.

In human pose recognition, Siddiq et al. used a single Red-Green-Blue (RGB) camera to capture some human poses into a spatiotemporal sequence to obtain training data. There are six poses in the test scene. These poses will be clustered using the K-Nearest Neighbor (KNN) method. Each recognized human gesture will be connected to an IoT device that controls the switching patterns of relays in the smart power system. Experimental results show that the system can successfully read each pose using a confusion matrix with good accuracy performance [4]. Zhang et al. proposed and evaluated a robust adaptive hybrid classifier. The classifier combined a pose-based adaptive signal segmentation algorithm with a Multilayer Perceptron Classifier (MLPC) that combines multiple voting methods, compared with single MLPC and static mixture classifiers from three perspectives (classification precision, recall, and F1 score). The results

The associate editor coordinating the review of this manuscript and approving it for publication was Chong Leong Gan .

show that the adaptive mixture classifier model provides slightly larger improvements than static mixture classifiers in classification accuracy and robustness and significantly outperforms single ones [5]. Licciardo et al. proposed a custom hardware design scheme of a Fully Convolutional Neural Network (CNN) to implement an embedded human gesture recognition (HGR) system. The system can achieve high recognition accuracy of lying and sitting postures. By using a limited number of pressure sensors, the optimized hardware implementation allows to stay close to the data source based on edge computing and supports the design of embedded HGR systems [6]. Nadeem et al. pointed out a new method for automatic human pose estimation. The method intelligently recognizes human poses by exploiting salient contour detection, robust body part models and multidimensional cues from whole body contours, and entropy Markov models. Four benchmark datasets are used to achieve better body part detection and higher recognition accuracy. The recognition results outperform existing state-of-the-art methods. The lightweight model implemented can achieve the effect of HGR, and the accuracy of action recognition is higher, considering the characteristics of the hardware platform. Wiley et al. [7] researched computer vision and image processing methods, extending from raw data recording to combining digital image processing, integrating the techniques and ideas of machine learning and computer graphics. The results showed that computer vision helped improving the accuracy of image detection. Maier et al. [8] researched Deep Learning (DL) in medical image processing and reviewed the basics of perceptrons and neural networks, as well as some fundamental theories that were often overlooked. The results suggested that by incorporating DL algorithms, the efficiency of physical simulation, modeling, and reconstruction was improved. Sungeetha [9] conducted research on image input processing technology for human-computer interaction (HCI) and machine learning. They projected the input image in all possible dimensions and visualized it for model checking. The results verified that the model's error, complexity, and recognition rate indicators could meet the actual needs.

With the continuous development of DL technology, the accuracy of HGR has improved, which can already meet the accuracy requirements in most scenarios. In practical applications, models based on DL have large parameters, high hardware requirements, and slow inference speed, so they cannot be used in scenarios that require high inference speed. Previous research work has mainly focused on a single module. There are a few solutions for streamlined HGR systems. A high-precision and low-latency modular HGR system are designed to improve this problem. Knowledge extraction compresses the human pose recognition model and reduces the number of parameters of the model. Furthermore, a novel knowledge distillation method is employed in model training. Compared with other knowledge extraction methods, the model training efficiency of this method is further improved. The main research contribution to the knowledge subject is to detect the image processing process in the fitness movement

through soft robot technology, and analyze the overall effectiveness of the system with High-Resolution Net (HRNet) as the skeleton model through the improvement of human posture recognition algorithm and knowledge extraction. This study has important reference values for knowledge extraction and IoT image recognition. Teacher annealing and semi-supervised learning are introduced to improve knowledge extraction and model compression. The proposed system can provide 3D human key point input for scenes such as robots and provide a solution for applying HGR systems.

II. MATERIALS AND METHODS

A. THE IMPROVED HUMAN POSTURE RECOGNITION ALGORITHM

Human posture recognition technology is mainly divided into traditional and DL methods. The traditional methods mainly focus on feature performance and the direct spatial position relationship of key points, and DL also pays attention to these two dimensions. However, to avoid independent disassembly, DL will model each element in the same network, so using DL methods to solve the problem of human posture recognition has become mainstream [10].

1) HIGH-RESOLUTION NET

HRNet starts from the input high-resolution image information and inserts sub-networks from high-resolution images to low-resolution images layer by layer in the network, forming more parallel stages. As the network depth slowly increases, parallel integrated connections are made to the multi-resolution sub-networks generated from high-resolution images [11], [12]. In this way, image information fusion of multiple scales is achieved. HRNet has a unique parallel multi-resolution sub-network structure, so it has a good effect on the task of human pose estimation [13]. Although HRNet can effectively improve prediction accuracy and achieve optimal results on many public datasets, the model parameters of HRNet are substantial and depend on many computing resources [14], [15]. Therefore, to avoid losing too much computational accuracy, HRNet needs to be improved to reduce its dependence on hardware resources.

2) THE IMPROVED HUMAN POSTURE RECOGNITION MODEL

The structure of the HRNet model is shown in Figure 1. The structure of the model is divided into four stages. Each stage is further divided into two parts, namely the multi-parallel and multi-resolution convolution module and the multi-scale feature fusion module. Each level adds a branch to the previous level. The newly added branch is the product of the convolutional fusion of all feature maps in the previous stage. The resolution is half the resolution of the previous branch. The number of channels is double the number at the last branch. Through comprehensive improvement and recognition, the model has important application value for optimizing and improving the embedded HGR system.

