Analyzing Exercise Repetitions: YOLOv8-enhanced Dynamic Time Warping Approach on InfiniteRep Dataset

Michal Slupczynski[1] [0000−0002−0724−5006],
Aleksandra Nekhviadovich[1] [0000−0001−9355−6443],
Nghia Duong-Trung[2] [0000−0002−7402−4166], and
Stefan Decker[1] [0000−0002−2296−3401]

[1] Information Systems and Databases, RWTH Aachen University, Aachen, Germany
slupczynski@dbis.rwth-aachen.de
https://dbis.rwth-aachen.de
[2] Educational Technology Lab,
German Research Center for Artificial Intelligence (DFKI), Berlin, Germany
https://www.dfki.de/web/forschung/forschungsbereiche/educational-technology-lab

Abstract. This paper presents a novel approach to exercise repetition analysis using the YOLOv8-pose model and Dynamic Time Warping (DTW) techniques applied to the InfiniteRep dataset. Our research addresses the challenges of accurate pose estimation and tracking in dynamic camera environments and with varying occlusions in synthetic datasets. By integrating YOLOv8's pose detection capabilities with the temporal analysis strength of DTW, we propose a method that significantly improves the detection and classification of exercise repetitions across diverse conditions. We demonstrate the effectiveness of this approach through rigorous experiments that test various scenarios, including changes in camera angles and exercise complexity. Our results indicate notable improvements in the accuracy and robustness of exercise recognition, suggesting promising applications in sports science and personal fitness coaching.

Keywords: Synthetic Dataset · InfiniteRep · Pose Tracking · Exercise Repetition Detection · Dynamic Time Warping (DTW) · YOLO · Fitness

1 Introduction and Motivation

Human action recognition is pivotal in various applications within multimedia computing, including intelligent surveillance, virtual reality, and human-computer interaction [10,1]. In sports and fitness, artificial intelligence (AI) assists humans in decision-making and problem-solving [17]. Garbett et al. conducted an intensive comparison and user evaluation of six AI fitness instructor applications [4]. This technology can track an individual's movements, analyze their performance data, and provide suggestions for improvement.

This is especially important for activities that require the learning of complex motor movements. Despite progress in computer vision for action recognition [6,3], the complexity and variability of human movements, combined with challenging datasets for pose tracking, limit the accuracy and efficiency of current models. However, many state-of-the-art models struggle to accurately recognize and track human poses with dynamic camera movements [20,21]. Additionally, the use of synthetic datasets poses another challenge. These datasets are often created by rendering images from parameterized 3D human models, involving complex processes like shaping, posing, dressing, and texturing. While these rendered images provide precise annotations, they can mislead pose estimation models due to their artificial nature.

1.1 Research Questions and contributions

To guide our investigation and address the challenges identified in pose estimation and exercise repetition analysis, we formulated the following research questions:

RQ 1 Are there specific exercise types or movement patterns within the InfiniteRep dataset that are more susceptible to inaccuracies in pose estimation using YOLOv8 default model?

RQ 2 Are there strategies to mitigate camera angle, position variations, and body occlusions to maintain effectiveness in detecting exercise repetitions?

RQ 3 How can missing values and occlusions be effectively handled in pose estimation for exercise repetition analysis?

This paper presents a novel approach for exercise repetition analysis utilizing the YOLOv8-pose model [16] and Dynamic Time Warping (DTW) techniques [7,12,11,15] applied to the InfiniteRep dataset. Our research tackles the challenges of accurate pose estimation and tracking in dynamic camera environments and under varying occlusions within synthetic datasets. By integrating YOLOv8's pose detection capabilities with the temporal analysis strengths of DTW, we propose a method that significantly enhances the detection and classification of exercise repetitions across various conditions.
We present several contributions as follows.

Application to InfiniteRep Dataset This dataset encompasses substantial environmental variations and comprehensive annotations, making it an invaluable resource for evaluating the proposed methods. This emphasis addresses gaps in current research, which frequently depends on less diverse and richly annotated datasets.

Integration of YOLOv8 and DTW This combination improves the accuracy and robustness of exercise repetition detection and classification, especially in dynamic camera environments and under varying occlusions, surpassing existing methods that typically manage these tasks independently.

These methods are critical for maintaining the accuracy of pose tracking under challenging conditions, addressing a common issue where occlusions and non-detections can significantly degrade performance.

Real-Time and Post-Workout Analysis Algorithms This study proposes two distinct algorithms for exercise repetition detection: one for real-time (during-workout) analysis and another for post-workout analysis. The real-time algorithm delivers immediate feedback and correction, essential for sports coaching and physical therapy applications. In contrast, the post-workout algorithm enables a comprehensive review and detailed analysis after the exercise session, enhancing utility for various user needs.

Rule Creation Interface A web-based application has been developed, enabling teachers and students to create and perform exercise routines. This interface leverages the motion detection and feedback mechanisms described in this paper, making advanced techniques accessible for practical use. The interactive rule creation tool allows users to define specific feedback rules, enhancing the educational and training value of the system.

2 Related Work on InfiniteRep Dataset

The InfiniteRep dataset[3] [20] is an open-source synthetic dataset designed for fitness and physical therapy applications. It features videos of diverse avatars performing multiple repetitions of common exercises, capturing significant variations in environment, lighting conditions, avatar demographics, and movement trajectories. This variability ensures that each repetition mimics real human performance differences. Key features of the InfiniteRep dataset include a comprehensive collection of 1,000 videos, distributed across 10 distinct exercises, each represented by 100 videos. The exercises covered in this dataset are pushups, alternating bicep curls, delt flys, squats, bird dogs, supermans, bicycle crunches, leg raises, front raises, and overhead presses. The dataset provides extensive annotations and metadata, including bounding boxes, segmentation masks, keypoints, joint angles, repetition counts, avatar characteristics, and camera settings.

