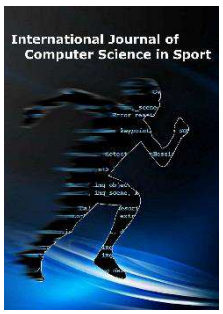
International Journal of Computer Science in Sport

Volume 23, Issue 1, 2024

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2024-0002



Automatic Detection of Faults in Simulated Race Walking from a Fixed Smartphone Camera

Tomohiro Suzuki[1], Kazuya Takeda[1] and Keisuke Fujii[1,2,3]*

1Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan; 2RIKEN Center for Advanced Intelligence Project, Fukuoka, Fukuoka, Japan; 3PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan

[*] *Corresponding author*

Abstract

Automatic fault detection is a major challenge in many sports. In race walking, judges visually detect faults according to the rules. Hence, automatic fault detection systems will help a training of race walking without experts' visual judgement. Some studies have attempted to use sensors and machine learning to automatically detect faults. However, there are problems associated with sensor attachments and equipment such as a high-speed camera, which conflict with the visual judgement of judges, and the interpretability of the fault detection models. In this study, we proposed an automatic fault detection system for non-contact measurement. We used pose estimation and machine learning models trained based on the judgements of multiple qualified judges to realize fair fault judgement. We verified them using smartphone videos of normal race walking and walking with intentional faults in several athletes including the medalist of the Tokyo Olympics. The results show that the proposed system detected faults with an average accuracy of over 90%. We also revealed that the machine learning model detects faults according to the rules. In addition, the intentional faulty walking movement of the medalist was different from that of other walkers. This finding informs realization of a more general fault detection model.

KEYWORDS: MACHINE LEARNING, MOTION ANALYSIS, RACE-WALKING, POSE ESTIMATION

Introduction

Race walking is one of the long-distance track and field Olympic events, which has been also studied as biomechanical and race analysis research (Gomez-Ezeiza et al., 2018; Miura et al., 2017; Menting et al., 2022; Pavei & La Torre, 2016). Race Walking is a progression of steps so taken that the walker makes contact with the ground, so that no visible (to the human eye) loss of contact occurs. The advancing leg must be straightened (i.e. not bent at the knee) from the moment of first contact with the ground until the vertical upright position (World Athletics, 2023). Violation of the first part is called loss of contact (LC) violation and that of the second part is called bent knee (BK) violation. The walker who commits these violations will be given a red card and disqualified after four red cards.

Judges in race walking check such faults visually (World Athletics, 2023). Then, the standard of judgement varies depending on the judges, calling into question the objectivity of the judgements (Hanley et al., 2019; Knicker & Loch, 1990). In addition, because judges check several walkers at the same time, they may not be able to fully observe all walkers. For spectators who are not familiar with race walking, it is difficult to understand whether a fault has been committed and to appreciate the sport. Therefore, it is important to realize objective and fair judgement. It is also necessary to clarify the reasons for the judgements and devise ways to make it easier for spectators to understand race walking.

To detect faults precisely, some approaches detected LC using piezoelectric sensors on the surface of the insole in contact with the sole (Santoso & Setyanto, 2013) or inertial measurement units (IMU) attached to the waist and lower limbs (Di Gironimo et al., 2016; Lee et al., 2013). Another study used machine learning to detect two faults using IMU data as input (Taborri et al., 2019). However, all of these studies used sensors attached to the body, which is not practical because they may affect the performance of walkers. In addition, some machine learning methods (e.g., support vector machine (Vapnik, 1999) used in Taborri et al., 2019) may make it difficult to interpret the reasons for the fault detection and sometimes be in conflict with the visual judgement of race-walk judges. Other sports use image recognition technology to solve similar problems. Examples include scoring in rhythmic gymnastics (Díaz-Pereira et al., 2014) and figure skating (Xu et al., 2019) and detecting offsides in soccer (Uchida et al., 2021). These studies use competition videos and do not require sensors. Similarly, in judging the race walking competition, automatic non-contact fault detection from the walking video would contribute to reducing the burden on judges and improving the objectivity of judgement.

The purpose of this study is to develop a fault detection system using a fixed smartphone camera video to realize non-contact (i.e., without sensors) and objective fault detection (an overview of the proposed system is shown in Figure 2). Note that we consider the usage as self-checking (e.g., in training) rather than during competition. This study explores the feasibility of non-contact fault detection through validation of the proposed system using intentional faulty walking videos. The system first estimates key points (joint locations) from walking videos by pose estimation. We used a pose estimation model whose performance was improved by the fine-tuning technique described below. Next, the input features of the fault detection model are calculated from the coordinate data estimated by the pose estimation. Finally, the feature vector is input to the classifier, which outputs the detection results. A smartphone camera is used to capture video to simplify the system, which can be used in a wide range of situations such as practice, competition, and judge's training. The effectiveness of the proposed system was verified by intentionally faulty walking videos of a university student race walker and a Tokyo Olympics medalist. We extend our previous short paper (Suzuki et al., 2022) by enhancing all sections including analysis of the reason for the fault detection and the differences in individual faulty walking movements, and evaluation of the validity of the data collection environment.

