

Analysis of Artificial Intelligence Technology and Its Application in Improving the Effectiveness of Physical Education Teaching

Rui Guo, Henan University of Technology, China*

ABSTRACT

To promote the construction of public physical education online courses in colleges and universities and the evaluation of the effectiveness of course teaching, this article combines 3D reconstruction techniques in computer vision to construct a set of human body shape reconstruction models and apply them to physical training exercises and teaching effectiveness assessment tasks. Specifically, first, the joint point location information of the human body in the input image is extracted using the human skeleton analysis algorithm, and modeling the foreground and background pose information of the target region using the Pix2Pix image transformation algorithm; second, multi-scale features such as nodal location features, foreground and background features, high-resolution detail features, and low-resolution global features are fused and the extracted multi-scale features are also decoded with the help of pixel-aligned implicit functions to generate a 3D model of the human body representing the human form.

KEYWORDS

Artificial Intelligence Technologies, Assessment of the Effectiveness of Physical Education, Body Reconstruction, Human Skeleton Analysis Algorithm

INTRODUCTION

Since the beginning of the new century, rapidly changing internet technology has penetrated all aspects of the economy and society, and the internet has gradually developed into an important hub to promote the development of global economization and social progress. Internet plus physical education evaluation (Internet+) is a product of the times (Ding et al., 2020). The concept of Internet+ is fully utilized in the application of physical-education evaluation, which constantly plays a role in intelligence, data integration, and combing. The concept of Internet+ is becoming more and more popular (Lonsdale et al., 2019). Combining different fields in education with the new achievements of the internet and exploring new forms of physical-education teaching evaluation will help to continuously enhance its quality and form a new trend in the development of teaching evaluation carried out in the context of the network environment.

DOI: 10.4018/IJWLTT.335115

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

In recent years, with the outbreak of the Covid-19 epidemic, online teaching has become mainstream (Shi et al., 2021). To strengthen learning for students at home and effectively guarantee the synchronization of physical education in schools, online teaching has become an important part of the teaching content in colleges and universities (Gu, 2022). The vigorous development of online teaching and learning activities also requires corresponding teaching and learning evaluation activities to support and guide the development trends of public physical education. In the context of Internet+, the evaluation of public physical education in colleges and universities is also facing new problems and indicators; how to make the design of online courses more complete and how to guarantee the effect of online physical education are urgent problems to be solved.

Along with the continuous updating of the teaching system, the content of physical-education evaluation in colleges and universities also needs to keep up with the times. At the same time, with the continuous deepening of education, many problems of the original teaching system have been gradually revealed. An important tool to measure the milestones and overall effect of teaching is an effective teaching evaluation, which can not only reflect the strengths and weaknesses of physical-education teachers but also identify students' weaknesses in the learning process and improve them. This is a great help to improve the teaching of public physical education as well as the learning efficiency of students.

This paper combines Internet+, artificial intelligence, and other related resources to conduct a detailed study on the content of the teaching evaluation index system to provide relevant advice for the reform and development of college physical education and to promote the operation of advanced and forward-looking reform of college physical education in the post-epidemic era. The purpose of this study is to provide suggestions for the construction and application of public online physical-education courses in colleges and universities.

The combination of human body shape reconstruction models and deep learning-based behavior detection has been applied to various industries and achieved certain results. It is feasible to apply deep learning-based behavior-detection methods and human body shape reconstruction models to video data to perform online physical education teaching evaluation. Currently, there is still a lack of research in this area. Therefore, this article combines three-dimensional reconstruction technology and computer vision to build a set of human body shape reconstruction models and apply them to physical training and teaching effectiveness evaluation tasks.

RELATED WORKS

To solve the specific problems in the development of education today, it is necessary to improve the educational function system, meet the actual needs of education stakeholders, promote the optimization and integration of teaching resources, strive to bring educational benefits to people in underdeveloped areas, and achieve modern educational equity. In 2019, relevant national departments gave instructions to accelerate the construction and development of the Internet+ education system (Uysal & Balci, 2018). Encourage regions to combine education with the development of online information. In the modern Internet environment, we should give full play to the advantages and role of Internet plus, lead the efficient evaluation of physical education teaching, and promote the reconstruction of social and campus education concepts (Feng, 2020).

Research Related to the Evaluation of Physical Education

The emergence of online courses and online teaching has made the content and teaching resources of public physical-education courses in colleges and universities more abundant and diversified. The construction of new teaching evaluation indexes can provide constructive ideas for physical education teaching reform and clear teaching and learning ideas for public physical-education teachers and students and can play a positive role in monitoring the effectiveness of teachers and the learning outcomes of students (Reece et al., 2021).

The evaluation of physical-education teaching has the functions of motivation and feedback, but many scholars have found that the current evaluation of physical education in colleges and universities has problems such as insufficient evaluation function, incomplete evaluation content, and unscientific evaluation method after investigation and research (Bedard et al., 2019). Some scholars also believe that university sports evaluation is a complex management activity, and any influencing factor may affect the final evaluation results. They have analyzed the influencing factors of sports evaluation, mainly due to people's insufficient understanding of university sports evaluation, unscientific evaluation standards and methods (Feng, 2020)[.

In the construction of the physical education teaching evaluation index system, some scholars have proposed coming up with a new system model and assigning weight coefficients to each index. For example, based on previous research and recent teaching evaluations conducted by oneself, the indicators for physical education teaching evaluation are designed into 4 primary indicators, 11 secondary indicators, and 32 tertiary indicators, and then experts are sought to determine the evaluation factors and their weights (Abdulhammed et al., 2019a).