Such detailed annotations are particularly valuable for various computer vision and machine learning tasks, enhancing the dataset's utility for research and application development. Regarding format and accessibility, the videos are provided in a 224x224 RGB format at 24 frames per second (fps).

To date, limited research has been conducted using the InfiniteRep dataset. We identified two notable studies in existing literature: Chang et al. [2] implemented a Spatio-Temporal Graph Convolutional Network (ST-GCN) for human action recognition to assess users' fitness statuses, utilizing skeleton data as input to model inter-skeleton connections. This approach was validated using the InfiniteRep dataset, demonstrating high accuracy.

[3] https://marketplace.infinity.ai/pages/infiniterep-dataset

Conversely, Pande et al. [9] developed Fitwave, a fitness application designed to monitor and correct users' exercise postures. They employed transfer learning techniques on a pre-trained MobileNet architecture, refining their model using the InfiniteRep dataset with a focus on three exercises: arm raises, bicep curls, and squats. Despite advances in computer vision, the complexity of human movements and the variability in pose tracking datasets present significant challenges. Current models often fail to accurately track human poses in dynamic environments, particularly when using synthetic datasets. Integrating YOLOv8 for pose estimation with DTW for temporal sequence analysis offers a promising solution, significantly improving the detection and classification of exercise repetitions.

3 Technical Background

In this section, we introduce the underlying algorithms used in our application. First, we examine the YOLOv8 algorithm and its application in pose detection. Following this, we describe two algorithms designed for post- and during-workout repetition detection.

3.1 YOLOv8-Pose

The YOLO (You Only Look Once) architecture [16] became a key object detection algorithm by framing the problem as a single regression task, directly predicting bounding boxes and class probabilities from full images in a single evaluation. This approach contrasted with previous methods that required region proposal networks or sliding windows, thereby significantly reducing computation time and enabling real-time performance. Subsequent versions, YOLOv1 through YOLOv10, introduced various improvements, such as batch normalization, anchor boxes, multi-scale training, and feature pyramid networks, which collectively enhanced the models' speed and accuracy [5,19,18]. YOLOv8-Pose, the latest pose estimation framework from the YOLO models, builds on these advancements and focuses on pose estimation, a complex task that involves detecting keypoints on human bodies and mapping their spatial relationships. This model incorporates several key innovations:

Enhanced Backbone Network: YOLOv8 employs an enhanced backbone network that leverages advancements in convolutional neural network (CNN) architectures, such as deeper networks with more efficient layers, to capture more intricate features from input images.

Keypoint Detection Figure 1 illustrates the YOLOv8 joint detection output, enhanced for exercise repetition analysis. Each keypoint corresponding to a body joint is marked and labeled with a unique identifier for precise tracking. The keypoints include the nose (1), eyes (2, 3), ears (4, 5), shoulders (6, 7), elbows (8, 9), wrists (10, 11), hips (12, 13), knees (14, 15), and ankles (16, 17).

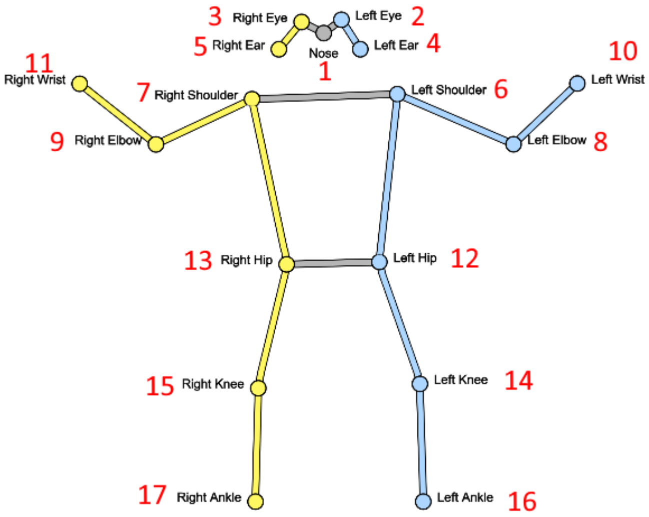


Fig. 1: Keypoint ID description.

Multi-Scale Feature Extraction: By integrating feature pyramid networks (FPN) and path aggregation networks (PAN), YOLOv8-Pose effectively extracts and utilizes features at multiple scales. This multi-scale approach is crucial for accurately detecting keypoints across a range of human poses and body sizes.

Keypoint Localization Segment Head: The pose estimation segment head in YOLOv8-Pose is designed to predict keypoints with high precision. It uses specialized loss functions and optimization techniques to ensure accurate localization of human joints and key body parts.

Real-Time Inference: Adhering to the YOLO philosophy, YOLOv8-Pose is optimized for real-time inference, making it suitable for applications requiring immediate feedback, such as motion capture, interactive fitness applications, and real-time video analytics. YOLOv8-Pose's application domains are diverse, spanning sports analytics, physical therapy, augmented reality, and human-computer interaction. Its ability to provide real-time, accurate pose estimation enables new interactive technologies and enhances user experiences. However, challenges persist, particularly when dealing with complex scenes, varying lighting conditions, and occlusions. The performance of YOLOv8-Pose is highly dependent on the quality and diversity of its training data. For instance, models pre-trained on standard datasets may struggle with domain-specific datasets like InfiniteRep, which includes various environmental variations such as occlusions.

Pose detection accuracy In some frames, YOLO fails to detect the person, resulting in no skeletal data. In other frames, certain joints are not detected correctly, causing their coordinates to be recorded as zeros. Consequently, those joints are marked as $[0.0, 0.0]$ in the skeletal data instead of having a valid position.

