Real-Time Posture Correction in Gym Exercises: A Computer Vision-Based Approach for Performance Analysis, Error Classification and Feedback

Hitesh Kotte[1], Miloš Kravčík[1] and Nghia Duong-Trung[1]

[1]*Educational Technology Lab, German Research Center for Artificial Intelligence (DFKI), Alt-Moabit 91C, 10559 Berlin, Germany.*

Abstract

Developing psychomotor skills is a major challenge, but modern technology offers novel solutions that can provide valuable support to trainees. This paper proposes a pioneering strategy that uses computer vision methods to monitor performance and provide real-time feedback on posture during fitness exercises, allowing for instant self-correction and motivation even without professional guidance. Our system utilizes a versatile learning framework to analyze live expert demonstrations or recorded video content. We leverage the deep learning YOLOv7-pose model to identify human keypoints and combine it with a human topology-oriented tracking procedure. Our system delivers immediate feedback to rectify posture by collecting comprehensive tracking data. Notably, we capitalize on transfer learning techniques to avoid extensive model retraining. To demonstrate the usefulness of our method, we benchmarked it to professional fitness videos and evaluated it with five inexperienced participants. The results showed a positive reaction from the participants, suggesting improvements to the user interface.

Keywords

Psychomotor Skills, Computer Vision, Performance Analysis, Error Classification, Feedback

1. Introduction

Augmented intelligence, an exciting blend of human and artificial intelligence, is redefining several fields, notably sports [1]. Advanced fitness training systems use machine learning, sensors, and cameras to provide tailored coaching and instant feedback, significantly improving performance in various activities [2]. Fitness applications are vital today, especially when access to a personal trainer is limited due to logistical or financial constraints. Hiring a trainer for every workout session can be expensive, and scheduling conflicts might make finding a suitable time for both parties difficult. These circumstances highlight the advantages of fitness applications, which can guide users through their workouts, reducing the risk of injuries due to incorrect form or overexertion [3]. The ubiquity of smartphones and wearable devices makes fitness applications accessible to virtually anyone, anywhere, providing personalized, expert guidance

MILeS'23: 18th European Conference on Technology Enhanced Learning (ECTEL2023), 3rd International Workshop on Multimodal Immersive Learning Systems, September 04–08, 2023, Aveiro, Portugal.
hitesh.kotte@dfki.de (H. Kotte); milos.kravcik@dfki.de (M. Kravčík); nghia$_{t}rung.duong@dfki.de$ (N. Duong-Trung)
https://sites.google.com/ismll.de/duongtrungnghia (N. Duong-Trung)
0009-0005-8885-889X (H. Kotte); 0000-0003-1224-1250 (M. Kravčík); 0000-0002-7402-4166 (N. Duong-Trung)

CEUR Workshop Proceedings (CEUR-WS.org)

at their fingertips. Despite these benefits, without appropriate guidance, fitness programs can result in injuries and lack of motivation [4]. Tools like safety equipment, sensors, and augmented reality devices can help prevent these issues, but they often limit the freedom of movement [5].

Our research aims to tackle these problems by incorporating computer vision into training systems. This technology can analyze an individual's movements in detail, identify potentially dangerous practices, provide real-time feedback, automate tasks like counting repetitions, and introduce a game-like element to workouts to increase engagement [6]. Previous research has shown AI's promise in sports, particularly in recognizing gym activities and weight training [7]. However, these models generally require extensive data collection and training. While proposals for augmented reality solutions exist, their high cost and impracticality often limit their application [8]. Multimodal Learning Analytics (MMLA) delves into AI's potential to decipher complex learning data [9]. Nevertheless, the reliable identification of findings and reproducibility of studies for generalizable results remains challenging, underscoring the need for more robust methodologies [10].

Aiming at the identification of mistakes in the execution of movements of the learners and generating helpful feedback for them, we employ the YOLOv7 model for pose estimation. This cutting-edge AI model is tasked with detecting the keypoints on a human body, as shown in Figure 1, providing an intricate understanding of the body's positioning and movement [11]. These keypoints are utilized further in our methodology to calculate the angle between them during various exercises. This distinctive classification approach allows us to differentiate between different movements, offering real-time feedback tailored to the individual's performance. The combination of AI technology and physical data promises unprecedented precision and personalization to fitness training. The first results were promising, showing the potential usefulness and attractiveness of this approach [12]. In this paper, we further elaborate on the feedback provision opportunities, which is still a work in progress. Detected errors must be processed in such a way that learners receive starting points for improving their processes, which they can cognitively process and implement psychomotorically.