Our research questions are summarized as follows. (1) Is our non-contact fault detection system from a fixed smartphone camera video accurate? (2) What improvements in pose estimation accuracy from the video can be achieved through fine-tuning techniques? (3) What are the specific characteristics that differeniate the intentional faulty walking movements of a medalist from those of university walkers?

Materials and Methods

Experiment

Participants

In this study, the participants were one 25-year-old Tokyo Olympics medalist (Walker A) with 11 years of experience and a personal best of 38′57”37 in the 10,000 m walk and four university track and field athletes (21 ± 0.8 years old; Walkers B, C, D, and E) with 5.5 ± 0.5 years of experience and a personal best of 47′20”37 ± 1′19”55 in the 10,000 m walk. The participants were fully informed about the study and their consent was obtained in advance. All the experimental procedures were performed after obtaining prior approval for “Experiments on Human Subjects” at Nagoya University.

Procedures

The experiment was conducted at the Nagoya University track and field ground (under the conditions shown in Figure 1). The walking section was approximately 20 m long, and we captured the side profiles of the participants during the experiment. The participants walked repeatedly with breaks in between. The participants were randomly instructed to walk with a normal race walking form, an intentional BK form, and an intentional LC form. Since walker E was not able to perform intentional BK walking, he walked with the normal race walking form and LC. The experiment was conducted over multiple dates to ensure that fatigue did not affect the walk.

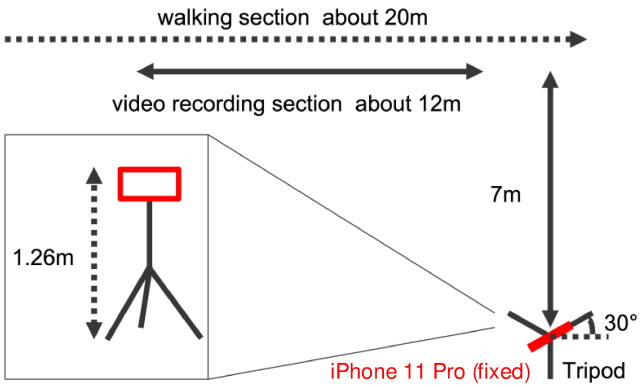


Figure 1. Experimental environment for walking video collection. We conducted the experiment on the track and field ground using the iPhone 11 Pro camera. The recording section was about 12 meters. The camera was always fixed at 30 degrees, as shown in the figure, to capture video.

Data Collection

Walking videos were captured by an iPhone 11 Pro camera at 60 fps. The resolution was 1080p HD. After capturing the video, we asked three judges (One person has an International Race Walking Judge qualification, and the other two have a Japan Race Walking Judge qualification) to judge whether faults were committed during the experiments by the participants. They

watched walking videos multiple times using their laptops. We annotated the videos according to the majority vote of the three judges. The judgement results were used as the correct labels when training the machine learning model. We captured 268 normal walking, 265 BK walking, and 268 LC walking. The number of videos and model input data by walker are shown in Table 1. The "videos" represents the videos collected during the experimental phase, and the number of data per walk type is based on the instructions given to each walker during the experiment. The "model input data" represents the input data to the machine learning model in which the videos are post-processed, and the number of data is based on the annotation results from the judges. Note that the number of data between the videos and the model input data is different due to the difference between the instructions in the experiment and the annotations from the judges. For example, 11 of the videos of Walker A walking with instructions to walk in BK were judged as Normal. In addition, some video data was excluded due to insufficient data length during post-processing.

Table 1. Number of valid data and videos (model input data / videos). There were differences in the number due to data elimination in the post-processing phase and differences between the instructions at the time of walking and the judges' judgements. For example, Walker A was instructed to perform 20 strides of normal, BK and LC walking; however, 11 of the BK strides were judged to be normal.

	Normal	BK	LC
Walker A	31 / 20	9 / 20	20 / 20
Walker B	60 / 63	94 / 84	56 / 63
Walker C	60 / 60	77 / 81	64 / 60
Walker D	40 / 61	69 / 80	55 / 60
Walker E	64 / 64	0 / 0	65 / 65

Proposed System

An overview of the proposed system is shown in Figure 2. The code and dataset are available at https://github.com/SZucchini/racewalk-aijudge.