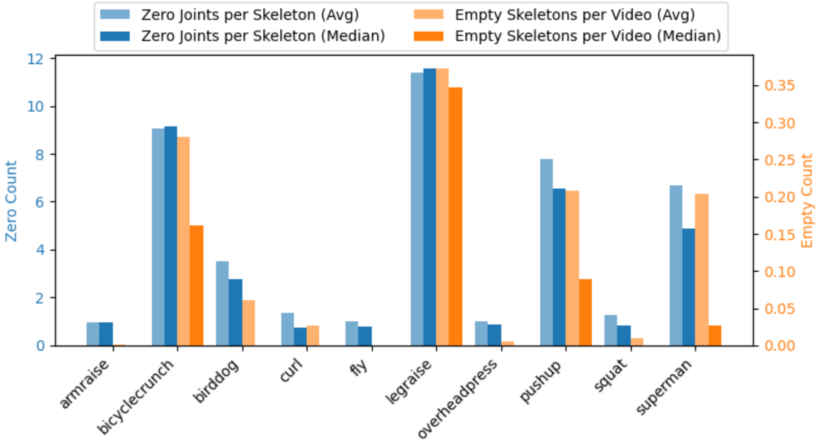


Fig. 2: YOLOv8l applied to InfiniteRep exercises: Comparison of average and median percentage of empty skeletons and zero joints *(higher is worse)*.

As illustrated in Figure 2, when applied to the InfiteRep dataset, YOLOv8 is not ideal for estimating poses in exercises where the person is lying on the floor rather than standing, such as Bicycle Crunches, Bird Dogs, Leg Raises, Pushups, and Supermans. However, YOLOv8 performs better at estimating poses in exercises where the person is standing, such as Front Arm Raises (with dumbbells), Alternating Bicep Curls (with dumbbells), Delt Flys (with dumbbells), Overhead Press, and Squats. Simultaneously, the occlusion percentage in some videos is so high that YOLOv8 cannot be expected to detect anything. Analyzing the average occlusion percentage per video reveals that, for arm raises, the maximum average occlusion percentage is 22.64%, whereas for bicycle crunches it reaches 97.82%. Additionally, for a standing person, occlusion often affects the legs. This is problematic for exercises that involve leg movements, such as squats.

3.2 The Post-workout and During-workout Algorithms

We employed slightly different algorithms for the post-workout and during-workout exercise repetition detection. The during-workout algorithm is designed for real-time analysis, where the learner performs actions in front of the camera.

The algorithm incrementally processes the data, continuously updating window parameters to detect exercise repetitions as they occur. This enables immediate feedback and correction. In contrast, the post-workout algorithm operates offline on uploaded videos, analyzing the entire sequence at once. It identifies multiple best matches and evaluates them against dynamic criteria, making it suitable for detailed post-exercise review and analysis without the need for immediate feedback. This distinction allows the during-workout algorithm to provide instant guidance, while the post-workout algorithm comprehensively evaluates the entire workout session.

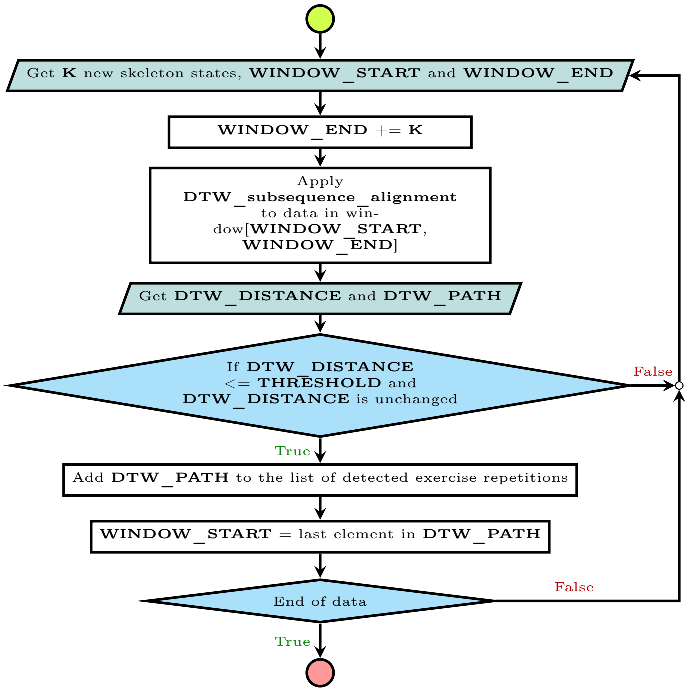

Get K new skeleton states, WINDOW_START and WINDOW_END
WINDOW_END += K
Apply DTW_subsequence_alignment to data in window[WINDOW_START, WINDOW_END]
Get DTW_DISTANCE and DTW_PATH
If DTW_DISTANCE <= THRESHOLD and DTW_DISTANCE is unchanged
False
True
Add DTW_PATH to the list of detected exercise repetitions
WINDOW_START = last element in DTW_PATH
End of data
False
True

Fig. 3: Flowchart for DTW-based during-workout algorithm.

During-workout algorithm - Continuous analysis The during-workout algorithm (see Fig. 3) operates using a data stream, continuously analyzing the learner's movements in real-time. Initially, the window's start is set to 0, and the end is advanced with each new frame received. When K new skeleton states are collected, the DTW alignment begins. The algorithm retrieves the DTW distance and path within this window. It detects an exercise repetition if the DTW distance is below a specified threshold and remains unchanged for I iterations or S seconds. The start of the window is then updated to the last frame of this detected repetition, allowing the process to continue.

If the DTW distance criterion is not met, the algorithm updates the window parameters and repeats the alignment and evaluation steps, ensuring continuous and immediate feedback for the learner.

Post-workout - Batch processing The post-workout algorithm (see Fig. 4) starts by applying Dynamic Time Warping (DTW) subsequence alignment to compare the learner's trajectory with an expert's trajectory. This algorithm generates a list of the K best matches. The algorithm then selects the best unchecked match, denoted as M, and checks if M's DTW distance is below a predefined threshold and within a factor T of the maximum DTW distance from the last checked match.