Abdulhammed et al. (2019b) collected a large amount of relevant information on classroom teaching skill evaluation both domestically and internationally, and ultimately determined a set of teaching skill evaluation index systems that include 6 primary indicators and 25 secondary indicators. Through the exploration of physical education teaching evaluation practice in China, Aldweesh et al. (2020) pointed out that there are six major development trends in physical education teaching evaluation in China: interactive evaluation subjects, emphasis on process evaluation, emphasis on personalized evaluation, diversified development of evaluation content, combination of qualitative and quantitative evaluation, and multi-dimensional evaluation system.

By summarizing the current teaching-evaluation situation, Safi et al. (2021) argue that teaching-evaluation reform trend is manifested as practical evaluation, multiple evaluation, and process-oriented. And pointed out the trend of future physical education teaching models in universities, which is the diversification of educational goals. Schools should achieve an organic combination of health and lifelong physical education, cultivate students' sports innovation ability, and establish diversified three-dimensional evaluation indicators.

At this stage, the evaluation of physical education has become a hot issue of academic concern, and after various scholars' in-depth research, many effective results have been achieved, which provides a certain theoretical basis for future research. But physical education is a complex process, and the reform of the teaching system is a process of continuous practice and exploration. How to evaluate physical education in a comprehensive, diversified, and practice-integrated manner is still an area that needs continued research.

Deep Learning–Based Behavior Detection

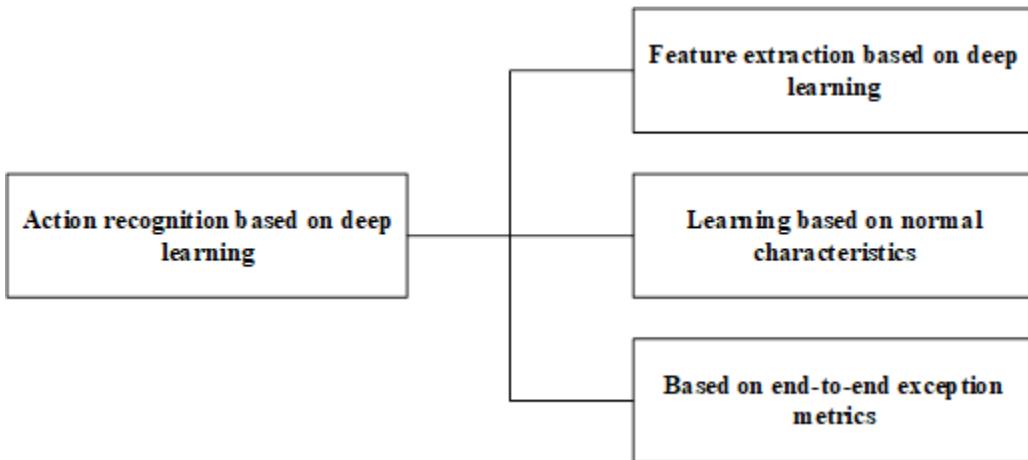
In the general process of anomalous behavior detection, deep learning–based anomaly detection can be divided into three major categories based on whether feature extraction and anomaly detection are separated: deep learning–based feature extraction, constant feature–based learning, and end-to-end anomaly metric learning. The classification of abnormal action detection based on deep learning is shown in Figure 1.

Deep Learning–Based Feature Extraction

In the deep learning–based feature extraction framework, the two modules of feature extraction and anomaly detection are independent of each other. The main work of such methods is to reduce the dimensionality of high-dimensional data using deep-learning techniques to obtain a low-dimensional feature representation.

The main idea of the feature-extraction module is to extract features with the help of pre-trained models. Chen et al. (2019) use VGG-f, a model pre-trained on the ILSVRC benchmark, to extract high-level features from video frames, which are then passed through an anomaly detection module

Figure 1. Classification of abnormal action detection based on deep learning



and the extracted features are fed to a single classification support vector machine for anomalous behavior detection. Similarly, Sun et al. (2019) propose the extraction of behavioral and appearance features using a trained VGG-f model, which is also trained on the ILSVRC benchmark, and then construct an anomaly-scoring model based on the extracted features. Another idea for the deep learning-based feature extraction method is to train a deep-feature extraction model specifically. Long et al. (2019) propose a new method based on stacked denoising autoencoders, where the low-level visual information consists of static image blocks and dynamic motion fields represented by optical streams. Finally, the three feature representations are input to three one-class support vector machines to predict anomaly scores and perform post-fusion for anomaly detection.

Learning Based on Constant Features

In constant feature-based learning, the two modules, feature extraction and anomaly detection, are in some form dependent on each other and are not completely separated. In recent years, constant feature-based learning has been widely used for video anomalous behavior detection, and such methods can be broadly divided into two types of models: reconstruction models and prediction models. Autoencoder (AE) networks are used very frequently in anomalous-behavior detection for constant-feature learning, and in the setting of unsupervised methods, AE is a powerful tool for processing high-dimensional data. Kim & Kwak (2022) designed a generic framework where the overall architecture is a deep autoencoder with parameter-density estimation used to improve the memory efficiency and minimize the reconstruction error of normal samples, where the parameter-density estimation learns the probability distribution of potential representations through an autoregressive process joint optimization of reconstruction errors and probability distributions to finally obtain anomaly scores.

Chiew et al. (2019) use two-dimensional human skeletal trajectories to detect abnormal events associated with human behavior. First, the human-pose map is calculated from the input video sequence by pose recognition. Then the pose map is encoded using Spatio-Temporal Graph Auto-Encoder (ST-GCAE), and finally the potential vectors of the encoded output are clustered by a deep clustering layer to obtain anomaly scores. Compared with appearance features, skeletal features are more compact, strongly structured, semantically rich, and highly descriptive of the body's movements.

Jain & Gupta (2018) propose an anomaly-detection framework—based on future frame prediction—that uses U-Net networks as generators to predict the next frame and error analysis of the predicted next-frame results with the real next-frame results. To generate high-quality images, the optical flow of predicted and real frames is also extracted using the flow network, and error analysis

is performed. Finally, discriminators in generative adversarial networks are used to discriminate those with large errors as anomalous behaviors.