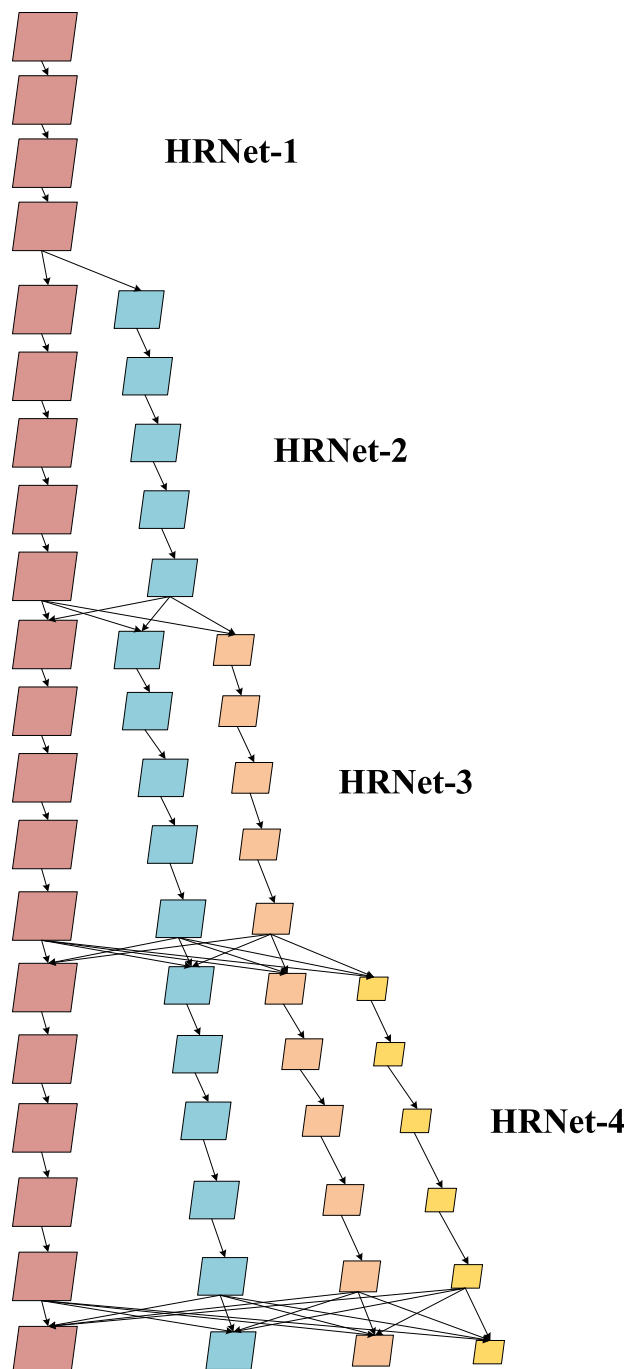


FIGURE 1. The structure of the HRNet model.

In Figure 1, the parameters of the HRNet model are relatively large. The model width of the last three stages is reduced to 18 to reduce the model’s size. Furthermore, this study uses HRNetV2 to improve the model. Feature maps of all resolutions are used for upsampling low-resolution and stitched with high-resolution feature maps. Figure 2 shows the feature map output of the original HRNet model and the improved HRNetV2. Figure 2(a) shows the feature map output of the original HRNet model, and Figure 2(b) shows the feature map output of the improved HRNetV2. The output of the improved feature graph can effectively improve the

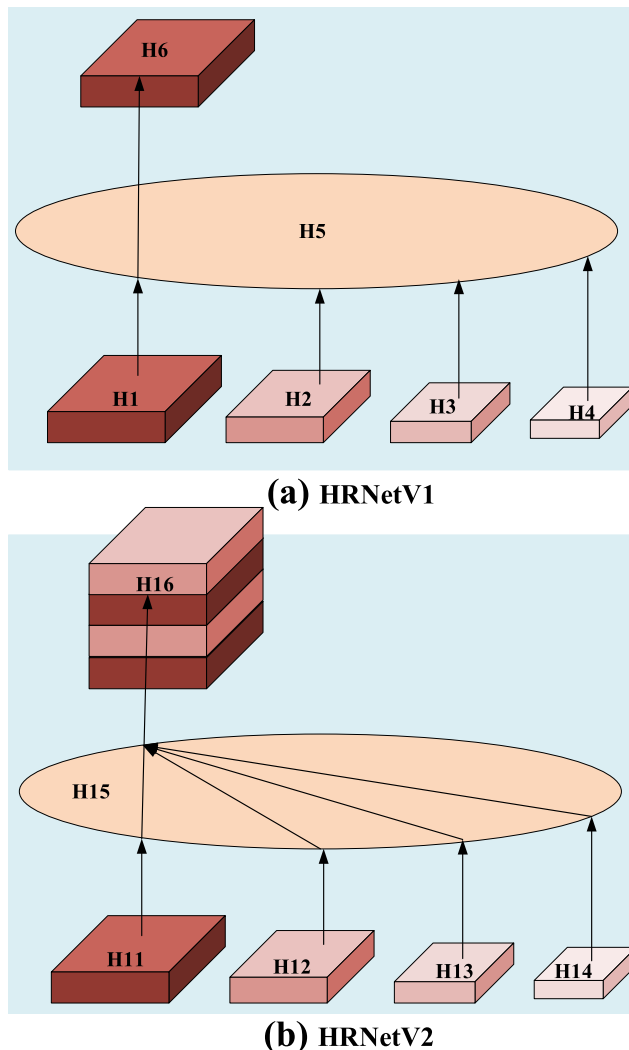


FIGURE 2. HRNet output map (a: HRNetV1; b: HRNetV2).

effect of the model and reduce the precision loss caused by the model parameters.