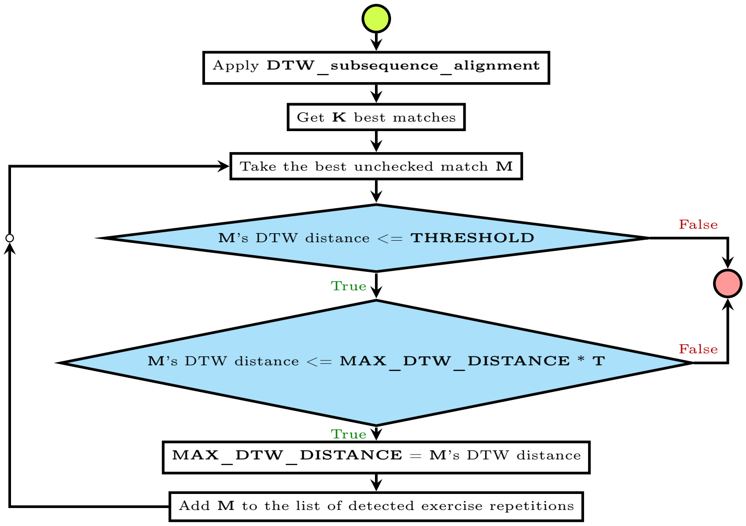

Apply DTW_subsequence_alignment
Get K best matches
Take the best unchecked match M
M's DTW distance <= THRESHOLD
False
True
M's DTW distance <= MAX_DTW_DISTANCE * T
False
True
MAX_DTW_DISTANCE = M's DTW distance
Add M to the list of detected exercise repetitions

Fig. 4: Flowchart for DTW-based post-workout algorithm

If M meets these criteria, it is added to the list of detected exercise repetitions, and the maximum DTW distance is updated to M's distance. If not, M and all remaining unchecked matches are discarded. This process continues until all relevant matches have been evaluated, ensuring a comprehensive post-workout analysis of the exercise repetitions.

Replacing YOLO non-detection frames For post-workout analysis, to address missing values caused by YOLO model non-detection, we interpolate unknown values by identifying the nearest known values before (see Fig. 5) and after the gap (see Fig. 6). Unknown angles or distances are replaced with values that transition from the nearest known value before the gap to the nearest known value after it. Similarly, zero joint coordinates are replaced with coordinates that gradually change from the nearest known value before the gap to the nearest known value after it.

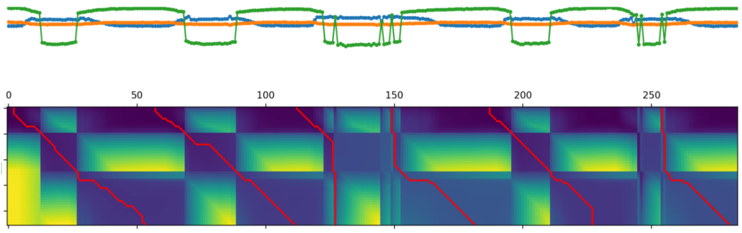


Fig. 5: DTW trajectories before filtering out outliers

Since DTW sub-trajectory alignment is sensitive to outliers, we filter them by examining data within a specified window size. If values in this window deviate significantly from the surrounding data, they are considered outliers and are replaced gradually to maintain data consistency. For during-workout analysis, we replace unknown angle or distance values with the last known angle; and zero joint coordinates with their last known coordinates. To filter outliers, we examine the data within a window size of 1. If an outlier is detected within this small window, it is replaced with the last known value, ensuring real-time consistency and accuracy in pose detection during the workout.

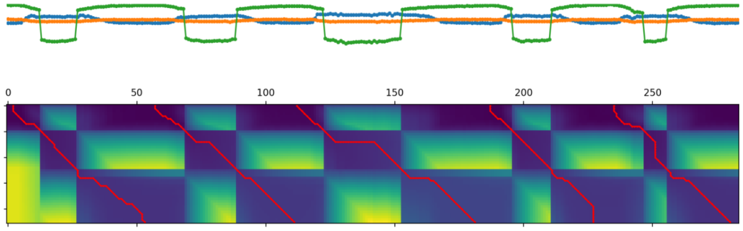
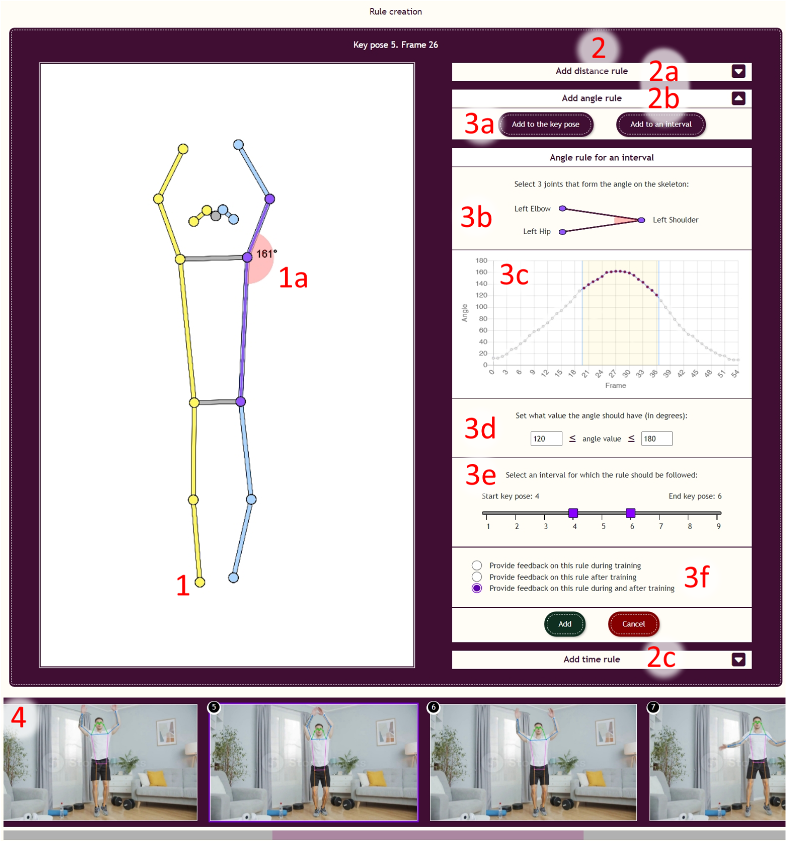