Heidari et al. (2023) combine the reconstruction model with the prediction model to propose a new model of Message-Passing Encoder-Decoder Recurrent Neural Network (MPED-RNN). The model consists of two RNNs for extracting global and local features, respectively, with the two branches processing their skeleton data and interacting through cross-branch messaging to extract the most compact normal behavior from the training data and efficiently detect abnormal events. Each branch is composed of three RNNs: encoder, reconstruction decoder, and prediction decoder, which generate past and future poses using reconstruction and prediction branches, respectively, and the anomaly score is ultimately determined by the reconstruction and prediction errors. This is the first paper to use skeletal pose for abnormal behavior detection.

End-to-End Anomaly Metric-Based Metric Learning

In end-to-end anomaly metric-based learning, feature extraction and anomaly detection are then fully unified. The method uses neural networks to directly compute anomaly scores, i.e., the raw data is used as input to directly learn and output anomaly scores to detect anomalous behavior.

Granjal et al. (2018) propose an end-to-end unsupervised video anomaly detection method with self-training depth-ordered regression, which can unify two modules, feature extraction and anomaly detection, to achieve end-to-end results. The original video is first detected to generate a pseudo-normal frame set and a pseudo-anomaly frame set, which are used as input to an end-to-end anomaly score learner to train a pre-trained model ResNet-50 and a fully connected network. Finally, the trained model is used to recalculate the anomaly scores of all frames. Chen et al. (2021) propose a deep multi-instance learning framework that first divides normal and abnormal videos into multiple video segments. Then each video is represented as a package and each video clip as an instance in the package. Next, the C3D features in the video clips are extracted. Finally, a fully connected neural network is trained using a new ranking loss function that computes the highest scoring instances in the package as a way to locate anomalies.

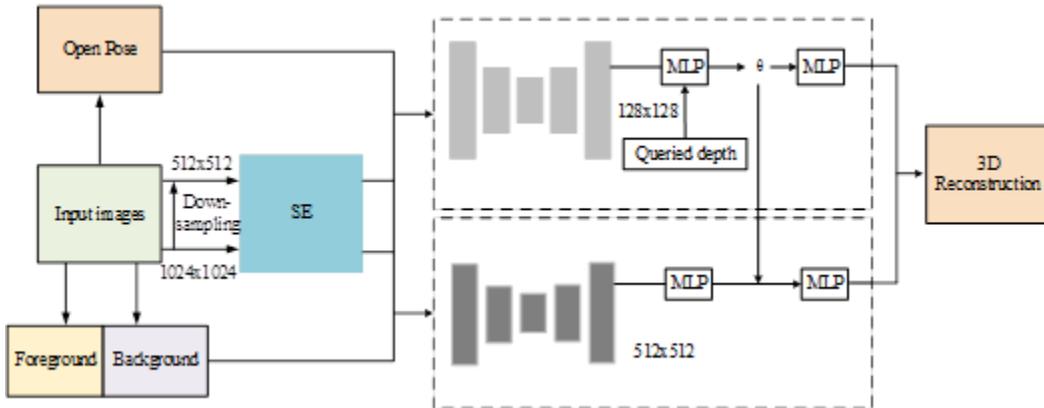
Currently, there are not many articles studying end-to-end anomaly metric-based learning methods, which have also been proposed in the past two years. End-to-end detection solves the issue of high data-labeling costs and obtains anomaly scores in a single step by using an unsupervised algorithm to identify anomalous frames from a large number of unlabeled films. Still, there is potential for greater performance improvement when compared to earlier supervised techniques; therefore, it is beneficial for researchers to carry out more thorough investigation. For example, a supervised method has been proposed to achieve 93.8% performance on the USCD dataset, while the unsupervised end-to-end method proposed in the literature has a performance of 71.7%, which is already far better than existing end-to-end methods, but there is still room for improvement (Chen et al., 2018).

METHODOLOGY

The proposed approach aims to improve the effectiveness of physical education by leveraging artificial-intelligence technology. Specifically, 3D reconstruction techniques in computer vision are combined to construct a set of human body shape reconstruction models and apply them to physical-training exercises and teaching-effectiveness assessment tasks. In this section, a detailed description of the multi-scale 3D reconstruction network based on the human-body shape proposed in this paper is provided. The network consists of multi-scale feature extraction, feature coding, and 3D reconstruction. The following subsections provide a step-by-step explanation of the proposed approach.

Figure 2 gives the structure of a multi-scale 3D reconstruction network based on the human-body shape proposed in this paper. It mainly includes multi-scale feature extraction, feature coding, and 3D reconstruction. Given a single static map V of the human body, extract the keypoint features of the human-body shape using OpenPose; obtain data M at different scales by downsampling; enhance the

Figure 2. The framework of the multi-scale human 3D reconstruction



extraction quality of the human contour features and 3D embedding features by combining channel attention; and predict the texture features of the back of the human body using the Pix2Pix algorithm to obtain the normal vector prediction maps F and B for the front and back of the human body. Then the feature information obtained from pre-processing and the multi-scale feature M are modeled using the coder and decoder and the multilayer perceptron (MLP). Finally, the generated human 3D models at different scales are fused to obtain the final human 3D model.

Multi-scale feature extraction refers to the process of extracting features from the human body at different scales, i.e., different levels of detail. Feature coding involves encoding the extracted features into a format that can be easily processed by a computer. 3D reconstruction is the process of creating a 3D model of the human body from the extracted and encoded features. OpenPose is a computer-vision library that can detect keypoints on the human body, such as the location of joints and limbs. Downsampling refers to the process of reducing the resolution of the image data to make it easier to process. Channel attention is a technique that enhances the importance of certain image features by selectively amplifying them. Pix2Pix is a type of neural network that can generate images based on input data. The normal vector prediction maps F and B are maps that provide information about the orientation of the human body in 3D space. The coder and decoder and MLP are types of neural networks that can process and manipulate the extracted feature data. Fusing the human 3D models at different scales involves combining the different levels of detail to create a more complete and accurate representation of the human body.

