

Evaluation System for Dancing Enlightenment Posture Training Using the Skeleton Tracking of Microsoft Common Objects in Context

Ruilong Huang*, Huifang Deng*, Ruei-Yuan Wang**, Bing-Yuh Lu*,
Hongwei Ren*, Yiheng Chen*, Jianwen Ye*, Jinhui Chen*, Yingbo Jia*, Leyang Lang*

*School of Automation, Guangdong University of Petrochemical Technology, Maoming City, Guangdong, China

**School of Science, Guangdong University of Petrochemical Technology, Maoming City, Guangdong, China

FranklinLu888@outlook.com

Abstract— Dance enlightenment education is of great significance to children's physical health. Therefore, we developed an evaluation system to correct the posture of the beginners for their dancing training. The pre-train network is based on the Open Neural Network Exchange. The pattern of human pose is the skeleton in the Microsoft Common Objects in Context dataset. We designed a quantitative presentation to calculate the similarity of the postures between the skeletons of the dancing beginner and target image of the training, and obtained objective evaluation indicators based on the recorded of the angle differences of limbs to calculate the score. 8 angles of the joints have been computed and presented in the evaluation system. The results show that the dancing beginner can correct her postures to approach the target image of the training. She improved the score from 94 to 96. Now, parents pay more and more attention to the quality education of their children. The AI-aided dancing training will make the beginners to learn the performances of the postures much easier in any time and any location. Therefore, the learning of dancing will become more interesting for the beginners.

Keywords—Dance, Deep Learning, Evaluation System, MatLab, Microsoft Common Objects in Context, Posture, Skeleton, Training.

I. INTRODUCTION

Dance enlightenment education is of great significance to children's physical and mental health. It can not only regulate children's behavior but also improve children's aesthetics. Dance enlightenment education requires scientific and

effective learning methods. In this diversified society, movement practice, form, dance posture and other training in dance enlightenment can avoid the bad forms of children's bodies during growth and development, forge children's good appearance, and lay a good foundation for children's healthy development. The application of human posture recognition technology in computer vision and deep learning in dance enlightenment teaching can help teachers better understand the learner's movement, correct wrong movement postures in time, and improve teaching.

Wu et al. [1] has proposed multi-stage pose network (MSPN) and optimized in three aspects of the stacked hourglass network, i.e. network structure, feature information transfer and heat map generation. Li et al. [2] deduced the downsampling and upsampling on the Resnet structure, and featured out that the different stages are connected to enhance the ability of feature expression. Scholars generated different heatmaps at their respective stages to gradually improve the estimation accuracy of human pose. As the accuracy of human posture estimation networks continues to improve, the complexity of the network structure also gradually increases, which makes it difficult to compare and analyze between networks. Xiao et al. [3] proposed the simple baseline network (SBN) in 2018. It mainly proposed a relatively simplified baseline for human posture estimation and posture tracking. This paper mentioned believes that the current popular human posture estimation methods are too complex. The various models proposed on the Internet seem to be quite different in structure, but they are very similar in performance. Compared with the complex network structures of traditional networks such as Hourglass [4], the basic SBN structure appears intuitive and simple. This network structure is constructed by adding several layers of denaturation after Resnet to directly

generate heatmaps. Compared to other models, this is done by replacing the oversampling structure through deconvolution.

Recently, there were many applications of the posture correction using the AI technology, especially, extracting the human skeleton to evaluate the posture. Carey et al. [9] compared the effectiveness of two different skeletal pose models for a near real-time, multi-stage classifier and find no significant difference between the 2 pose models. Therefore, we select Microsoft Common Objects in Context (MS COCO) dataset for this study. Liu et al. [10] proposed a mechanism for estimating and correcting fitness posture based on deep learning. The 14 keypoints of the human body can be obtained after correction. Jangade et al. [11] pointed out that Human Pose Estimation (HPE) will be a wide range of applications and enter human daily living step by step. In fact, many papers mentioned the potential applications of HPE [12-15].