2. Our Method

2.1. Real-time Pose Detection and Tracking

The creation of intelligent tutoring systems that merge the power of human intelligence with artificial intelligence presents a unique set of challenges. Such systems rely on machine and deep learning algorithms to parse data, identify patterns, and enhance human performance. YOLO (You Only Look Once) is a standout in this domain, a real-time object detection system renowned for its speed and accuracy. YOLO has found favor in various applications, including human pose estimation, given its efficiency and real-time processing capability [13, 14, 15]. However, the importance of visual feedback in this setup varies based on the complexity and intensity of the exercises being performed. Complex movements typically require continuous visual feedback to ensure correct form and posture, while high-intensity workouts demand less frequent feedback. Users have preferred visible skeleton overlays to understand their body positions better. Interestingly, they find it more useful when specific incorrect body parts are highlighted, rather than having the entire skeleton light up, helping them to make targeted

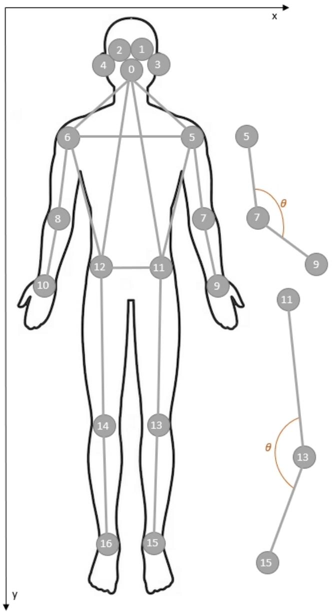


Figure 1: Our keypoints used in the experiments.

corrections during their workouts [16].

2.2. Calculating Joint Angles from Human Pose Key Points

Our approach utilizes a 17-keypoint pose topology to track and analyze body movements during fitness exercises. Given three key points $u(x_u, y_u), v(x_v, y_v), p(x_p, y_p)$, the joint angle $\theta(u, v, p)$ (in degrees) between two rays formed by the three mentioned points is calculated as follows:

$$\theta(u, v, p) = \frac{180(\phi(y_p - y_v, x_p - x_v) - \phi(y_u - y_v, x_u - x_v))}{\pi} \tag{1}$$

where the angle $\phi(y, x)$ (in radians) between the ray from the origin to the point (x, y) and the positive x-axis in the Cartesian plane is calculated as follows:

$$\phi(y, x) = \begin{cases} \arctan(\frac{y}{x}), & \text{if } x > 0, \\ \frac{\pi}{2} - \arctan(\frac{x}{y}), & \text{if } y > 0, \\ -\frac{\pi}{2} - \arctan(\frac{x}{y}), & \text{if } y < 0, \\ \arctan(\frac{y}{x}) \pm \pi, & \text{if } x < 0, \\ \text{undefined}, & \text{if } x = 0 \text{ and } y = 0. \end{cases} \tag{2}$$

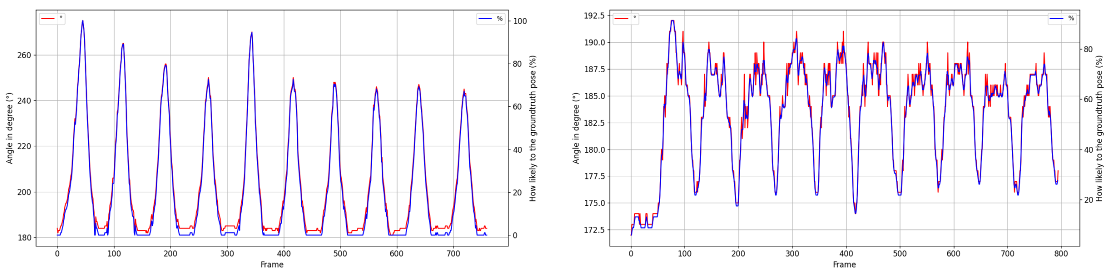
Table 1 showcases the angles between two rays formed by three joint points on different exercises. The correct angles were defined by meticulously collecting data from McFIT's publicly available videos. These videos feature professional trainers performing gym exercises, and their form and technique were used as the ground truth reference for determining the correct angles. This approach provided accurate and reliable measurements, crucial for developing and calibrating our model [12].

Table 1
Keypoint combination for selected sports exercises.