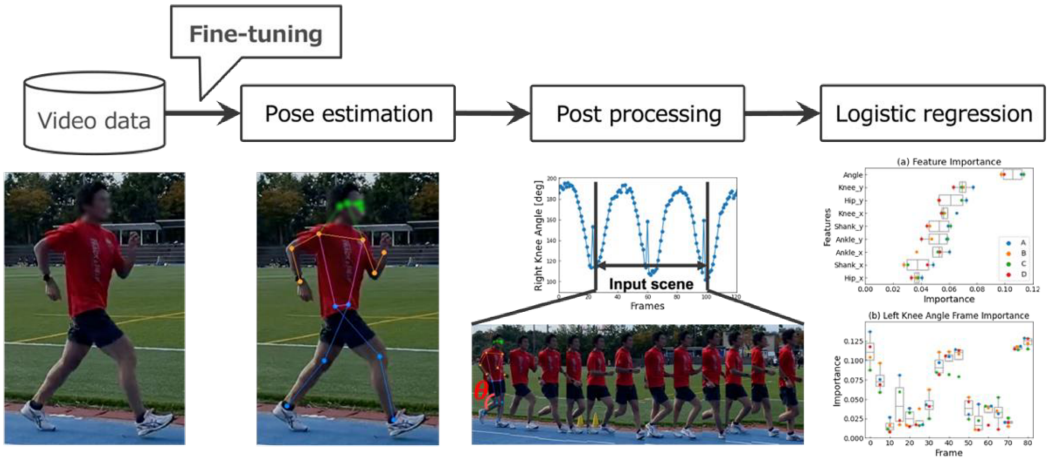


Figure 2. Overview of the proposed system. Our system extracts key points from videos using a fine-tuned pose estimation model. After post-processing the key points, a logistic regression model detects the faults.

Pose Estimation

There are two types of pose estimation models: a top-down model that estimates the key points separately after estimating the positions of all persons in the image, and a bottom-up model that estimates all key points in the image and then groups them for each person. In this study, we used HigherHRNet (Cheng et al., 2020), a bottom-up pose estimation model. The performance of the bottom-up model is superior to that of the top-down model in multiple-person estimation. Since the proposed system is meant to be applied to racing, the performance of multi-person estimation is important. Our system estimates 17 key points: nose, left and right eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. The input feature for the fault detection model is computed from pose estimation data, indicating that the performance of the pose estimation model affects the accuracy of the system. Therefore, we fine-tuned a pre-trained HigherHRNet model by race walking images to improve estimation performance.

Fine-tuning of the Pose Estimation Model

Fine-tuning is a learning method in which a model is transferred to another domain by updating the weights of a pre-trained model using other training data. In track and field pose estimation, fine-tuning has been used to improve the estimation performance of long jump and triple jump movements (Ludwig et al., 2021). We fine-tuned a pre-trained HigherHRNet based on common human pose images using race walking images to improve its estimation performance for walking movement. We adopted fine-tuning to reduce the cost of annotating images, aiming for a simpler system that can be used by a large number of people.

Post-processing of Pose Estimation Data

The key point coordinate data obtained by pose estimation must be corrected for the effects of differences in the walker's physique. Therefore, all data were normalized so that the length from the nose to the hip key point was 1. In addition, all points were translated so that the coordinates of the nose were set to $(0, 0)$.

Next, the key point coordinates of the shank and the knee angle were calculated from the key point coordinates of the hip, knee, and ankle. The shank position was defined as the midpoint of the knee and ankle, and the knee angle was defined as θ counterclockwise from the thigh to the calf. After the knee angle was calculated, outliers were identified based on the standard deviation of the frame-to-frame change in the right knee angle for all data, and the outlier data were removed.

After removing the outlier data, we extracted the interval of "two steps from the most bent right knee" from each data set to unify the walking scene of the input data. The definition of the knee angle and an example of a walking scene are shown in Figure 3. The number of frames for all walking data was normalized into 85 frames by interpolation to clarify the frames that contribute to the fault detection.

Figure 3. The definition of the knee angle and the waking scene that we extracted in post-processing.

Since we do not install the camera perpendicular to the walker's walking direction, the knee angle varies slightly depending on the position. However, we speculate that this may have less effect on the fault detection performance because we created input features that include differences in position. Figure 4 shows that maximum angle of the right knee is approximately 195° to 205°, becoming slightly smaller as the target moves to the right of the image. Therefore, we can say that the variation of the knee angle depending on the position is small and not random. This variation is also included in the input data for the machine learning model (because we used the angles for two step cycles extracted from the various positions), and the model can be trained to account for angle variation. For these reasons, if the input data for the judgement does not deviate significantly from the image position range of the training data, the judgement could not be affected.

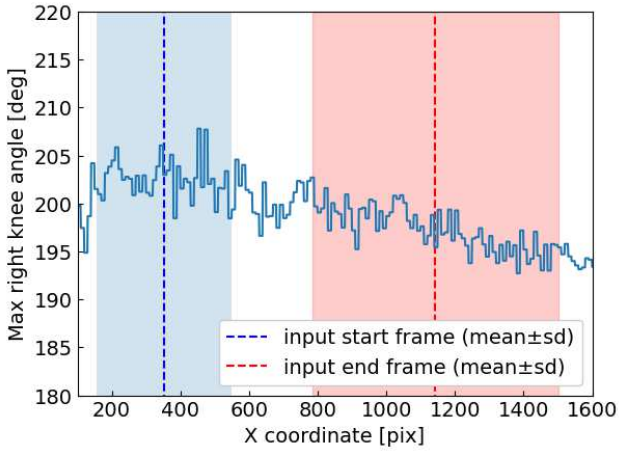


Figure 4. Variation of the maximum angle of the right knee angle. The x-coordinates of the walker's nose key point on the image coordinates were plotted on the horizontal axis and the maximum right knee angle at each position was plotted on the vertical axis. The mean of the start and end coordinates of the two steps walking extracted as input to the model is indicated by the dashed line, and the SD is indicated by the color-filled area.