However, little researches mentioned about the pose correction of training dancing. Intuitively, dancing is a kind of art. Therefore, little scientists pay attention to the dancing. The number of the studies related to AI-aided correction of dancing is less. Consequently, this study tried to develop a system to implement AI-aided correction of dancing. The operations of the evaluation system in this study are as follows: (1) Construct a deep learning human posture detection algorithm network to realize human posture recognition, extract and compare the coordinates of human body key points among learners and standards; (2) Based on the coordinate information of the key points of the human body of the standard person, the system determine the correlation and angle difference with the key points of the learner's human body, and finally obtain the similarity of dance movements between the learner and the standard person; and (3) Build a dance enlightenment evaluation system GUI platform on MATLAB software to achieve the purpose of assisting children's teaching and enlightenment education.

II. METHOD

A. Estimate Body Pose Using Deep Learning Using MatLab

This study follows an example of MatLabTM (The MathWorks Inc., USA) [5]. Therefore, the pre-train network is based on the Open Neural Network Exchange (ONNX) which is an open ecosystem for interoperable AI models [6] at first. The pattern of human pose is the MS COCO dataset which is a large-scale object detection, segmentation, key-point detection, and captioning dataset [7]. Then, the tested image was predicted the heatmaps and part affinity fields (PAFs), which are output from the 2-D output convolutional layers. The post-processing part of the algorithm identifies the individual poses of the people in the image using the heatmaps and PAFs returned by the neural network.

B. MS COCO human skeleton

The MS COCO data set was proposed in 2014 and is the mainstream in the field of human posture recognition. It focuses on solving large datasets such as object detection, key point detection, object segmentation and subtitle generation in natural environments by using computer vision technology. In

2016, we added a new task to the MS COCO dataset using 2D human skeleton key point detection. The data set contains human samples labeled with coordinates of key points on the human body. These samples are labeled with 17 key points on the human skeleton. Below we combine pictures and specific tables to learn more. Figure 1 shows the 17 key points of the MS COCO human skeleton.

C. Indicators of human posture estimation

For the COCO data set, the officially designated joint point similarity measurement method is OKS object keypoint similarity (OKS), which is used to calculate the similarity between the true value and the predicted key point. This data set is inspired by the target detection metric. Different from other evaluation indicators, it uses the average accuracy mean based on the OKS as a more scientific evaluation indicator. The value of OKS is between 0 and 1. The closer it is to 1, the closer the predicted human joint points are to the real values of the data and annotations, and the better the prediction effect.

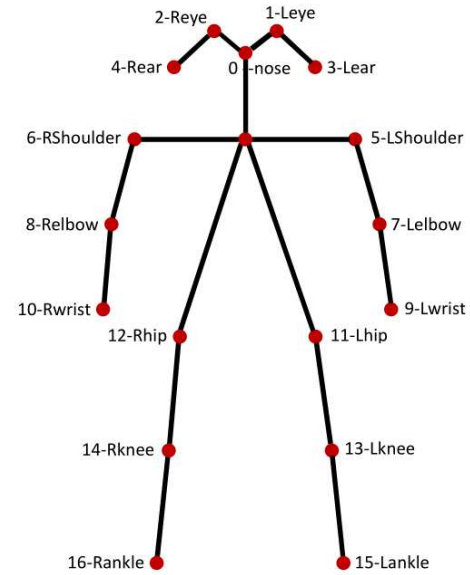


Figure 1. The 17 key points of the MS COCO human skeleton [8].

The calculation of OKS is based on a non-standardized Gaussian distribution.

Mathematically, the keypoint similarity (KS) for keypoint i is written as:

$$KS_i = \exp\left(-\frac{1}{2} \left(\frac{d_i}{s_i}\right)^2\right) \quad (1)$$

where the Euclidean distance between the ground truth and predicted keypoint i is denoted by d_i , the constant for keypoint i is denoted by k ; the scale of the ground truth object is denoted by s ; thus s^2 becomes the object's segmented area. For each keypoint, the KS lies between 0 and 1.