Fig. 6: DTW trajectories after filtering out outliers

4 Rule Creation Interface description

A web-based application was built to enable learners and teachers to apply the proposed algorithm in practice. The user interface is designed to streamline the creation and performance of exercise routines, leveraging the described motion detection and feedback mechanisms. The back-end architecture of this system follows a multimodal sensor-based cloud pipeline [13,14]. The UI design for defining feedback rules focused on using angles to detect motion differences [8], ensuring consistency regardless of body shape. Expert feedback led to incorporating relative distances between body parts for more detailed feedback. Preliminary interviews with sports trainers and students provided useful insights, but a small sample size limited the statistical significance of the findings. Nonetheless, this feedback played a crucial role in iterating the design to better meet the needs of the users and ensure the effectiveness of the feedback system in real-world training scenarios. Students interact with the system by selecting exercises based on their thumbnails and descriptions and performing them in front of a webcam. The system then uses YOLOv8 to extract their skeleton data, match exercise repetitions, and provide immediate and summative feedback based on rules predefined by the teachers. Teachers initiate the exercise creation process by providing an expert recording of the exercise. From this recording, they select key poses and define specific rules for the algorithm to check, ensuring that learners perform the exercises accurately.

The underlying algorithm interpolates between the key poses defined in the exercise, identifying the closest match to apply the relevant rules to the learner's motion. The rule creation interface for teachers (see Figure 7) allows for versatile input of feedback guidelines.


Rule creation

Key pose 5. Frame 26

Angle rule for an interval

Select 3 joints that form the angle on the skeleton:

Set what value the angle should have (in degrees):

120 ≤ angle value ≤ 180

Select an interval for which the rule should be followed:

Start key pose: 4

End key pose: 6

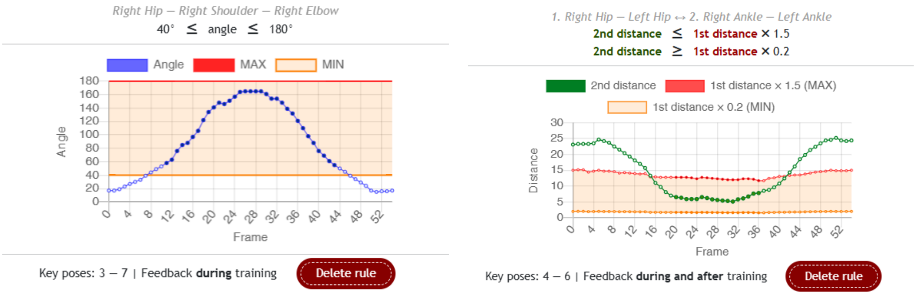
○ Provide feedback on this rule during training
○ Provide feedback on this rule after training
● Provide feedback on this rule during and after training

Fig. 7: Rule Creation UI.

Teachers can view a list of key poses **(4)** and select any key pose to see the corresponding extracted skeleton on the left side of the UI **(1)**. For each key pose or interval between key poses, teachers can define various rules **(2)**: distance rules **(2a)** specify the required distance between certain body joints (e.g., the distance between the left and right hand must be at least 1.5 times the distance between the shoulders). To do this, two distances are selected. For visualization purposes and to assist the user, the first distance is used to calculate a unit of measurement. The unit of measurement is 1/10 the first distance. The second distance is calculated using this unit of measurement. Then a chart shows the change of these two distances over time.

Angle rules **(2b)** set intervals for the angles between body joints (e.g., the angle between the arm and hip at the shoulder joint must be at least 120°); and time rules **(2c)** determine the timing requirements for holding or transitioning between key poses. Teachers can choose to apply these rules to a single key pose or across an interval between selected key poses **(3a)**. By clicking on the skeleton, teachers can select the body parts to be considered for angle rules **(1a)**. The application pre-computes the expert motion angles and visualizes the target motion range, simplifying the rule-creation process. Teachers then specify the angle interval for the motion **(3d)** and determine the range of key poses to which the rule applies **(3e)**. Additionally, they can decide whether feedback should be provided during the execution of the learner's motion or afterward in the summative section **(3f)**.

Once a rule is created, the system displays a list of active rules with their parameters and provides a visualization of the motion range relative to the expert recording (see Figure 8).


Right Hip – Right Shoulder – Right Elbow

40° ≤ angle ≤ 180°

Key poses: 3 – 7 | Feedback **during** training

1. Right Hip – Left Hip ↔ 2. Right Ankle – Left Ankle

2nd distance ≤ 1st distance × 1.5

2nd distance ≥ 1st distance × 0.2

Key poses: 4 – 6 | Feedback **during and after** training

(a) Angle Rule Description.

(b) Distance Rule Description.

Fig. 8: Rule Description UI.

In the depicted angle rule example (see Figure 8a), we can see a rule for the angle formed by the "Right Hip – Right Shoulder – Right Elbow". The system shows the acceptable angle range as 40° to 180°. A graph illustrates the recorded angle over time, with the x-axis for frame number and the y-axis for angle in degrees. The blue line represents the actual angle, while the red and orange lines mark the maximum (180°) and minimum (40°) allowable angles, with the allowed range highlighted in light orange. Points within the specified interval between key poses, where the rule must be followed, are filled in blue, and points outside this interval are filled in white. Additionally, the interface specifies that this rule applies to specific key poses and provides feedback during training. This allows teachers to review and adjust rules as necessary, ensuring that the exercises are both precise and effective. Overall, this structured and interactive UI ensures that teachers can create detailed and accurate exercises while students receive precise real-time feedback, enhancing their learning experience.