Classifier

A logistic regression model was used as the classifier for fault detection. The logistic regression model is one of the simplest machine learning models, and it can analyze the detection criteria from feature importance based on the standard regression coefficients. Since it is undesirable not to understand the process of judging sports, we use the model that allows analysis of judging criteria. The detection model was created separately for two faults (BK and LC). The model takes the key points coordinates and knee angles as input, and outputs the results of the fault detection.

System Verification

Pose Estimation Model

Some walking images extracted from the collected videos were used to fine-tune the pose estimation model. The walking images were annotated by the COCO Annotator (Brooks, 2019). We used 108 images of walkers A, D, and E as training data, 25 images of walker B as validation data, and 52 images of walker C as test data. We used MMPose (Contributors, 2020), an open-source framework for pose estimation, for fine-tuning.

The average Precision (AP) and time-series changes in knee angle were used to evaluate the pose estimation models. Representative pose estimation and object detection models, including the

HigherHRNet used in this study, such as Openpose (Cao et al., 2019) and Detectron2 (Wu et al., 2019), use AP for evaluation. AP is calculated from object keypoint similarity (OKS), which represents the degree of proximity between the correct and estimated points at each key point. OKS is defined by the following equation:

$$OKS = \frac{\sum_i \exp(-d_i^2/2s^2k_i^2)\delta(v_i>0)}{\sum_i \delta(v_i>0)}, \quad (1)$$

where d_i is the Euclidean distance between the detected key point and the corresponding ground truth, v_i is the visibility flag of the ground truth, s is the object scale, and k_i is a per-keypoint constant that controls falloff.

In the AP calculation, a threshold value is defined, and the average precision of all key points is calculated, assuming that the OKS of key point i exceeds the threshold value as the correct answer. For general evaluation, the average value of the AP for each threshold value is used when the threshold value is changed in 10 steps of 0.05 from 0.50 to 0.95. The time-series variation in the knee angle allows for the identification of angle outliers. Since outliers are caused by incorrect key point estimation, the occurrence of incorrect estimation can be visually determined by checking the time-series variation of the knee angle.

Fault Detection Model

The fault detection model was evaluated by dividing the data by walker and cross-validating. In this method, the training walker data does not include the test data, thus the model performance for unknown walkers can be evaluated. In the training and evaluation, since we used all data except the test target walker, the number of data for training the model was sufficient as shown in Table 1. We used the accuracy and F-score to evaluate the model performance. The accuracy is the percentage of the estimated results that are correct. The F-score is expressed as F-score = (2 × Precision × Recall) / (Precision + Recall), where the Recall is equal to the true-positive rate, and the Precision is defined as the ratio of the sum of true positives and true negatives to false positives. The BK detection model was evaluated using data from walkers A to D. The LC detection model was evaluated using data from walkers A to E.

To determine why faults were detected, we used the standard regression coefficient for the logistic regression model as a measure of feature importance. In the analysis of feature importance by standard regression coefficients, the input features with larger absolute values of the coefficients contribute more to the detection results. We classified the input features into nine categories: features related to the x and y coordinates of the hip, knee, shank, and ankle, and features related to the knee angle. For the analysis of the detection reason, we used the average of the absolute values of the standard regression coefficients of the four (BK detection) or five (LC detection) models created for cross-validation.

Results

In this section, we first compared the performance of the pre-trained and fine-tuned pose estimation models. Second, we evaluated the performance of the fault detection model. In addition, we analyzed the feature importance of the model to clarify the reason for fault detection. Finally, we explained why the LC detection performance of Walker A was worse than the others by showing the difference in movement between the training and test data.

Pose Estimation Performance

First, we show the results of the pose estimation performance. AP increased by 0.031 from 0.961 to 0.992 before and after fine-tuning. For comparison between the pre-trained and fine-

tuned models, Figure 5 shows mean ± SD curves for correct and fault walking separately for the right knee angle calculated with each model. These results show that the result of fine-tuned model has less variability than the result of the pre-trained model, indicating that fine-tuning improves the performance of the pose estimation model.