The OKS is expressed by

$$OKS_i = \frac{\sum_i KSi \cdot \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \quad (2)$$

where the keypoint similarity for keypoint i is denoted by OKS_i ; v_i is the ground truth visibility flag for keypoint i . $\delta(v_i > 0)$ is the Dirac-delta function which computes as 1 if the keypoint i is labeled, otherwise 0.

D. Dance Enlightenment Evaluation System Based on Posture Estimation

A new method for the dance enlightenment evaluation system is introducing in this section. This method follows the steps:

- Use the pre-trained COCO model to extract human skeleton information from standard dance pose parameters and student dance pose parameters.
- Obtain position information of human body key points and calculate based on the obtained coordinate information of human body key points.
- Compute the angle differences between each limb, and infer the angle difference between the limb angle of the image of student's dance pose and that of standard dance pose.
- Design quantitative presentation to calculate pose similarity, and obtain objective evaluation indicators based on the recorded of the angle differences of limbs.
- Select the largest error to advice the corrections of the student's poses.

In summary, the computing processes of the dance enlightenment evaluation system based on posture estimation are presented in Fig. 2.

E. Equipment

1. Hardware

The experiment is performed by a laptop computer with Intel® Core™ i5-8250U CPU @ 1.60GHz, 4 cores, 8 logical processors and 8GB memory. The operating system is Windows 10.

2. Software

The program is run in MatLab 2020a integral developing environment. The MatLab is a high-level computer programming language and interactive environment for developing algorithms, data analysis, and visualization as well as provides a large number of toolboxes and functions to help users to efficiently handle various complex mathematical operations and data analysis.

III. RESULTS

The initial page of the evaluation system is presented in Fig. 3. There are 3 frames in the figure. The left frame shows the picture of the target pose of this train which can be selected by the items of the list under the picture. The middle frame is null in the figure. The frame will be filled with a picture of the learner when the train begins. The rightmost frame is designed for showing the differences of the joint

angles between the target and learner's MS COCO skeletons which is denoted by $\Delta\theta_i$.

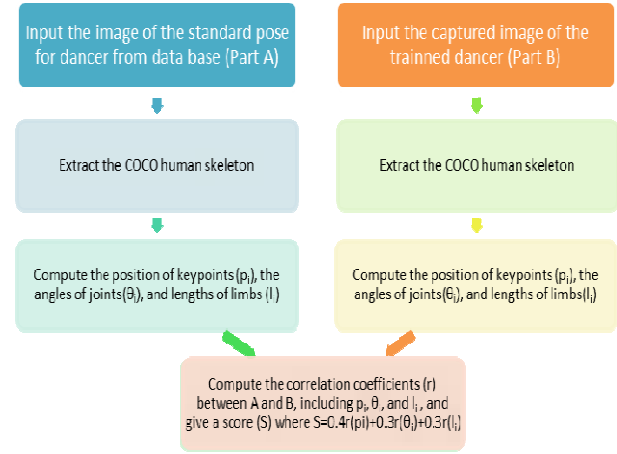


Figure 2. The computing processes of the dance enlightenment evaluation system based on posture estimation

In Fig. 4, the null space in Fig. 3 has been filled. The image of the learner and the corresponding skeleton is uploaded and displayed. Furthermore, the $\Delta\theta_i$ are measured and shown in the blocks in the rightmost frame. They are 17, 12, 2, 6, 9, 0, 5, and 6 degrees, respectively. The approaching time is 0.15 sec. and the score of the training is 94 points.

In Fig. 5, the learner corrects her postures to improve the dancing based on the $\Delta\theta_i$ of the Fig. 4. Therefore, the $\Delta\theta_i$ improve to 5, 3, 3, 0, 9, 0, 5, and 6 degrees. Concurrently, the score goes ahead to 96.