5 Experimental Results

We evaluated the effectiveness of our proposed method by analyzing the detection and classification accuracy of exercise repetitions using both the YOLOv8 pose model and DTW techniques. For our evaluation, we considered a detected repetition to be a true positive (TP) if the left and right boundaries differed by no more than 30% of the length of the repetition from the corresponding boundaries in the InfiniteRep dataset. We specifically used accuracy as our primary metric for evaluation, defined as Accuracy = $^{TP}/_{N}$, where TP is the number of true positives, and N is the total number of repetitions.

We analyzed 30 videos for each exercise, focusing on various angles and distances. Specifically, we identified the top 10 angles and distances for front arm raises. The optimal angles and relative distances are summarized in Table 1.

Table 1: 10 Best angles and rel. dist. for “Front Arm Raises (With Dumbbells)”

	Accuracy	IDs of joints	ID decoding	
Angles				
1	0.9029	(10, 11, 12)	Angle: Left Wrist - Right Wrist - Left Hip	
2	0.8854	(6, 9, 11)	Angle: Left Shoulder - Right Elbow - Right Wrist	
3	0.8769	(12, 9, 17)	Angle: Left Hip - Right Elbow - Right Ankle	
4	0.8739	(9, 10, 1)	Angle: Right Elbow - Left Wrist - Nose	
5	0.8724	(7, 8, 10)	Angle: Right Shoulder - Left Elbow - Left Wrist	
6	0.8719	(10, 6, 13)	Angle: Left Wrist - Left Shoulder - Right Hip	
7	0.8717	(12, 9, 15)	Angle: Left Hip - Right Elbow - Right Knee	
8	0.8704	(7, 10, 8)	Angle: Right Shoulder - Left Wrist - Left Elbow	
9	0.8629	(10, 8, 15)	Angle: Left Wrist - Left Elbow - Right Knee	
10	0.8585	(11, 9, 1)	Angle: Right Wrist - Right Elbow - Nose	
Relative Distances (1st distance and 2nd distance – used to calculate a unit of measurement)				
1	0.9505	(6, 11) | (9, 13)	Dist.: Left Shoulder - Right Wrist	Rel. to: Right Elbow - Right Hip
2	0.9472	(11, 1) | (9, 12)	Dist.: Right Wrist - Nose	Rel. to: Right Elbow - Left Hip
3	0.9418	(11, 1) | (9, 13)	Dist.: Right Wrist - Nose	Rel. to: Right Elbow - Right Hip
4	0.9406	(7, 11) | (9, 13)	Dist.: Right Shoulder - Right Wrist	Rel. to: Right Elbow - Right Hip
5	0.9317	(11, 1) | (11, 15)	Dist.: Right Wrist - Nose	Rel. to: Right Wrist - Right Knee
6	0.9291	(9, 1) | (9, 13)	Dist.: Right Elbow - Nose	Rel. to: Right Elbow - Right Hip
7	0.9277	(8, 1) | (9, 13)	Dist.: Left Elbow - Nose	Rel. to: Right Elbow - Right Hip
8	0.9241	(6, 9) | (9, 13)	Dist.: Left Shoulder - Right Elbow	Rel. to: Right Elbow - Right Hip
9	0.9238	(7, 14) | (9, 13)	Dist.: Right Shoulder - Left Knee	Rel. to: Right Elbow - Right Hip
10	0.9228	(11, 17) | (6, 1)	Dist.: Right Wrist - Right Ankle	Rel. to: Left Shoulder - Nose

The results suggest that our method can effectively be used for detecting and classifying exercise repetitions. Accuracy metrics reveal that certain angles and relative distances are more reliable for correct repetition detection in dynamic camera environments and with varying occlusions. For instance, the angle formed by between the Left Wrist - Right Wrist - Left Hip achieved the highest accuracy of 0.9029, showcasing the algorithm’s robustness in correctly identifying front arm raises. Additionally, the best relative distances, such as Left Shoulder - Right Wrist relative to Right Elbow - Right Hip 0.9505, offer further context for enhancing detection accuracy.

Findings indicate that specific joint configurations are essential for accurate recognition of exercise movements. Analyzing the joint angles, such as Wrist-Elbow-Shoulder, shows that these angles can be affected by the person’s orientation relative to the camera. For more accurate results, selecting joints that are not directly involved in the movement seems advantageous. For instance, when lifting arms, using the left and right wrist along with the nose, hip, or knee can improve detection accuracy (see Table 1).

Additionally, we compared the accuracy of after-workout and during-workout algorithms, utilizing combinations of the best angles and distances, in detecting exercise repetitions using the InfiniteRep dataset (see Figure 9). The evaluation included both ideal dataset data and data obtained with YOLOv8 models. The highest accuracy is achieved with the after-workout algorithm on ideal data, followed by the during-workout algorithm on ideal data. Accuracy decreases significantly when using YOLOv8 pose detection, especially for exercises performed on the floor, such as Bicycle Crunches, Bird Dogs, Leg Raises, Pushups, and Supermans, highlighting YOLOv8's limitations in scenarios where body parts of the trainees were occluded, e.g. due to them laying on the ground.