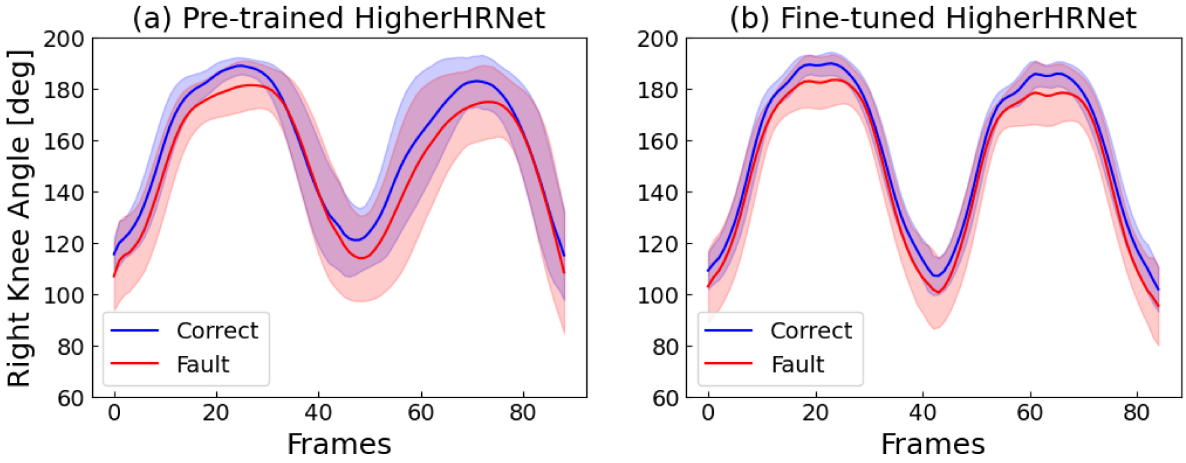


Figure 5. Comparison of mean ± SD curves in the right knee angle to all data. (a) Knee angles calculated from pre-trained HigherHRNet results; (b) knee angles calculated from fine-tuned HigherHRNet results.

We also show the time-series change in the right knee angle of the test data by the estimation model in Figure 6. In the fine-tuned model, there are no angle outliers due to mis-estimation of key positions, which existed in the pre-trained model result. This result indicates that accurate joint position estimation is achieved through fine-tuning.

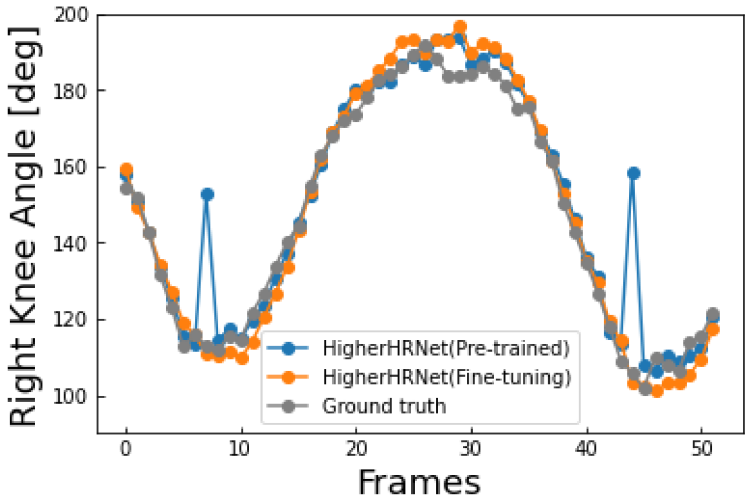


Figure 6. Comparison of time-series changes in right knee angle of the test data. The results from the fine-tuning model are closer to the ground truth.

Fault Detection Model Performance

Table 2 shows the results of the fault detection performance for the input data generated from the HigherHRNet (fine-tuned). For each input data (i.e., from each video), we determined whether fault or not. The average accuracy was over 90%. However, the detection of faults in the video of Walker A tended to be lower than that of the other walkers. The LC detection result for Walker A was worse than the BK result.

Table 2. Fault detection model evaluation for each walker and fault.

	Bent Knee		Loss of Contact	
	Accuracy	F-score	Accuracy	F-score
Walker A	0.850	0.893	0.700	0.727
Walker B	0.966	0.961	0.960	0.959
Walker C	0.915	0.905	0.967	0.967
Walker D	0.972	0.966	1.000	1.000
Walker E	-	-	0.969	0.969

Feature Importance Analysis

We show the feature importance of the BK detection model in Figure 7(a). This figure reveals that the knee angle is the most important feature in BK detection. Next, we analyzed the importance by frame of the input feature of knee angle. Figures 7(b) and 7(c) show the average importance for each of the five frames for the left and right knee angles. Frames 0–9, 35–49, and 75–84 had high importance for the left knee angle, while frames 15–24 and 55–69 had high importance for the right knee angle. From the above results, the BK detection of the model was considered to be based on the knee angle in specific frames. Figures 7(d) and 7(e) show the average time-series change in knee angle for all walkers during normal walking and BK walking. The frames with high feature importance indicate the pose from when the front foot touches the ground until it becomes vertical. In BK walking, the knee at this time is more bent than in normal walking. An example image of the pose is shown in Figure 8(a). Therefore, the BK detection of the model would follow the rules of race walking (e.g., the timing and the knee angle).