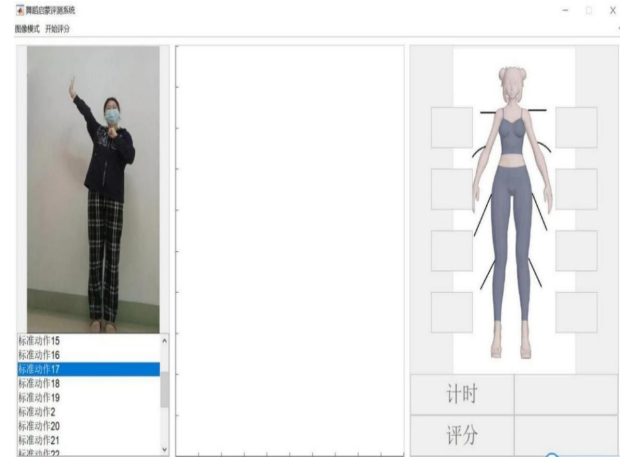


Figure 3. The initial page of the evaluation system

IV. DISCUSSION

In Fig. 4, the difference angle of the right elbow (8-Relbow) between target and learner's images is 17 degree. Therefore, the learner raises her arms slightly and stands slightly straighter. Then, the corrected learner's dance movements and postures are photographed and uploaded to the dance

enlightenment evaluation system. Figure 5 shows the corrected human posture model. From the picture, we can know that after raising the right arm, the left arm also needs to be further adjusted, such as relaxing the arm slightly and raising the left arm slightly, so as to meet the pose of the target picture as much as possible, so that the similarity will also be increased. Therefore, the posture and score have been improved.

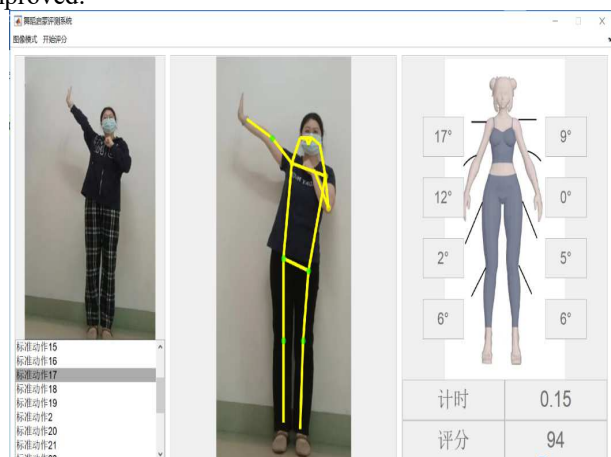


Figure 4. The training phase of the evaluation system

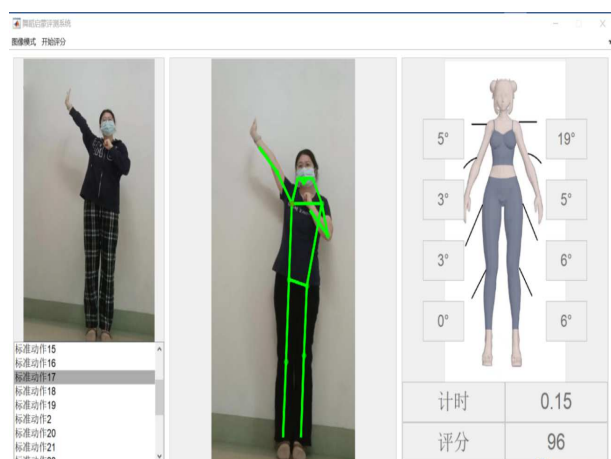


Figure 5. The correcting phase of the evaluation system

The computation of the dance posture similarity is an open question. However, we proposed a method to compute the score. Furthermore, the evaluation of the pose focuses on the limbs, because we think the poses of limbs are the main postures to influence the touch of the dancers' expressions. However, the evaluation of total skeletons might be tested in the near future.

If the color of the learner's clothes is similar to their skin color, the background is cluttered or the learner wear the loose-fitting clothes, the recognitions of the human bones and joints become less accurate. Therefore, it is necessary to improve the training data systematically and improve the human posture estimation algorithm.