1. **Multi-scale feature extraction:** Here, to reduce the time overhead, depth-separable convolution is used instead of traditional convolutional layers to improve the real-time performance of the model by reducing the number of model parameters. First, the different scale representations M of the input data V are obtained by the downsampling method. Second, the human skeletal keypoint feature information J of the input data V is obtained by the opposing method, and then the normal vector prediction maps (F, B) of the front and back of the human body are obtained by the Pix2Pix method. Finally, the attention of M input channels of different scales is used to achieve the reinforcement of the focus on the human-contour features and 3D embedding features in the pictures during the network-training process and then combined with the generated human keypoint feature information J and the normal vector prediction maps F and B as the input of the next module.
2. **Feature coding:** In the feature-extraction module, the input information at different scales is passed through the graph encoder and the MLP for feature extraction. For low-resolution inputs,

the extraction of human 3D embedding features as well as human-contour information is combined mainly with the sampling point depth information, and the representation of human models is carried out by pixel-aligned implicit functions. For the high-resolution input, the extraction of human surface texture detail information is mainly performed, and the human 3D embedding features generated from the low-resolution images are combined to generate a human 3D model.

3. **Three-dimensional reconstruction:** After feature extraction of the input data at different scales, two kinds of human 3D models expressed by pixel-aligned implicit functions are obtained, and the two kinds of human 3D models with different reconstruction centers of gravity are combined to obtain a high-precision human reconstruction model.

Channel Attention

In a study of human reconstruction based on a single image, the model needs to focus only on the area where the human body is located in the image, while the rest of the area is useless. Thus, the model uses a channel-attention mechanism, the squeeze and excitation (SE) block, to model the dependency of each channel of the image to improve the feature-extraction capability of the network (Li et al., 2020). With channel attention, features can be adjusted on a channel-by-channel basis, allowing network learning to selectively improve feature areas with valuable information by global data while suppressing feature regions that are useless. Reduction of the computing effort of the reconstruction process and improved human body region representation can be achieved by the use of features obtained during the reconstruction process.

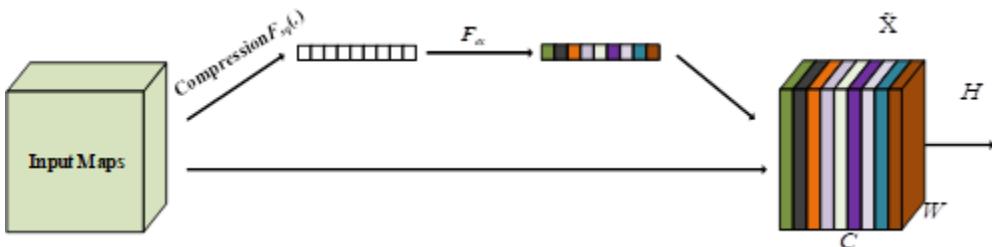
The basic structure of channel attention is shown in Figure 3. First, the squeeze operation is performed on the input image and the global average pooling is performed based on the width as well as the height of the feature map to reduce the spatial features to 1×1 , as shown in (1). Second, two fully connected layers and a nonlinear activation function are used to establish the connection relationship between each channel of the input data, and the global spatial features of the respective channel are used as the representation of that channel to form the channel descriptor, as shown in (1):

$$z_c = F_{sq}(x_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

where F_{sq} denotes the squeeze operation and H and W denote the height and width of the image, respectively.

$$\hat{z} = T_2(\text{ReLU}(T_1(z))) \quad (2)$$

Figure 3. Channel attention



where T_1 and T_2 denote fully connected functions.

Then the excitation operation is performed to realize learning the dependence of the model on each channel and adjusting the feature map according to the different dependences of the model on the channels. The normalized weights are first obtained by the sigmoid activation function, and then the channel-by-channel weighting is applied to each channel of the original feature map by multiplication to complete the recalibration of the channel attention to the original features. The calculation is shown in (3), and the adjusted feature map is the final output of the channel attention.

$$\hat{X} = X \cdot \sigma(z) \quad (3)$$

After global averaging pooling, the global perceptual field can be obtained, and the parameters and computation are greatly reduced by reducing the dimensionality of the feature map during the first full connection. After a nonlinear activation function, the number of channels is restored to the original number of channels by one full connection to complete the establishment of interchannel correlation.

Multi-Scale Feature Fusion

The commonly used human-shape representations are based on display functions, and Sun et al. (2020) propose using implicit functions for 3D reconstruction of objects to infer the overall shape of the target object from the global and contextual information of the image, but with the pitfall of spatial unalignment. In the literature, an implicit function is added to the human 3D reconstruction work for the first time, and a pixel-aligned implicit function is proposed to express 3D surfaces, which aligns the feature vectors of individual pixels with global features through convolutional networks with high memory efficiency (Hoza et al., 2021). Through the training of a graph encoder and the spatial alignment of features based on global contextual information, the researchers are able to determine the unique feature vectors for every pixel in the target image. An implicit function is utilized to determine whether the point corresponding to the depth is within or outside the human-body surface for the extracted pixel feature vector and the supplied depth acquired in the direction of that vector.

The implicit function defines the surface of the human-body surface as the level set f of a function.

$$f, f(X) = 0 \quad (4)$$

where X is the set of 3D points on the human-body surface; the space embedded in the surface does not need to be displayed for storage, achieving efficient storage. The pixel-aligned implicit function consists of a fully convolutional graph encoder and a continuous implicit function f represented by a multilayer perceptron.

$$f(F(x), z(X)) = s : s \in R \quad (5)$$

where X is the set of 3D points on the body surface, $z(X)$ is the specified depth value corresponding to the feature vector, and $F(x)$ is the feature vector extracted with the full convolutional map encoder.

$$F(x) = g(I(x)) \quad (6)$$

where g denotes the full convolutional graph encoder and I denotes the input image.