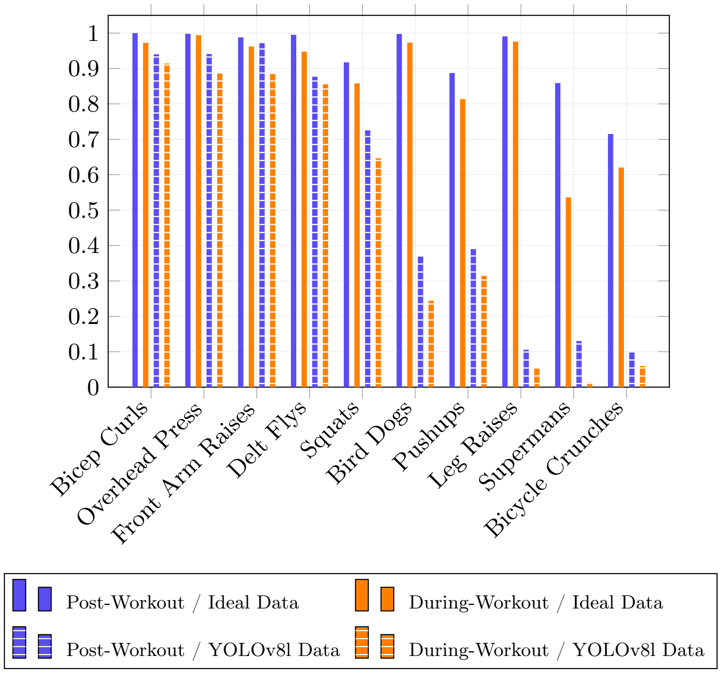


Fig. 9: Highest Accuracy Comparison Across Exercises Using Post- and During-Workout Algorithms with Ideal and YOLOv8l Pose Detection

6 Limitations

The design and implementation of our psychomotor learning application exhibit several limitations that may impact the system's effectiveness and applicability in diverse contexts. The manual rule creation process in the application is susceptible to expert error, and users may struggle to determine which expert rules to trust.

While the application's approach to analyzing angles and relative distances is designed to be non-discriminatory regarding body shape, it does not account for the user's flexibility or fitness level. Another limitation of our contribution is the lack of comparison of YOLOv8 with existing human pose models used in kinematic analysis, such as MediaPipe [4], OpenPose [5], or other available open source models. This comparison would be crucial for validating the system's performance against established benchmarks and could help address the limitations of YOLOv8. Additionally, while our method is not directly comparable with other studies utilizing the InfiniteRep dataset, as those typically focus on detecting exercise types without tracking the exact number of repetitions or their timing, applying our algorithm to other datasets would make such a comparison feasible. As mentioned, the default YOLOv8 models present challenges in detecting poses for users who are lying down or are heavily occluded. Furthermore, YOLOv8 does not detect finer skeleton details such as fingers or toes, which could be beneficial for certain exercises requiring detailed analysis. Finally, the UI evaluation was conducted with a small sample size, limiting the statistical significance of the findings. Further studies with a larger participant pool are necessary to draw more robust conclusions and to drive a thorough UI design process.

7 Future Work

Future work could validate our approach across multiple heterogeneous datasets to understand its generalizability in various real-world scenarios. Given that our algorithm's performance on the InfiniteRep dataset isn't directly comparable with other studies, further evaluations on additional datasets and against other algorithms are necessary to establish broader applicability and effectiveness.

Additionally, introducing a rating system for exercises and fitness experts could enhance trust in the exercise creation process, mitigating potential errors in manual rule creation. To better accommodate users' varying fitness and flexibility levels, future iterations could incorporate detailed difficulty or expertise levels into exercises.

The system could be extended to support group training sessions, where real-time feedback is provided to multiple users simultaneously. This could be particularly useful for sports teams or fitness classes, where individual and group performance can be monitored and adjusted on the fly.

Integrating biomechanical models with pose estimation could improve accuracy by considering joint constraints and physical properties. This would imply implementing a model that assesses the risk of injury based on detected pose and movement patterns. Such a system could alert users to potential risks and suggest safer alternatives or modifications to exercises based on their form and physical condition.

[4] https://github.com/google-ai-edge/mediapipe

[5] https://github.com/CMU-Perceptual-Computing-Lab/openpose

Developing personalized feedback models that adapt to individual performance histories could tailor workouts to specific needs and goals. An adaptive feedback model could not only provide feedback but also adapt their coaching style based on the user's emotional state and motivation levels.

Evaluating user progress over time would provide valuable data to refine the algorithms, enhancing their predictive capabilities and adapting to long-term trends in user performance.

A critical area for future work involves comparing our pose estimation approach with existing kinematic analysis models like MediaPipe, OpenPose, or other open source models. This comparison could reveal areas for improvement or potential alternatives that might outperform YOLOv8 in certain scenarios. Moreover, addressing the limitations of YOLOv8, such as its difficulty in detecting poses for individuals lying down or occluded, as well as its inability to detect finer details like fingers or toes, is essential. Future work could also explore placing cameras on ceilings, fine-tuning YOLOv8, or exploring other pose detection models like MediaPipe to overcome these challenges.

Finally, a thorough UI evaluation with a larger and more diverse sample size is needed to gather statistically significant data. This feedback would inform a potential redesign, ensuring the interface is user-friendly and effective for a wider audience.

8 Summary and Conclusions

In this paper, we introduced an approach to exercise repetition analysis by integrating the YOLOv8-pose model with DTW techniques, specifically applied to the InfiniteRep dataset. This combination enhances the detection and classification of exercise repetitions, especially in dynamic environments and under varying occlusions, addressing the limitations of current state-of-the-art models.

We found that floor exercises like Bicycle Crunches, Bird Dogs, Leg Raises, Pushups, and Supermans are more susceptible to pose estimation inaccuracies using the YOLOv8 default model.

These exercises result in significant occlusions and complex body orientations, posing challenges for accurate keypoint detection and tracking (see **RQ1**).

Our method leverages the detailed annotations and environmental variations of the InfiniteRep dataset, including diverse lighting conditions, mirroring duplication, avatar demographics, and movement trajectories. We handled missing values and occlusions by proposing robust methods for interpolating unknown values and filtering outliers to improve accuracy (see **RQ2**). By developing during-workout and post-workout analysis algorithms, we offer solutions for immediate feedback during exercises and detailed reviews post-exercise, ensuring broad applicability from sports coaching to physical therapy. Zero joint coordinates can be substituted with their last known coordinates to maintain data continuity (see **RQ3**). For post-workout analysis, interpolation estimates missing values by identifying the nearest known values before and after gaps, creating a seamless transition.