B. HUMAN POSTURE RECOGNITION ALGORITHM BASED ON KNOWLEDGE DISTILLATION

Currently, the human posture recognition model has defects in the inference speed and is compressed by knowledge distillation. The model parameters are compressed under the premise of similar model accuracy to improve the inference speed of the model. Firstly, knowledge distillation obtains a model with many parameters and high accuracy based on the training data. Then, on the training data, labels are generated for the small model to learn from the large, which in turn successfully transfers the knowledge of the large to the small model [16]. The classic process of distillation training is shown in Figure 3.

The Mean Square Error (MSE) loss function for supervised learning on the task of human posture recognition is shown

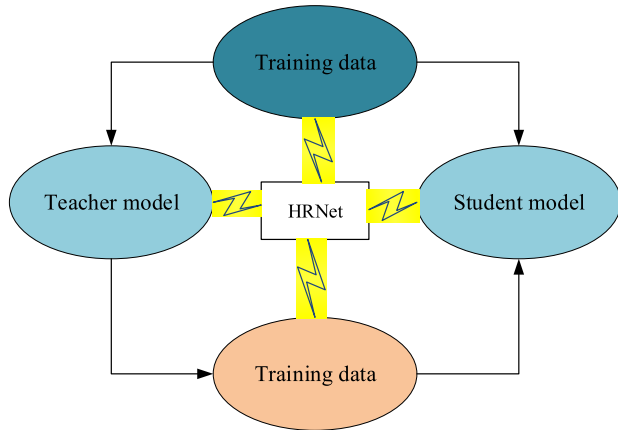


FIGURE 3. The training process of knowledge distillation.

in Eq. (1):

$$L_{SL}(\theta) = \sum_{x \in D} \sum_j \sum_p W(p) \cdot \|H_j(p) - P_j(p; \theta)\|_2^2 \quad (1)$$

In Eq. (1), $W(p)$ indicates whether the label point is visible or not; $H_j(p)$ represents the heat map label of the key point j ; $P_j(p; \theta)$ denotes the output of the model θ .

In the task of human posture recognition, it is difficult to converge only the heat map containing key points as training labels. Therefore, the key point is taken as the center, one pixel is used as the standard deviation, and a two-dimensional Gaussian function is applied to obtain the heat map. The heat map is used as the label learned by the model to guide the model in predicting the output close to the real label coordinates.

The loss function trained by the knowledge distillation on the training data D is shown in Eq. (2):

$$L_{KD}(\theta) = \sum_{x \in D} \sum_j \sum_p W(p) \cdot \|Q_j(p; \theta_T) - P_j(p; \theta)\|_2^2 \quad (2)$$

$Q_j(p; \theta_T)$ refers to the output of the teacher model θ_T ; $W(p)$ represents whether the labeled point is visible or not; $P_j(p; \theta)$ means the output of the model θ .

The loss function of knowledge distillation can be combined with the loss of real label training, and the combined loss is shown in Eq. (3):

$$L_{Combine}(\theta) = (1 - \lambda)L_{SL}(\theta) + \lambda L_{KD}(\theta) \quad (3)$$

λ means the weight of knowledge distillation, which combines the training of real labels and the loss of knowledge distillation.

Although the teacher model can help the model converge to a better position, the model's output has error information. Hence, Eq. (3) is improved, and λ is changed from a fixed value to a linear function $\lambda(\text{Epoch})$. The improved equation is shown in Eq. (4):

$$L_{r, \dots, kmo}(\theta) = (1 - \lambda(\text{Epoch}))L_{ct}(\theta) + \lambda(\text{Epoch})L_{vm}(\theta) \quad (4)$$

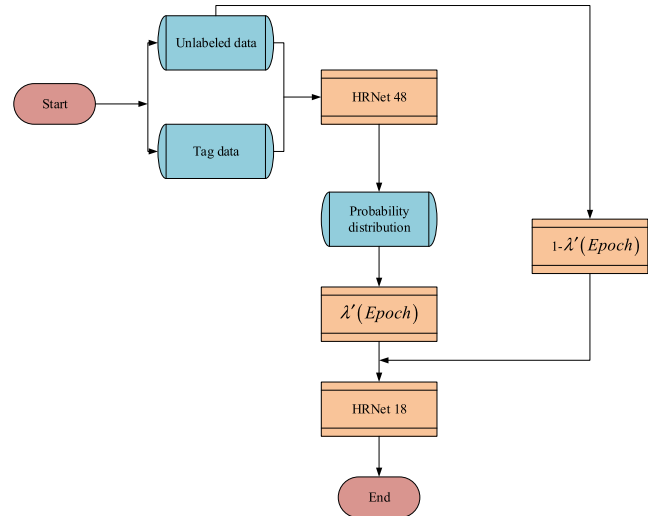


FIGURE 4. The training process of knowledge distillation introduces teacher annealing and SSL.