In this test, the learner danced with the face mask. Therefore, some errors of the keypoints positions on the face

are obvious. However, the target image of the training is in the same condition, i.e. the dancer also posed with the face mask. Therefore, the "noise", the face mask, has been cancelled. The computations of $\Delta\theta_i$ keep accurate. In fact, the computation of the keypoints on the body and limbs are independent to those on the face. Therefore, this study verified the computing accuracy when the face mask is necessary to be with the learner

V. CONCLUSION

Now, parents pay more and more attention to the quality education of their children. Now, parents pay more and more attention to the quality education of their children. The AI-aided dancing training will make the beginners to learn the performances of the postures much easier in any time and any location. Therefore, the learning of dancing will become more interesting for the beginners.

ACKNOWLEDGMENTS

The authors thank for the support from the Guangdong University of Petrochemical Technology, Guangdong, China (Project number: 702/72100003142), the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515010168, Grant 2019A1515010830, and Maoming Green Chemical Industry Research Institute (Project number: MMGCIRI-2022YFJH-Y-042), Guangdong, China.

REFERENCES

- [1] Wu, H., Lu, X., Lei, B. and Wen, Z., 2021. Automated left ventricular segmentation from cardiac magnetic resonance images via adversarial learning with multi-stage pose estimation network and co-discriminator. *Medical Image Analysis*, 68, p.101891.
- [2] Li, B. and He, Y., 2018. An improved ResNet based on the adjustable shortcut connections. *Ieee Access*, 6, pp.18967-18974.
- [3] Xiao, B., Wu, H. and Wei, Y., 2018. Simple baselines for human pose estimation and tracking. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 466-481).
- [4] Newell, A., Yang, K. and Deng, J., 2016. Stacked hourglass networks for human pose estimation. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14* (pp. 483-499). Springer International Publishing.
- [5] Estimate Body Pose Using Deep Learning Using MatLab (<https://www.mathworks.com/help/deeplearning/ug/estimate-body-pose-using-deep-learning.html#EstimateBodyPoseUsingDeepLearningExample-3>)
- [6] Open Neural Network Exchange (<https://onnx.ai/>)
- [7] Microsoft Common Objects in Context (<https://paperswithcode.com/dataset/coco>)
- [8] The 17 key points of the MS COCO human skeleton. (<https://www.stubbornhuang.com/525/>)
- [9] Carey, K., Abruzzo, B., Lowrance, C., Sturzinger, E., Arnold, R. and Korpela, C., 2020, April. Comparison of skeleton models and classification accuracy for posture-based threat assessment using deep-learning. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II* (Vol. 11413, pp. 671-678). SPIE.
- [10] Liu, X., Xiao, H. and Cheng, J., 2021, July. Human posture estimation and correction based on the CPM and the Pearson correlation coefficient. In *International Conference on Sensors and Instruments (ICSI 2021)* (Vol. 11887, pp. 385-390). SPIE.
- [11] Jangade, J. and Babulal, K.S., 2023, March. Study on Deep Learning Models for Human Pose Estimation and its Real Time Application. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1-6). IEEE.

- [12] Andriluka, M., Pishchulin, L., Gehler, P. and Schiele, B., 2014. 2d human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition (pp. 3686-3693).
- [13] Toshev, A. and Szegedy, C., 2014. Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653-1660).
- [14] Zhang, F., Zhu, X. and Ye, M., 2019. Fast human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3517-3526).
- [15] Mune, T.L., Jembre, Y.Z., Weldegebriel, H.T., Chen, L., Huang, C. and Yang, C., 2020. The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation. IEEE Access, 8, pp.133330-133348.



Ruilong Huang received his B.Eng in industry automation from Northeast Petroleum University in 2002, M.Eng in control theory and control engineering from South China University of Technology in 2005. He is currently a lecturer with Faculty of Automation, GDUPET Maoming City, Guangdong, China. His academic interests focus on intelligent detection and control technology.