In the surface reconstruction work, since the human body is a nonrigid object with a complex surface structure, the curved surface cannot be well-represented by only one function, but the

position information of the surface 3D points is obtained by the full convolutional encoder and multilayer perceptron, so the ground truth is represented by a 0.5 level set with a continuous 3D occupancy domain.

$$f_v^*(X) = \begin{cases} 1, & \text{if } X \text{ is inside} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where, when $f_v^*(X) > 0.5$, the point is in the inner space of the human-body grid, and when $f_v^*(X) < 0.5$, the point is in the outer space of the human-body grid; the closer the value of $f_v^*(X)$ is to 1, the greater the chance that the point is inside, and the closer the value of f_v^* is to 0, the greater the chance that the point is outside.

Our goal is to obtain the implicit function f_v such that the function can fit the ground-truth function $f_v^*(X)$ as closely as possible. The implicit function f_v for pixel alignment is trained by minimizing the mean of the mean square error.

$$L_V = \frac{1}{n} \sum_{i=1}^n |f_v(F_V(x_i), z(X_i)) - f_v^*(X_i)|^2 \quad (8)$$

To obtain higher accuracy 3D models of the human body, this paper proposes a multi-scale training method based on pixel-aligned implicit functions. Based on the ideas in the previous subsection, a higher resolution 1024×1024 image is used as the input to the model and the human keypoint location information and positive and negative normal vector prediction maps of the image are obtained by preprocessing the image, which is the input into the model together with the original image.

After the image is input to the model, a 512×512 image is obtained after one downsampling as one layer of input to obtain the 3D embedded feature information of the human body at low resolution, the original image is used as another layer of input to obtain higher resolution detail information, and the location information of the keypoints of the human body and the normal vector prediction map of the front and back of the human body are used as feature inputs for pixel prediction at each scale to enhance the prediction capability of the network model for global 3D features of the human body. Due to the increase in input information, the multi-scale pixel alignment implicit function is shown in (9).

$$f(X, I) = g(\varphi(x, I), Z, J) \quad (9)$$

where I denotes the input image, X denotes the set of 3D points on the human surface, Z denotes the set of depths of x in the ray direction, J denotes the position information of keypoints on the human body, φ denotes the full convolutional map encoder, and g denotes the MLP.

The aim is to obtain an implicit function f_v so that it can fit the ground-truth function $f_v^*(X)$ as closely as possible. The first layer of the low-resolution representation is shown in (10).

$$f_v^L(X) = g^L(\varphi^L(x_L, I_L, F_L, B_L), Z, J) \quad (10)$$

where F_L and B_L represent the normal vector prediction maps of the front and back sides of the human body at low resolution, respectively. The second layer of high-resolution representation is shown in (11).

$$f_v^H(X) = g^H(\varphi^H(x_H, I_H, F_H, B_H), 0(X)) \quad (11)$$

where F_H and B_H denote the normal vector prediction maps of the front and back sides of the human body at high resolution, respectively. $0(X)$ denotes the global embedding features obtained by learning at low resolution. After obtaining the human-feature information at two resolutions, the zero-value surface of the implicit spatial surface is extracted from the obtained feature point coordinate information to obtain the reconstructed 3D model of the human body.

Loss Function

To improve the accuracy of human-morphology modeling, this paper uses the cross-entropy loss function to achieve end-to-end optimization of the model, which is calculated as shown in (12).

$$loss = \frac{1}{N} \sum_{i=1}^N (-B_i \log \bar{B}_i - (1 - B_i) \log(1 - \bar{B}_i)) \quad (12)$$

where N is the total number of samples, B_i is the i th-sample true value, and \bar{B}_i is the i th-sample predicted value.

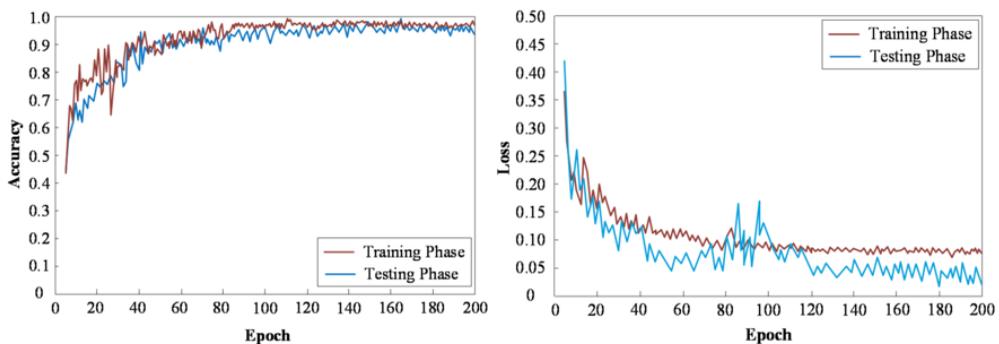
EXPERIMENTAL RESULTS AND ANALYSIS

Experimental Environment

The CPU used in this paper is the 48-core Intel Xeon Gold 5118, the server running memory is 128GB, the graphics card model is NVIDIA Tesla V100 GPU with 32GB video memory with NVIDIA CUDA 11.3, cuDNN v8.2.1 deep learning acceleration library with GPU-accelerated computing support of the PyTorch deep-learning framework that supports GPU-accelerated computing, etc. The training data for the model in this article is the human posture estimation dataset from the public dataset MPII Human Pose and COCO Keypoints. The 3DPW and Human3.6M datasets are public datasets used to evaluate three-dimensional human posture estimation methods. The 3DPW datasets contain character activities from multiple perspectives, while the MHuman datasets contain a large number of character movements and changing scenes.