Exercises	Keypoints (Angle Ranges)											
	5	7	9	11	13	15	6	8	10	12	14	16
Pushups		[210°,280°]						[210°,280°]				
Bicep Curls		[010°,150°]						[010°,150°]				
Lunges					[142°,321°]						[153°,066°]	
Squats					[220°,280°]						[220°,280°]	
Shoulder Press		[026°,180°]						[026°,180°]				
Shoulder Lateral Raise		[171°,194°]						[171°,194°]				

3. Performance Analysis, Error Classification, and Feedback

Our research proposed ideal joint angle ranges for five exercises, configurable as hyper-parameters by fitness experts. We tracked the joint angles of participants during these exercises, comparing them to a ground-truth pose to assess correctness. A real-time monitoring system provided immediate feedback, allowing participants to improve their technique. Figure 2 illustrates angle variations between keypoints during the exercises. The x-axis denotes the frame number, acting as a time-based index capturing the progression of the movement. At the same time, the y-axis measures the angle in degrees (°), offering insights into how the keypoints' angles change. A red line traces these angle variations, visually representing the exercise's evolving pose. Examining these plots allows us to understand how these angles change dynamically during the training, helping to analyze movement patterns and identify key moments of interest.



(a) Squats

(b) Shoulder Lateral Raise

Figure 2: Performance tracking of different exercises.

AI algorithms and computer vision offer the potential to improve fitness tutoring systems by providing assessments and feedback. While challenges persist in enhancing pose estimation accuracy and refining feedback mechanisms, integrating these technologies into real-world

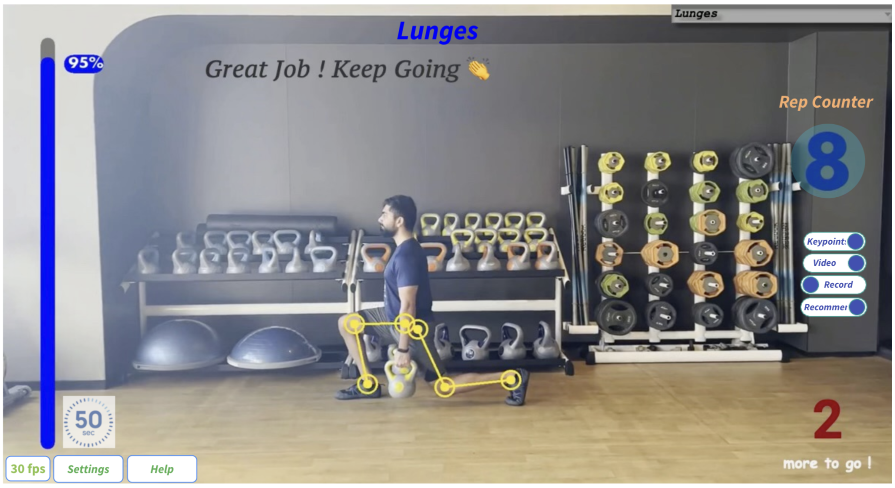


Figure 3: Prototype of the GUI.

training environments shows significant promise. The potential benefits include promoting proper form, reducing injury risk, and enhancing overall performance. It makes it a vibrant field of ongoing research and development.

In the scope of our ongoing research, we are making significant strides in creating a Graphical User Interface (GUI) for our innovative fitness tutoring system. In the prototype stage, our GUI already demonstrates functionality such as loading pre-recorded videos and processing them to display a performance progress bar and a repetition counter. Yet, it's important to note that the integration with live webcam feeds is still in development. These capabilities are rooted in a comprehensive user study encompassing a broad spectrum of end-users, ranging from fitness enthusiasts and professional trainers to tech-savvy individuals. This user-focused design methodology ensures our GUI development aligns with user needs, preferences, and expectations [17].

Through our diverse user study, we have gained valuable insights into the complexities of designing a fitness tutoring system. Thanks to various methods such as interviews, surveys, and interactive sessions, we have amassed a wealth of information on the GUI's desired functionalities, visual aesthetics, and usability. Adopting an iterative design strategy grounded in Human-Computer Interaction (HCI) principles, we continually refine our GUI prototype based on user feedback and interactions. Figure 3 represents a prototype of the graphical user interface (GUI) we strive to create. This cycle of design, evaluation, and refinement positions us to better align each iteration of the GUI with the evolving user needs and expectations.

The ongoing development of our fitness tutoring system's GUI demonstrates an integration of critical features like a performance bar and rep counter, which signal the progress of the pose prediction model. In this user-centric design, an embedded feedback mechanism is pivotal in establishing an interaction loop with the users, allowing real-time input that further refines the system.

Other aspects, such as Feedback, Webcam, and additional User Interface (UI) functionalities, are currently being scrutinized in the user study phase. It is all done with the overarching goal of creating an accessible, intuitive, and user-friendly experience. Features like displaying the remaining reps under the rep counter and strategic positioning of GUI functionalities are essential in achieving this goal. We are considering integrating audio feedback as part of our future enhancements, based on the results of our user study, which included a variable related to this aspect. Previous research has also demonstrated that audio feedback contributes to improved performance in training [18]. By incorporating audio cues, we aim to further engage users effectively and facilitate precise pose prediction, ultimately offering enhanced performance assessment in the athletic realm.