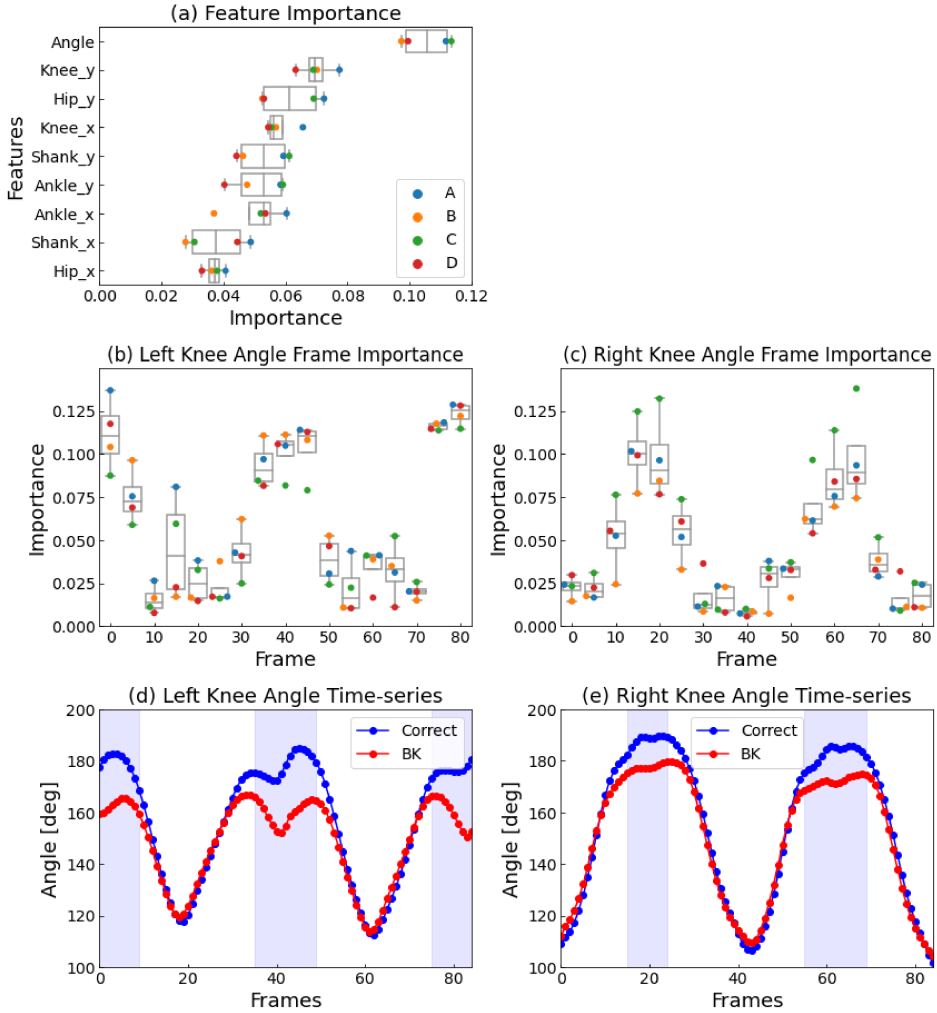


Figure 7. Feature importance analysis of BK. (a) Feature importance of the BK detection model. (b)–(c) Knee angle's frame importance of the BK detection model. (d)–(e) Comparison of time series changes in the knee angle between normal and BK walking (The frames with relatively high feature importance are shown in light blue).

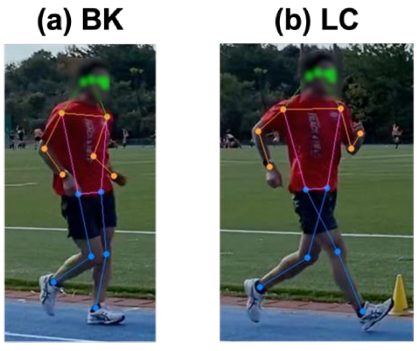


Figure 8. The pose examples of the important frame in (a) BK walking and (b) LC walking.

In LC detection, the importance of the input features is shown in Figure 9(a), which shows that the importance of the y-coordinate of each key point is high for LC detection. The importance of the knee angle is also relatively high, but this will be discussed later. Next, we analyzed the feature importance of the knee key point's y-coordinate by frame, which is one of the most important features. Figures 9(b) and 9(c) show the average importance for each of the five frames

for the left and right knee key point's y-coordinate. There is no significant difference in importance between frames for the left side, while frames 45–49 and 70–74 are more important for the right side.