The loss and accuracy curves of the training and testing phases of the model in this paper are given in Figure 4. It can be seen that after the number of iterations reaches 120, the accuracy and loss curve regions of the training and testing phases are smooth and the model reaches stable convergence.

Figure 4. Curves of the training and testing phases



Backbone Network Testing

In this paper, several different sets of backbone networks are used as feature extractors for 3D morphological reconstruction. The comparison results are shown in Figure 5. This is because SE is better able to direct the reconstruction of the human-stance model by focusing more intently on the salient features of the picture.

It can be seen that the ResNet50 reconstruction is better compared to the feature reconstruction extracted using the backbone network VGG. Therefore, ResNet50 was chosen as the backbone network of the model. Furthermore, by comparing the reconstruction effect of a single VGG and ResNet with that of the hybrid model with the addition of SE attention, it can be seen that the hybrid model of ResNet+SE can achieve 85.09 mean per joint position error (MPJPE) and 63.21 Procrustes-alignment (PA)-MPJPE on the 3DPW dataset. MPJPE of 65.14 and PA-MPJPE of 42.16 can be achieved on the human dataset. The advantage over all comparison models is obvious.

Comparison of Different Models

To verify the reconstruction performance of the model in this paper with the current mainstream models, several experiments were conducted with the same datasets, 3DPW and Human3.6M, and the same evaluation metrics; the results are shown in Figure 6., where the comparison models are denoted as A, B, C, and D, respectively.

Figure 5. The performance of different variant models in 3DPW and MHuman datasets

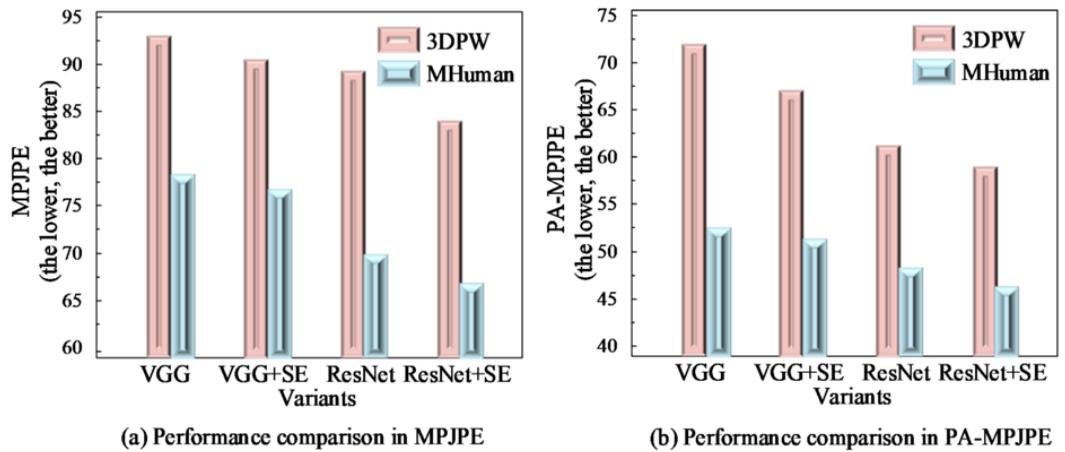
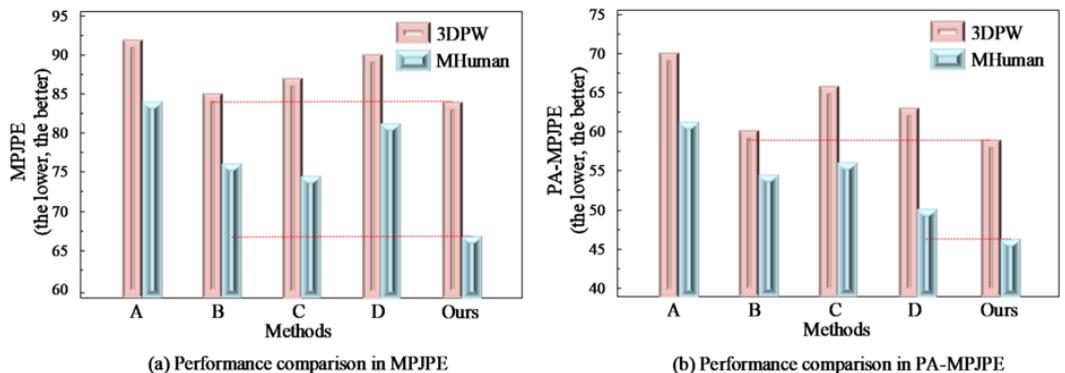


Figure 6. The performance of different models in 3DPW and MHuman datasets



By analyzing Figure 6, it can be seen that this paper’s model works best in the MHuman dataset with simple structure and clear scenes and is 3.02% and 1.53% lower than the mainstream Model A under MPJPE and PA-MPJPE evaluation indexes, respectively. In addition, the existence of severe occlusions in the 3DPW dataset leads to the poor reconstruction of existing models. The model in this paper is improved by reconstruction in the more challenging 3DPW dataset compared to current mainstream models, indicating that the model in this paper has good robustness to complex actions and occlusions.

Figure 7 gives a comparison of the time overhead of different models. From Figure 7, it can be seen that the model in this paper can achieve a generation rate of 12.4ms/image, Model A can achieve a generation rate of 15.8ms/image, Model B can achieve a generation rate of 16.3ms/image, Model C can achieve a generation rate of 14.6ms/image, and Model D can achieve a generation rate of 17.1ms/image. The aforementioned statistics also demonstrate that the model in this work can reach a higher generation rate. This is mostly because the discriminator stage of the model employs depth-separable convolution rather than conventional convolution in order to reduce model time overhead by lowering the model parameters.