For real-time analysis, continuously replacing unknown values with the most recent known values ensures consistency and accuracy, while filtering techniques identify and correct outliers in real-time. These approaches enhance the detection and classification of exercise repetitions under challenging conditions.

Our web-based application enhances practical utility by providing an interactive platform for creating exercise routines and performance assessments. This tool, with a rule creation interface, allows educators and trainers to tailor feedback and ensure precision in exercise execution. Our evaluation demonstrates the robustness and versatility of our approach across various exercises in the InfiniteRep dataset.

The experimental results highlight significant improvements in exercise recognition accuracy and robustness, suggesting promising applications in sports science and personal fitness coaching. This research advances the state-of-the-art in exercise repetition analysis and provides practical tools and methods for real-world settings. Integrating advanced pose estimation with temporal analysis opens new avenues for enhancing human motion analysis, significantly contributing to sports science.

Acknowledgments. This research was funded by the German Federal Ministry of Education and Research (BMBF) through the project "Multimodales Immersives Lernen mit künstlicher Intelligenz für Psychomotorische Fähigkeiten" ("MILKI-PSY"[6]) (grant no. 16DHB4015). We extend our gratitude to the BMBF for their generous support and investment in this research.

References

1. Beddiar, D.R., Nini, B., Sabokrou, M., Hadid, A.: Vision-based human activity recognition: a survey. Multimedia Tools and Applications **79**(41), 30509–30555 (2020)
2. Chang, J.W., Liu, H.R.: Applying 5PKC-Based skeleton partition strategy into spatio-temporal graph convolution networks for fitness action recognition. In: Hung, J.C., Chang, J.W., Pei, Y. (eds.) Innovative Computing Vol 2 - Emerging Topics in Future Internet. pp. 728–737. Springer Nature Singapore
3. Gammulle, H., Ahmedt-Aristizabal, D., Denman, S., Tychsen-Smith, L., Petersson, L., Fookes, C.: Continuous human action recognition for human-machine interaction: a review. ACM Computing Surveys **55**(13s), 1–38 (2023)
4. Garbett, A., Degutyte, Z., Hodge, J., Astell, A.: Towards Understanding People's Experiences of AI Computer Vision Fitness Instructor Apps. In: Designing Interactive Systems Conference 2021. pp. 1619–1637. ACM, Virtual Event USA (Jun 2021). https://doi.org/10.1145/3461778.3462094
5. Hussain, M.: Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection. Machines **11**(7), 677 (2023)
6. Morshed, M.G., Sultana, T., Alam, A., Lee, Y.K.: Human action recognition: A taxonomy-based survey, updates, and opportunities. Sensors **23**(4), 2182 (2023)

[6] https://milki-psy.de/

7. Müller, M.: Dynamic time warping. Information retrieval for music and motion pp. 69–84 (2007)
8. Paaßen, B., Baumgartner, T., Geisen, M., Riedl, N., Kravčík, M.: Few-shot keypose detection for learning of psychomotor skills (Oct 2022)
9. Pande, V., Mokashi, A., Patil, S., Singh, A., Jadhav, N.: Fitwave: A Posture Correction System Based on Machine Learning. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization, vol. 569, pp. 410–418. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-19958-5_38
10. Pareek, P., Thakkar, A.: A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. Artificial Intelligence Review **54**(3), 2259–2322 (2021)
11. Schneider, P., Memmesheimer, R., Kramer, I., Paulus, D.: Gesture recognition in rgb videos using human body keypoints and dynamic time warping. In: RoboCup 2019: Robot World Cup XXIII 23. pp. 281–293. Springer (2019)
12. Senin, P.: Dynamic time warping algorithm review. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA **855**(1-23), 40 (2008)
13. Slupczynski, M.P., Klamma, R.: MILKI-PSY Cloud: MLOps-based Multimodal Sensor Stream Processing Pipeline for Learning Analytics in Psychomotor Education. MILeS 2022 : Multimodal Immersive Learning Systems 2022 : proceedings of the Second International Workshop on Multimodal Immersive Learning Systems (MILeS 2022) at the Seventeenth European Conference on Technology Enhanced Learning (EC-TEL 2022) : Toulouse **France**, pages 8–14 (2022). https://doi.org/10.18154/RWTH-2022-09814
14. Slupczynski, M.P., Sanusi, K.A.M., Majonica, D., Klemke, R., Decker, S.J.: Implementing cloud-based feedback to facilitate scalable psychomotor skills acquisition. https://doi.org/10.18154/RWTH-2023-09769
15. Tchane Djogdom, G.V., Otis, M.J.D., Meziane, R.: Dynamic time warping–based feature selection method for foot gesture cobot operation mode selection. The International Journal of Advanced Manufacturing Technology **126**(9), 4521–4541 (2023)
16. Terven, J., Córdova-Esparza, D.M., Romero-González, J.A.: A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. Machine Learning and Knowledge Extraction **5**(4), 1680–1716 (2023)
17. Venkatachalam, P., Ray, S.: How do context-aware artificial intelligence algorithms used in fitness recommender systems? a literature review and research agenda. International Journal of Information Management Data Insights **2**(2), 100139 (2022)
18. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458 (2024)
19. Wang, C.Y., Yeh, I.H., Liao, H.Y.M.: Yolov9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616 (2024)
20. Weitz, A., Colucci, L., Primas, S., Bent, B.: Infiniteform: A synthetic, minimal bias dataset for fitness applications. arXiv preprint arXiv:2110.01330 (2021)
21. Xiu, Y., Yang, J., Cao, X., Tzionas, D., Black, M.J.: Econ: Explicit clothed humans optimized via normal integration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 512–523 (2023)