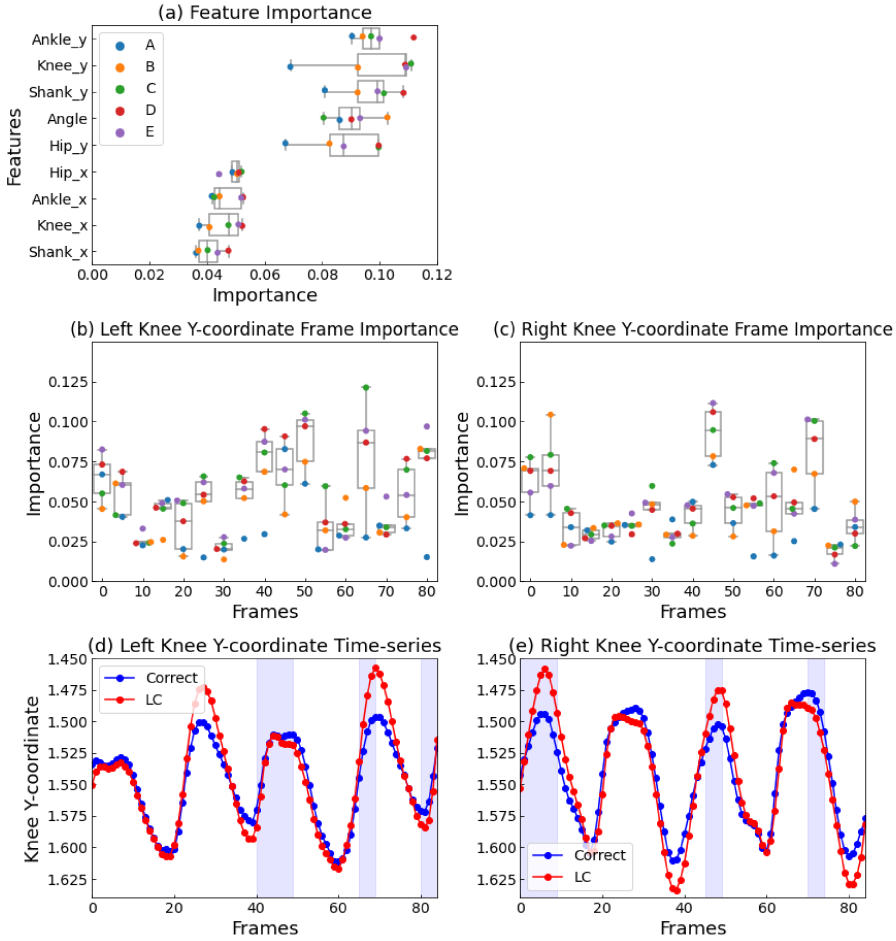


Figure 9. Feature importance analysis of LC. (a) Feature importance of the LC detection model. (b)–(c) Knee Y-coordinate's frame importance of the LC detection model. (d)–(e) Comparison of time series changes in the knee Y-coordinate between normal and LC walking (The frames with relatively high feature importance are shown in light blue).

From the above results, we considered that the LC detection model focuses on the y-coordinate of each key point to detect the LC. Figure 9(e) shows the average time-series change in the y-coordinate of the right knee for all walkers during normal walking and LC walking. In the important frames, especially in frames 45–49, the knee position is higher in LC walking than in normal walking. The left knee y-coordinate shown in Figure 9(d) also differed between normal and LC walking. In the relevant frames, the walker is swinging his feet forward, a situation where both feet may be off the ground. An example image of the pose is shown in Figure 8(b). Therefore, the LC detection of the model would follow the rules of race walking (e.g., the timing and the height of each joint).

Differences in Movement during LC Walking

In Figure 9, the knee angle, which according to the rules should not be the focus of attention during LC detection, became more important. Moreover, the model could not detect LC well in the movements of Walker A. We determined the cause of these results by comparing the movements of Walker A and other walkers. First, to clarify the features of the detection model of Walker A (an Olympian), we compared the feature importance of the detection model of

Walker A (the model trained on data other than Walker A's) with the feature importance of the detection models of the other walkers. Figure 10(a) shows that the feature importance of the knee angle was the second highest in the detection model for Walker A. This result suggests that there is a difference in knee angle between normal and LC walking in the training data of the detection model of Walker A (data of walkers other than Walker A), whereas there is no difference in the data of Walker A.