Real-Case Testing

The confusion matrix of this paper’s method in six sets of real data is given in Figure 8. A confusion matrix is a matrix used to evaluate the performance of a classifier, where rows represent real tags and lists represent tags predicted by the classifier. Each element in the matrix represents a combination of real tags and predicted tags, which can be used to calculate indicators such as the accuracy rate, recall rate, F1-score, and so on of the classifier. In this article, the confusion matrix is used to evaluate the success and accuracy of the model in generating a human-form model from six sets of real data. From the confusion matrix, it can be seen that the successful accuracy rates of six testers in generating human body shape models in six groups of experiments are 93.32%, 92.22%, 94.88%, 94.05%, 93.38%, and 94.98%. The aforementioned information demonstrates that the model in this work has good real-time performance and tends to perform steadily throughout several sets of experimental data, confirming the model’s strong durability.

Figure 7. Comparison of time overhead between different models

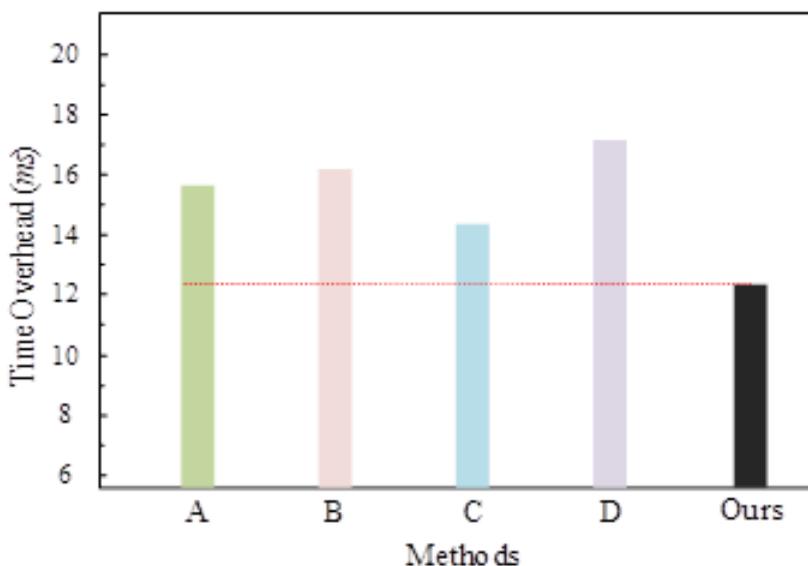
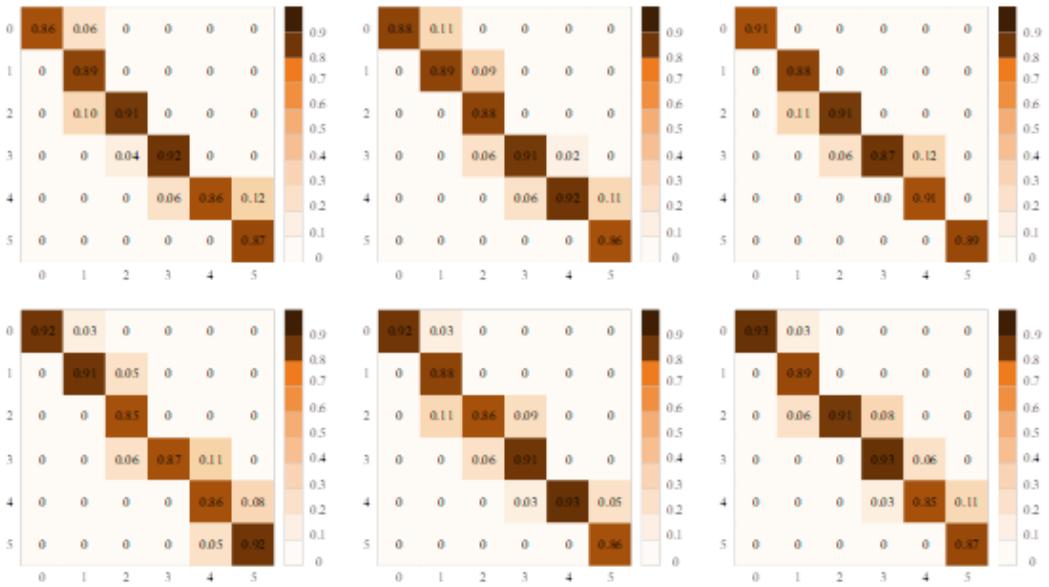


Figure 8. Confusion matrix



CONCLUSION

In this paper, a new multi-scale-based human body shape reconstruction model is proposed and applied to the online assessment task of sports lessons. The problem of ambiguity in the reconstruction of the human back under single-view conditions is alleviated by drawing on knowledge of the target's foreground and background regions. In addition, multi-scale features are constructed and fused from two dimensions, high resolution and low resolution, to obtain key information and edge information of human joint points at fine granularity, further improving the performance of reconstruction. The efficiency of the model design in this paper is verified by conducting experiments on real and open-source datasets.

AUTHOR NOTE

The author declares that she has no conflicts of interest. This work was not supported by any outside funding. The author thanks those who have contributed to this research.