As the number of training increases, it decreases linearly from 1 to 0. The probability distribution of the teacher model can be used to guide the student model to converge to a better position in the early stage of training, helping student models achieve better results.

Knowledge distillation has great advantages in utilizing unlabeled data. Teacher models can annotate unlabeled data, providing cheap training data for student models [17], [18]. Therefore, knowledge distillation is enhanced with unlabeled data by borrowing ideas from Secure Socket Layer (SSL).

The loss of knowledge distillation by introducing unlabeled data D' from SSL is shown in Eqs. (5) and (6):

$$L_{Semi}(\theta) = (1 - \lambda)L_{SL}(\theta) + \lambda L_{KD}(\theta) + \lambda L_{USL}(\theta) \quad (5)$$

$$L_{USL}(\theta) = \sum_{x \in D'} \sum_j \sum_p W(p) \|Q_j(p; \theta_T) - P_j(p; \theta)\|_2^2 \quad (6)$$

λ demonstrates the weight of knowledge distillation; $Q_j(p; \theta_T)$ represents the output of the teacher model θ_T ; $W(p)$ shows whether the label point is visible; $P_j(p; \theta)$ stands for the output of the model θ .

By learning the output of the teacher model on unlabeled data, the student model can gain more knowledge from the teacher model and converge to output close to the teacher model. The SSL method can improve the effectiveness of the model.

The teacher annealing algorithm and SSL are borrowed. HRNet is used as a skeleton model to compress through knowledge extraction, improve the inference speed of the model, and provide help for the further promotion of high-precision human pose recognition models [19], [20]. The training process of teacher annealing algorithm and SSL is introduced, as shown in Figure 4:

In Figure 4, the selected teacher model is the HRNet-48 model with the input image dimension of 256×192 , and the student model is the HRNet-18 with the input image

dimension of 256×192 , as shown in Eq. (7).

$$\lambda'(Epoch) = \begin{cases} 1, & Epoch < E \\ 0, & Epoch \geq E \end{cases} \quad (7)$$

E stands for a certain round. When $Epoch < E$, the model learns the output of the teacher model; when $Epoch \geq E$, the model learns ground truth labels. After the unlabeled and the labeled data are subject to a probability distribution, the model is trained. The target result value is obtained after the training is completed.

C. HUMAN POSTURE RECOGNITION SYSTEM BASED ON DL AND ROBOTICS

1) ANALYSIS OF THE SCENARIOS

Human posture recognition has always been a hot task in computer vision. With the continuous improvement of human posture recognition technology, recognition accuracy has also been improved. Meanwhile, with the continuous progress of robot technology, many application scenarios of robots have emerged, such as bionic robots, HCI, etc. The robot system needs human posture recognition information to perform behavior analysis or control robots. Therefore, the requirements are high, but there are still difficulties [21], [22].

(1) The accuracy of human posture recognition needs to be improved. The basis for controlling the robot is accurate human posture recognition results. When human posture recognition is accurate, the robot can obtain accurate coordinate information to move smoothly and accurately [23].

(2) The speed of human posture recognition needs to be improved. In some scenarios that require robot control, the requirements for the robot are very high. The robot needs to respond in real time, which requires the human posture recognition system to generate recognition results quickly. Currently, the accuracy of human posture recognition based on the DL model can meet the requirements of some scenarios. Still, the parameter network of the DL model is very large, which slows down the inference speed and limits the application scenarios of the human posture recognition model [24].

Moreover, the robot system runs in three-dimensional (3D) space. It needs to generate 3D coordinates for use, but the current human posture recognition system is more concerned with generating two-dimensional (2D) coordinates. Therefore, based on these difficulties, a human posture recognition system by DL and robotics is designed.

2) DESIGN OF SYSTEM

The purpose of the HGR system is to detect the key points of the human body. Robotic systems require continuous 3D coordinates of key points on the human body. Therefore, the designed HGR system takes video as input and divides HGR into four steps, namely video frame extraction, target detection, 2D HGR, and the conversion module of human gesture 3D. The designed human target detection system is shown in Figure 5.

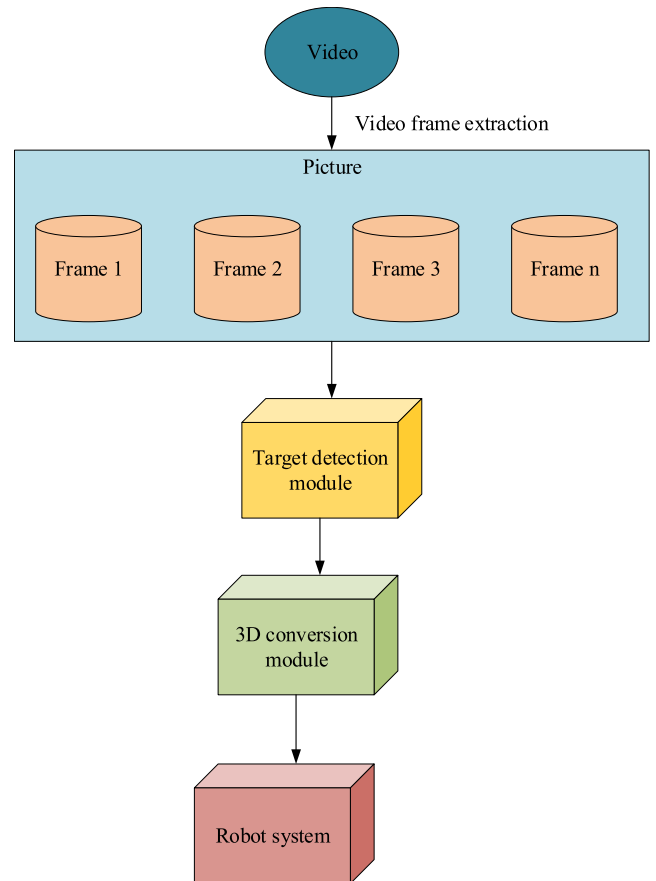


FIGURE 5. The recognition system of human posture.