Huifang Deng received her B.S. from the Faculty of Automation, Guangdong University of Petrochemical Technology (GDUPET), Guangdong, China in June, 2023. She got the scholarships of excellent learning at GDUPET many times. She is currently a junior engineer at Chinese Huadian Corporation, Guangdong, China. Her research interests include AI and image processing technologies.



Ruei-Yuan Wang received his PhD in Science from the Institute of Geosciences at the Chinese Culture University in Taiwan in 2010; he has been served as a postdoctoral researcher at the Spatial Information Research Center (SIRC) of Taiwan University (NTU) and the Center for Space and Remote Sensing Research (CSRSR) of Central University (NCU). He is currently serving as an associate professor in the Department of Geographic Sciences at the Faculty of Science, Guangdong University of Petrochemical and Chemical Technology (GDUPET), Maoming City, Guangdong, China. He primarily engaged in teaching courses in geographic science, remote sensing (RS), and geographic information systems (GIS). His academic interests focus on geographic science related fields, environmental ecology, remote sensing carbon monitoring, tourism geography, etc. His application technology mainly focuses on decision support (DSS), knowledge management (KM), artificial neural networks (ANN), big data, remote sensing technology, and geographic information system technology (GIS). He previously served as expert peer reviewers of various journals for Sustainability; Land; Forests; Expert Systems with Application; Business Management and Economics, Journal of Coastal Research (JCR), etc.



Bing-Yuh Lu received his BS in electrical engineering from National Central University in 1988, MS in electrical engineering from National Taiwan University in 1993, and PhD in electrical engineering from National Taiwan University in 2000. He is currently a professor with Faculty of Automation, Guangdong University of Petrochemical Technology (GDUPET), Maoming City, Guangdong, China. He has been an instructor (1993 to 2000), an

associate professor (2000 to 2016), and a full professor (2016 to 2019) with the Department of Electronic Engineering, Tunghai University, New Taipei City, Taiwan, from 1993 to 2019. He is a member of IEEE, has been a member of the Technical Committee IEEE International Conference on Advanced Communication since 2015, and served as a reviewer for some international journals. His academic interests include electronic circuits and systems, medical engineering, acoustics, modeling, and signal measurement and processing.



Hongwei Ren received her B.S. degree in Industry Automation from Northeast Electric Power University, Jilin, China, in 1998, her M.S. degree and Ph.D. degree in Systems Engineering from South China University of Technology, Guangzhou, China in 2002 and 2017, respectively. She is currently an Associate Professor in the School of Automation, Guangdong University of Petrochemical Technology, Maoming, China. Her research interests include synchronization of complex networks, consensus of multi-agent systems and stochastic dynamic system analysis and control.



Yiheng Chen is a student with Faculty of Automation, Guangdong University of Petrochemical Technology, Guangdong, China. He won the third prize of 2022 "Cossberg Cup" Engineering College Students' Comprehensive Experimental Skills Competition of Guangdong University of Petrochemical Technology, and participated in the 2022 "Application Innovation" project of "Sail Plan" of Maoming Green Chemical Research Institute. His research interests include PCB hardware design and fuzzy control.

Jianwen Ye, is a student majoring in Electrical Engineering and Automation at Guangdong University of Petrochemical Technology, has won university-level scholarship and the second prize of Guangdong Provincial Engineering College Students' Comprehensive Experimental Skills Competition.

Jinhui Chen is a student majoring in Electrical Engineering and Automation at Guangdong University of Petrochemical Technology, has won university-level scholarship and the second prize of university-level innovation and Entrepreneurship Competition.

Yingbo Jia is a student majoring in measurement and control technology and instrumentation at Guangdong University of Petrochemical Technology, has won first-class scholarships for many times

Leyang Lang is a student with Faculty of Automation, Guangdong University of Petrochemical Technology, Guangdong, China. He has won the third prize of the Guangdong Provincial Research Cup. His main research interests are geographic remote sensing monitoring, machine learning, and deep learning.