4. Conclusion

In summary, our research emphasizes prioritizing users and continually improving the design of our fitness tutoring system's graphical user interface (GUI). By integrating augmented intelligence and computer vision technologies, our system offers personalized coaching, instant feedback, progress tracking, and customized training programs, ultimately enhancing the overall user experience. Advanced pose estimation models and human topology-based tracking give users real-time guidance on exercise poses, even without a physical trainer. Furthermore, implementing transfer learning techniques has significantly increased the system's efficiency by eliminating the necessity for extensive model retraining. We have observed positive engagement and learning outcomes by aligning professional fitness videos as benchmarks, suggesting potential applicability to similar sports exercises. Our ongoing efforts will focus on refining this robust, engaging, and inclusive design based on user interactions, feedback, and evolving requirements.

Acknowledgment

The authors would like to thank the German Federal Ministry of Education and Research (BMBF) for their kind support within the project *Multimodal Immersive Learning with Artificial Intelligence for Psychomotor Skills* (MILKI-PSY) under the project ID 16DHB4014.

References

[1] D. Maji, S. Nagori, M. Mathew, D. Poddar, Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2637–2646.

[2] K. N. Mileva, L. Zaidell, Sport and exercise science and health, in: Health studies: An introduction, Springer, 2022, pp. 85–124.

[3] Y. Kowsar, M. Moshtaghi, E. Velloso, L. Kulik, C. Leckie, Detecting unseen anomalies in weight training exercises, in: Proceedings of the 28th Australian Conference on Computer-Human Interaction, OzCHI '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 517–526.

[4] A. Farrokhi, R. Farahbakhsh, J. Rezazadeh, R. Minerva, Application of internet of things and artificial intelligence for smart fitness: A survey, Computer Networks 189 (2021) 107859.

[5] Ç. B. Erdaş, S. Güney, Human activity recognition by using different deep learning approaches for wearable sensors, Neural Processing Letters 53 (2021) 1795–1809.

[6] M. Xie, Design of a physical education training system based on an intelligent vision, Computer Applications in Engineering Education 29 (2021) 590–602.

[7] H. Novatchkov, A. Baca, Artificial intelligence in sports on the example of weight training, Journal of sports science & medicine 12 (2013) 27.

[8] M. Ueta, Improving mirror fitness through augmented reality technology, in: 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), IEEE, 2022, pp. 186–194.

[9] M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, R. Hammad, The Multimodal Learning Analytics Handbook, Springer Nature, 2022.

[10] J. A. Ruipérez-Valiente, R. Martínez-Maldonado, D. Di Mitri, J. Schneider, From sensor data to educational insights, 2022.

[11] C.-Y. Wang, A. Bochkovskiy, H.-y. Liao, Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.

[12] N. Duong-Trung, H. Kotte, M. Kravčík, Augmented intelligence in tutoring systems: A case study in real-time pose tracking to enhance the self-learning of fitness exercises, in: Proceedings of the 18th European Conference on Technology Enhanced Learning (ECTEL), 2023.

[13] A. C. Tran, N. C. Tran, N. Duong-Trung, Recognition and quantity estimation of pastry images using pre-training deep convolutional networks, in: Future Data and Security Engineering, Springer, 2020, pp. 200–214.

[14] A. C. Tran, H. T. Nguyen, V. L. N. Huu, N. Duong-Trung, Pesticide label detection using bounding prediction-based deep convolutional networks, in: Future Data and Security Engineering, Springer, 2021, pp. 195–210.

[15] A. C. Tran, P. K. Thoa, N. C. Tran, N. Duong-Trung, et al., Real-time recognition of medicinal plant leaves using bounding-box based models, in: 2020 International Conference on Advanced Computing and Applications (ACOMP), IEEE, 2020, pp. 34–41.

[16] A. Garbett, Z. Degutyte, J. Hodge, A. Astell, Towards understanding people's experiences of ai computer vision fitness instructor apps, in: Proceedings of the 2021 ACM Designing Interactive Systems Conference, DIS '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 1619–1637.

[17] M. Geisen, S. Klatt, Real-time feedback using extended reality: A current overview and further integration into sports, International Journal of Sports Science & Coaching 17 (2022) 1178–1194.

[18] F. Sors, M. Murgia, I. Santoro, T. Agostini, Audio-based interventions in sport, The Open Psychology Journal 8 (2015) 212–219.