In Figure 5, the extraction of image features needs to go through different structural frameworks. Through image feature extraction and recognition, target detection and action gesture recognition, image recognition, and application in the system are carried out. Therefore, the steps and processes of HGR are as follows:

(1) Extraction of video frames. The input of both the object detection module and the 2D human pose recognition module are images. So, frames are extracted from the video and divided into frame-by-frame images. Furthermore, the HGR system has a certain time overhead, and it is necessary to appropriately reduce the sampling frequency of video frame extraction to realize real-time processing.

(2) Target detection. Through the target detection module, the system selects the person's location from the image and provides location information for subsequent modules [25].

(3) Three-dimensional human gesture recognition module.

Although human gesture recognition technology has been continuously improved, most previous research methods are based on the residual network for 3D human gesture recognition. They tend to only rely on in-depth features extracted from residual networks without further enhancing feature information. Therefore, these methods are unsuitable for complex and changeable human pose recognition. Compared with the actual human pose, there is a large error in joint positioning. This work integrates multi-scale features and

attention distribution and enhances feature information to study the network structure design based on the residual network to achieve higher human pose recognition accuracy in 3D. The human pose detection module is the core of the design system. After the target detection module obtains the target position information of the person, the human pose recognition module needs to identify the key points in the image [26]. In this module, the basic skeleton model selected for human pose recognition is the HRNet model.

(4) Human posture 3D conversion module. Since the robot system cannot directly use 2D human key points, the coordinates of 2D human key points need to be converted by a 3D human pose transformation module. The transformed 3D human key point coordinates are transferred to the robot system. [27]. Continuous video is selected as an input, and the 2D human key point coordinates are converted into 3D human key point coordinates through the VideoPose3D model. The timing information is added to the human posture 3D conversion module.

D. EVALUATION INDICATORS

(1) Object Keypoint Similarity (OKS) calculates the similarity between the real and the human key points of the prediction model, which is the evaluation index of the key point detection of human poses commonly used at present. The calculation of OKS is shown in Eq. (8):

$$OKS_p = \frac{\sum_i \exp\left(-d_{pi}^2 / 2S_p^2 \sigma_i^2\right) \delta(v_{pi} > 0)}{\sum_i \delta(v_{pi} > 0)} \quad (8)$$

In Eq. (8), i indicates the i -th key point; p represents the person numbered p in the current picture. d_{pi} means the Euclidean distance between the predicted and true value of the i -th key point of the p -th person. σ_i shows the standard deviation between manually labeled and true values in all samples. S_p stands for the scale factor of the p -th person. $\delta(v_{pi} > 0)$ demonstrates the visible key points, where 1 means that the key points are visible, and 0 means that the key points are not visible.

(2) Average Precision (AP): The calculation of AP is shown in Eq. (9):

$$AP = \frac{\sum_p \delta(OKS_p > T)}{\sum_p 1} \quad (9)$$

T represents the threshold. When $OKS > T$, the predicted value is considered to be accurate.

(3) Mean Average Precision (MAP). The main detection metric used in this work is MAP. The threshold value T is set to different values. Then, the average value of AP under different thresholds is obtained, and the MAP can be obtained. The value range of the threshold T is shown in Eq. (10):

$$T \in [0.5 : 0.05 : 0.95] \quad (10)$$

(4) Giga Floating-point Operations Per Second (GFLOPs). GFLOPs are used to assess the capabilities of computer sys-

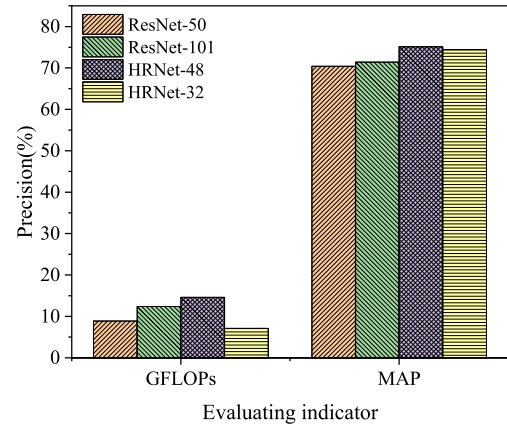


FIGURE 6. Comparison of HRNet and ResNet models on the COCO dataset.

tems. System simulation evaluates the memory architecture and cache memory consistency.