REFERENCES

- Abdulhammed, R., Faezipour, M., Abuzneid, A., & AbuMallouh, A. (2019a). Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic. *IEEE Sensors Letters*, 3(1), 1–4. doi:10.1109/LSENS.2018.2879990
- Abdulhammed, R., Musafir, H., Alessa, A., Faezipour, M., & Abuzneid, A. (2019b). Features dimensionality reduction approaches for machine learning based network intrusion detection. *Electronics (Basel)*, 8(3), 322. doi:10.3390/electronics8030322
- Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*, 189, 105124. doi:10.1016/j.knosys.2019.105124
- Bedard, C., Bremer, E., & Cairney, J. (2019). Evaluation of the Move 2 Learn program, a community-based movement and pre-literacy intervention for young children. *Physical Education and Sport Pedagogy*, 25(1), 101–117. doi:10.1080/17408989.2019.1690645
- Chen, S., Zhu, X., Androzzi, J., & Nam, Y. H. (2018). Evaluation of a concept-based physical education unit for energy balance education. *Journal of Sport and Health Science*, 7(3), 353–362. doi:10.1016/j.jshs.2016.06.011 PMID:30356610
- Chen, Y., Ma, T., Yang, X., Wang, J., Song, B., & Zeng, X. (2021). MUFFIN: Multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics (Oxford, England)*, 37(17), 2651–2658. doi:10.1093/bioinformatics/ctab169 PMID:33720331
- Chiew, K. L., Tan, C. L., Wong, K. S., Yong, K. S. C., & Tiong, W. K. (2019). A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Information Sciences*, 484, 153–166. doi:10.1016/j.ins.2019.01.064
- Ding, Y., Li, Y., & Cheng, L. (2020). Application of Internet of Things and virtual reality technology in college physical education. *IEEE Access : Practical Innovations, Open Solutions*, 8, 96065–96074. doi:10.1109/ACCESS.2020.2992283
- Feng, Y. (2020). An evaluation method of PE classroom teaching quality in colleges and universities based on grey system theory. *Journal of Intelligent & Fuzzy Systems*, 38(6), 6911–6915. doi:10.3233/JIFS-179769
- Granjal, J., Silva, J. M., & Lourenço, N. (2018). Intrusion detection and prevention in CoAP wireless sensor networks using anomaly detection. *Sensors (Basel)*, 18(8), 2445. doi:10.3390/s18082445 PMID:30060498
- Gu, Y. (2022). Deep integration of physical education and multimedia technology using Internet of Things technology. *Wireless Communications and Mobile Computing*, 2556142, 1–13. doi:10.1155/2022/2556142
- Heidari, A., Jafari Navimipour, N., Unal, M., & Zhang, G. (2023). Machine learning applications in internet-of-drones: Systematic review, recent deployments, and open issues. *ACM Computing Surveys*, 55(12), 1–45. doi:10.1145/3571728
- Hoza, B., Shoulberg, E. K., Tompkins, C. L., Meyer, L. E., Martin, C. P., Krasner, A., Dennis, M., & Cook, H. (2021). Meeting a physical activity guideline in preschool and school readiness: A program evaluation. *Child Psychiatry and Human Development*, 52(4), 719–727. doi:10.1007/s10578-020-01055-9 PMID:32914291
- Jain, A. K., & Gupta, B. B. (2018). Towards detection of phishing websites on client-side using machine learning based approach. *Telecommunication Systems*, 68(3), 687–700. doi:10.1007/s11235-017-0414-0
- Kim, C. M., & Kwak, E. C. (2022). An exploration of a reflective evaluation tool for the teaching competency of pre-service physical education teachers in Korea. *Sustainability (Basel)*, 14(13), 8195. doi:10.3390/su14138195
- Li, H., Qiu, K., Chen, L., Mei, X., Hong, L., & Tao, C. (2020). SCAAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 18(5), 905–909. doi:10.1109/LGRS.2020.2988294
- Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, 164, 163–173. doi:10.1016/j.knosys.2018.10.034

- Lonsdale, C., Lester, A., Owen, K. B., White, R. L., Peralta, L., Kirwan, M., Diallo, T. M. O., Maeder, A. J., Bennie, A., MacMillan, F., Kolt, G. S., Ntoumanis, N., Gore, J. M., Cerin, E., Cliff, D. P., & Lubans, D. R. (2019). An internet-supported school physical activity intervention in low socioeconomic status communities: Results from the Activity and Motivation in Physical Education (AMPED) cluster randomised controlled trial. *British Journal of Sports Medicine*, 53(6), 341–347. doi:10.1136/bjsports-2017-097904 PMID:28993404
- Reece, L. J., Foley, B., Bellew, W., Owen, K., Cushway, D., Srinivasan, N., Hamdorf, P., & Bauman, A. (2021). Active kids: Evaluation protocol for a universal voucher program to increase children's participation in organised physical activity and sport. *Public Health Research & Practice*, 31(2), 10–21. doi:10.17061/phrp30122006 PMID:34104931
- Safi, A., Cole, M., Kelly, A. L., & Walker, N. C. (2021). An evaluation of physical activity levels amongst university employees. *Advances in Physical Education*, 11(2), 158–171. doi:10.4236/ape.2021.112012
- Shi, X., Li, X., & Wu, Y. (2021). The application of computer-aided teaching and mobile internet terminal in college physical education. *Computer Aided Design*, 18(S4), 163–174. doi:10.14733/cadaps.2021.S3.163-174
- Sun, M., Konstantelos, I., & Strbac, G. (2019). A deep learning-based feature extraction framework for system security assessment. *IEEE Transactions on Smart Grid*, 10(5), 5007–5020. doi:10.1109/TSG.2018.2873001
- Tan, C., Gu, G., Ruan, T., Wei, S., & Zhao, Y. (2020). Dual-gradients localization framework with skip-layer connections for weakly supervised object localization. In C. W. Chen, R. Cucchiara, X. S. Hua, G. J. Qi, E. Ricci, Z. Zhang, & R. Zimmermann (Eds.), *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 1976–1984). ACM. doi:10.1145/3394171.3413622
- Uysal, G., & Balci, S. (2018). Evaluation of a school-based program for internet addiction of adolescents in Turkey. *Journal of Addictions Nursing*, 29(1), 43–49. doi:10.1097/JAN.0000000000000211 PMID:29505460
- Wang, Z. (2018). Fuzzy comprehensive evaluation of physical education based on high dimensional data mining. *Journal of Intelligent & Fuzzy Systems*, 35(3), 3065–3076. doi:10.3233/JIFS-169661

Rui Guo holds a Master of Physical Education and Training and is an associate professor at the Henan University of Technology. She was born in Zhengzhou, Henan Province, China, and graduated from Henan University in 2003. Her research interests include the teaching and training of basketball.