E. EXPERIMENTAL PARAMETERS AND DATASETS

HRNet is taken as the skeleton model, and HRNet-48 model is the teacher model. The HRNet-18 model is a learning model, and the effect of HRNet-32 is used as a reference for the experiments. Among them, 18, 32, and 48 refer to the width of the high-resolution subnetworks in the last three stages. The main experimental object is the HRNet-18 model, which implements the algorithm based on the Pytorch framework. The ratio of the height and width of the human detection frame is fixed at 4:3. The size of the selected part of the frame is adjusted to 256×192 . Using the Adam optimization operator, the learning rate is $1e-3$, which is reduced to $1e-4$ in the 170-th round and $1e-5$ in the 200-th round for 210 rounds. The training batch is 128, and training is performed on four modules of Bvidiap100. Bvidiap100 is a data processing accelerator in the research center [28]. The research technology has been verified through model training and iteration by evaluating and comparing different models such as HRNet-48, HRNet-18 and HRNet-32.

The data set used the Common Objects in Context (COCO) data set, including annotating multiple tasks such as target recognition and detection. Among them, there are about 200,000 standard images and about 250,000 human key points [29], [30]. The 2017 COCO dataset uses 118,000 images as the training set, 41,000 images as the test set, and 5,000 images as the validation set.

III. RESULTS AND DISCUSSION

A. COMPARATIVE ANALYSIS OF HRNet AND ResNet MODELS ON THE COCO DATASET

Figure 6 shows the comparison results of the parameters and accuracy of the HRNet model and the ResNet model on the COCO dataset of human posture recognition.

In Figure 6, when the parameters of the model and the amount of calculation are close. The HRNet model improves the accuracy of the human posture recognition task to a large extent. However, the calculation amount of the HRNet model

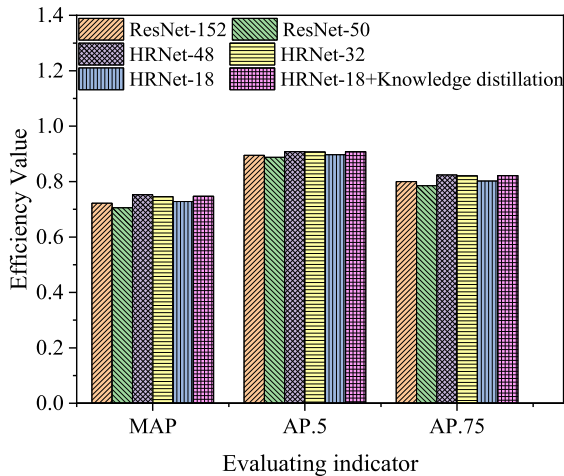


FIGURE 7. Comparison of overall effectiveness.

is relatively large, and the GFLOPs of the HRNet-32 model has reached 7.12. The inference speed restricts the practical application scenarios of the HRNet model. Therefore, it is necessary to compress the HRNet model through teacher annealing and SSL to reduce the model's parameters and maintain the original accuracy of the model.

B. COMPARISON OF OVERALL EFFECTIVENESS

The effect of the overall method on the HRNet-18 student model is shown in Figure 7. HRNet-18 means the result obtained by training directly on the labeled training set. HRNet-18+ knowledge distillation means the effect of the overall text method on HRNet-18.

In Figure 7, compared with the ResNet model, the HRNet model has great advantages in the task of human posture recognition. Under the level of the same parameter, the accuracy of the HRNet model is much improved, even if the parameter amount is relatively small. The HRNet-18 model has a MAP of 0.728. This model outperforms the ResNet-152 model with a relatively large number of parameters. In Figure 7, the calculation is conducted according to the estimated values of different model evaluation indicators. When the AP.5 value of the HRNet-32 model is 0.9, the AP.75 value of the model is 0.78, and the MAP value is 0.72. The running speed of the model is obtained by averaging the AP.5 value and AP.75 value of different models, which is calculated according to equation (9) above. The number of parameters compressed is calculated by MAP value. MAP values of different models will fluctuate. Therefore, the HRNet model is chosen as the skeleton. When the accuracy is close to HRNet-32, the inference speed of the model is increased by about 2.4 times, and the number of parameters is compressed by about 67%. The results demonstrate the effectiveness of the method.

C. COMPARISON OF STUDENT MODELS WITH DIFFERENT PARAMETERS

The knowledge distillation method is tried for compression onto a smaller HRNet-18-small model, and the results are shown in Figure 8.

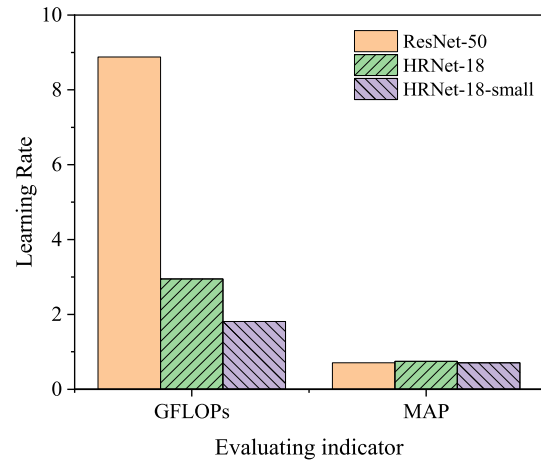


FIGURE 8. Comparison of student models with different parameters.

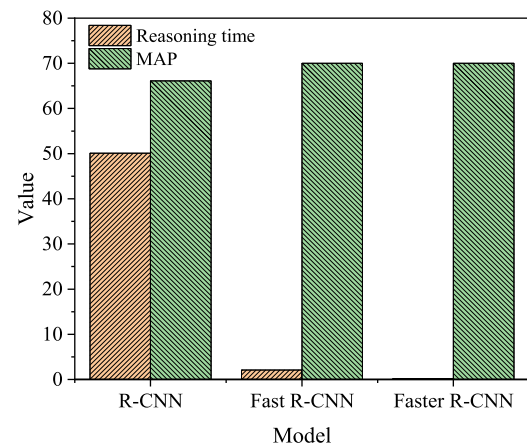


FIGURE 9. Comparison of results of different models.

Figure 8 shows that the MAP value of the HRNet-18-small model after knowledge extraction is 0.729, the MAP value of the HRNet-18 model is 0.747, and the MAP value of the ResNet-50 model is 0.704. Although the HRNet-18-small model performs better than the ResNet-50 model, it is not as good as the HRNet-18 model. It means that the task of human posture recognition is relatively complex, and when the model is too small, it is difficult to fit the training data. The models in Figure 6, Figure 7 and Figure 8 are compared. The results show that the learning ability of the HRNet-18-small student model is worse than other models. When the GFLOPs value of the ResNet-50 model is 9, the GFLOPs value of the HRNet-18-small model is only 1.8, and the learning efficiency value of the student model is relatively low. When the learning efficiency value of the ResNet-50 model is 0.9, the learning efficiency value of the student model is only 0.68. In addition, when the learning ability of the student model is too poor, it is very difficult to improve the effectiveness of the small model through knowledge distillation and make it achieve the effect of the large model.

D. ANALYSIS OF OBJECT DETECTION MODULE

The human target detection module selects the Region-Convolutional Neural Networks (Faster R-CNN) based on

ResNet-50 for comparison with other models. The results are shown in Figure 9.

In Figure 9, the inference time of a single image of the R-CNN, Fast R-CNN, and Faster R-CNN models is 50.1s, 2.1s, and 0.2s, respectively; MAP is 66.1, 70, and 70, respectively. In contrast, the inference time of a single image of the Faster R-CNN model has been greatly improved. The accuracy is relatively high, so the Faster R-CNN model is selected as the basic model used by the target detection module.

IV. CONCLUSION

The application of IoT technology in human motion monitoring has a long history, and human posture recognition has a relatively popular research direction in computer vision and multimedia. It mainly analyzes images or videos to realize the key points of the human test. By introducing DL technology, a high-precision human posture recognition system is constructed [31]. The human posture recognition model is compressed through knowledge distillation, thereby reducing the number of parameters of the model. Some methods, such as teacher annealing and SSL, are introduced to improve knowledge distillation, thereby further improving the effect of model compression. In the case of similar accuracy, the model effect is greatly improved. The results show that when the accuracy is close to HRNet-32 on the COCO data set. The inference speed is increased by about 2.4 times. The compression of model parameters is increased by about 67%. Experimental data demonstrates the effectiveness of the proposed method.

REFERENCES

- [1] J. Yin, "The method of table tennis players posture recognition based on a genetic algorithm," *Int. J. Biometrics*, vol. 13, nos. 2–3, pp. 243–257, 2021.
- [2] Z. Hong, M. Hong, N. Wang, Y. Ma, X. Zhou, and W. Wang, "A wearable-based posture recognition system with AI-assisted approach for healthcare IoT," *Future Gener. Comput. Syst.*, vol. 127, pp. 286–296, Feb. 2022.
- [3] Z. Chen, X. Chen, Y. Ma, S. Guo, Y. Qin, and M. Liao, "Human posture tracking with flexible sensors for motion recognition," *Comput. Animation Virtual Worlds*, vol. 32, no. 5, p. e1993, Sep. 2021.
- [4] M. Siddiq, I. Wibawa, and M. Kallista, "Integrated Internet of Things (IoT) technology device on smart home system with human posture recognition using kNN method," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1098, no. 4, 2021, Art. no. 042065.
- [5] S. Zhang and V. Callaghan, "Real-time human posture recognition using an adaptive hybrid classifier," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 2, pp. 489–499, Feb. 2021.
- [6] G. D. Licciardo, A. Russo, A. Naddeo, N. Cappetti, L. Di Benedetto, A. Rubino, and R. Liguori, "A resource constrained neural network for the design of embedded human posture recognition systems," *Appl. Sci.*, vol. 11, no. 11, p. 4752, May 2021.
- [7] V. Wiley and T. Lucas, "Computer vision and image processing: A paper review," *Int. J. Artif. Intell. Res.*, vol. 2, no. 1, pp. 29–36, 2018.
- [8] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift Für Medizinische Physik*, vol. 29, no. 2, pp. 86–101, May 2019.
- [9] A. Sungeetha, "3D image processing using machine learning based input processing for man-machine interaction," *J. Innov. Image Process.*, vol. 3, no. 1, pp. 1–6, Feb. 2021.
- [10] Q. Wan, H. Zhao, J. Li, and P. Xu, "Hip positioning and sitting posture recognition based on human sitting pressure image," *Sensors*, vol. 21, no. 2, p. 426, Jan. 2021.
- [11] S. Ahmed, T. Hossain, O. B. Hoque, S. Sarker, S. Rahman, and F. M. Shah, "Automated COVID-19 detection from chest X-ray images: A high-resolution network (HRNet) approach," *Social Netw. Comput. Sci.*, vol. 2, no. 4, pp. 1–17, Jul. 2021.
- [12] S. Seong and J. Choi, "Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates," *Remote Sens.*, vol. 13, no. 16, p. 3087, Aug. 2021.
- [13] K. Sengupta and P. R. Srivastava, "HRNET: AI-on-edge for mask detection and social distancing calculation," *Social Netw. Comput. Sci.*, vol. 3, no. 2, pp. 1–15, Mar. 2022.
- [14] M. Hidaka, D. Matsuoka, D. Sugiyama, K. Murakami, and S. Kako, "Pixel-level image classification for detecting beach litter using a deep learning approach," *Mar. Pollut. Bull.*, vol. 175, Feb. 2022, Art. no. 113371.
- [15] R. Tao, Y. Zhang, L. Wang, Q. Liu, and J. Wang, "U-high resolution network (U-HRNet): Cloud detection with high-resolution representations for geostationary satellite imagery," *Int. J. Remote Sens.*, vol. 42, no. 9, pp. 3511–3533, May 2021.
- [16] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *Int. J. Comput. Vis.*, vol. 129, pp. 1789–1819, Mar. 2021.
- [17] D. Martínez-Muñoz, J. V. Martí, J. García, and V. Yepes, "Embodied energy optimization of buttressed earth-retaining walls with hybrid simulated annealing," *Appl. Sci.*, vol. 11, no. 4, p. 1800, Feb. 2021.
- [18] P. Van Molle, C. De Boom, T. Verbelen, B. Vankeirsbilck, J. De Vylder, B. Diricx, P. Simoens, and B. Dhoedt, "Data-efficient sensor upgrade path using knowledge distillation," *Sensors*, vol. 21, no. 19, p. 6523, Sep. 2021.
- [19] M. Farajzadeh-Zanjani, E. Hallaji, R. Razavi-Far, M. Saif, and M. Parvania, "Adversarial semi-supervised learning for diagnosing faults and attacks in power grids," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3468–3478, Jul. 2021.
- [20] X. Hao, J. Liu, Y. Zhang, and G. Sanga, "Mathematical model and simulated annealing algorithm for Chinese high school timetabling problems under the new curriculum innovation," *Frontiers Comput. Sci.*, vol. 15, no. 1, pp. 1–11, Feb. 2021.
- [21] Z. Fan, X. Hu, W.-M. Chen, D.-W. Zhang, and X. Ma, "A deep learning based 2-dimensional hip pressure signals analysis method for sitting posture recognition," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103432.
- [22] X.-B. Fu, S.-L. Yue, and D.-Y. Pan, "Camera-based basketball scoring detection using convolutional neural network," *Int. J. Autom. Comput.*, vol. 18, no. 2, pp. 266–276, Apr. 2021.
- [23] P. Zhu, J. Zhu, X. Xue, and Y. Song, "Stretchable filler/solid rubber piezoresistive thread sensor for gesture recognition," *Micromachines*, vol. 13, no. 1, p. 7, Dec. 2021.
- [24] Z. Yue, Y. Zhu, J. Xia, Y. Wang, X. Ye, H. Jiang, H. Jia, Y. Lin, and C. Jia, "Sponge graphene aerogel pressure sensors with an extremely wide operation range for human recognition and motion detection," *ACS Appl. Electron. Mater.*, vol. 3, no. 3, pp. 1301–1310, Mar. 2021.
- [25] Z. Liu, K. Wan, T. Zhu, J. Zhu, J. Xu, C. Zhang, and T. Liu, "Superelastic, fatigue-resistant, and flame-retardant spongy conductor for human motion detection against a harsh high-temperature condition," *ACS Appl. Mater. Interfaces*, vol. 13, no. 6, pp. 7580–7591, Feb. 2021.
- [26] C. Liu, Y. Du, Y. Ge, D. Wu, C. Zhao, and Y. Li, "New generation of smart highway: Framework and insights," *J. Adv. Transp.*, vol. 2021, pp. 1–12, Dec. 2021.
- [27] Z. Zhang, C. Wang, W. Qiu, W. Qin, and W. Zeng, "AdaFuse: Adaptive multiview fusion for accurate human pose estimation in the wild," *Int. J. Comput. Vis.*, vol. 129, no. 3, pp. 703–718, Mar. 2021.
- [28] M. Ahmad, I. Ahmed, and G. Jeon, "An IoT-enabled real-time overhead view person detection system based on cascade-RCNN and transfer learning," *J. Real-Time Image Process.*, vol. 18, no. 4, pp. 1129–1139, Aug. 2021.
- [29] A. Iazzi, M. Rziza, and R. O. H. Thami, "Fall detection system-based posture-recognition for indoor environments," *J. Imag.*, vol. 7, no. 3, p. 42, Feb. 2021.
- [30] J. Liu, S. Tsujinaga, S. Chai, H. Sun, T. Tateyama, Y. Iwamoto, X. Huang, L. Lin, and Y.-W. Chen, "Single image depth map estimation for improving posture recognition," *IEEE Sensors J.*, vol. 21, no. 23, pp. 26997–27004, Dec. 2021.
- [31] C. Yu, Z. Li, D. Yang, and H. Liu, "A fast robotic arm gravity compensation updating approach for industrial application using sparse selection and reconstruction," *Robot. Auto. Syst.*, vol. 149, Mar. 2022, Art. no. 103971.

• • •