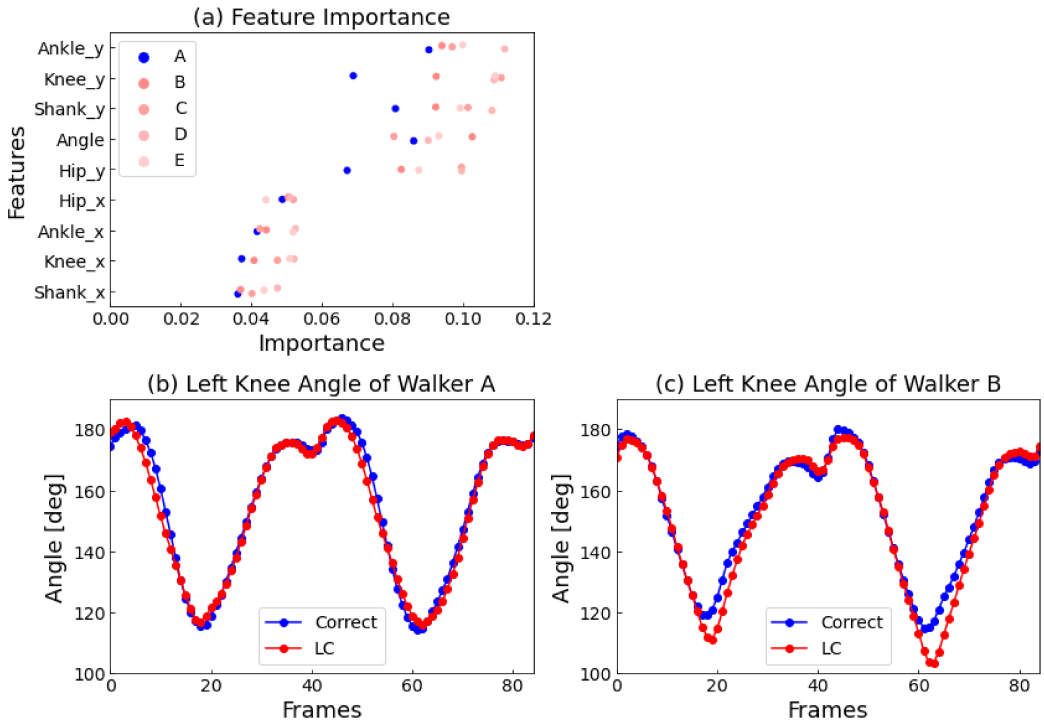


Figure 10. Comparison of Walker A and other walkers in LC. (a) Feature importance comparison of the detection models for Walker A and other walkers (we removed boxplots because of the comparison). (b)–(c)Left knee angle of Walker A and Walker B.

To clarify a difference in the knee angle between Walker A and the other walkers, we analyzed the average time-series change in the left knee angle of Walkers A and B. Figures 10(b) and (c) show that there was almost no difference in the knee angle between normal and LC walking for Walker A, but there was a difference between normal and LC walking for Walker B. Walker B lifted his hind leg during LC walking, and as a result, his knee was bent more than in normal walking. Walkers C through E showed the same tendency as Walker B. These results indicate that Walkers B to E tended to lift their hind legs when they intentionally walked in LC walking, while Walker A did not. This difference may explain why the detection model of Walker A could not use the features that were focused on during the learning process, and the detection of faults in the movements of Walker A was poor. Differences in the ability to control walking movements based on athletic experience and athletic ability might affect the differences in leg lifting movements during intentional faulty walking.

Discussion

In this study, a system for detecting faults from walking videos was constructed to realize non-contact fault judgement in race walking. Here, we discuss the effectiveness of fine-tuning the pose estimation model. We also discuss that the movement differences between the training and test data may cause failure of fault detection.

Typical pose estimation models are trained on large amounts of data, such as the COCO dataset (Lin et al., 2014). However, such datasets are mainly composed of daily activity movements, and trained models cannot accurately estimate sports movements. In this study, we were able to improve the performance of a pre-trained model by fine-tuning it with approximately 100 race walking images. The training was successful even with a small amount of data because the race walking movement consists of repetitions of specific periodic movements. In this sense, this method may also be effective for training a high-performance pose estimation model with low annotation cost for running and sprinting movements.

In detecting the faults in the movements of Walker A, the features in the training data did not match the rules, resulting in poor performance. In this case, even if both feet were off the ground, the model could fail to detect LC when walkers do not lift their legs. In other words, the model may have selected features that do not match the rules. We also need to take care to avoid bias in training data when building a fault detection model using machine learning. In addition, additional validation with large-scale data is necessary to clarify what kind of data is needed to create a more general detection model and accurate detection criteria.

Conclusion

The fault detection system proposed in this study achieved highly accurate non-contact fault detection under the constraint of using intentional faulty walking video data. This accuracy is insufficient for use in competition but may be sufficient for use in training self-checks. We also showed that the reason for the detection of a fault can be analyzed by feature importance analysis of the machine learning model. If a system is created by collecting faulty walking videos during a race, a judgement assistance system for a single walker could be relatively easy to implement. On the other hand, applying the system to a race is more difficult because it is necessary to improve the performance of pose estimation for multiple walkers and to solve the problem of occlusion between walkers. To realize a system that can be used in races, the following issues must be addressed: (1) to collect as many natural faulty walking videos as possible, (2) to verify the versatility of the system using the collected videos, and (3) to achieve high-performance pose estimation for multiple persons.

Acknowledgement

This work was financially supported by JSPS Grant Number 20H04075, JST PRESTO Grant Number JPMJPR20CA, and JST SPRING, Grant Number JPMJSP2125. The author (ND) would like to take this opportunity to thank the “Interdisciplinary Frontier Next-Generation Researcher Program of the Tokai Higher Education and Research System.”

References

World Athletics (2023), C1.1 & C2.1 – Competition Rules & Technical Rules. Retrieved from https://www.worldathletics.org/about-iaaf/documents/book-of-rules (Accessed: 2023/09/04).

Brooks, J. (2019). COCO Annotator. https://github.com/jsbroks/coco-annotator/.

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T. S., & Zhang, L. (2020). HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation. In *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5386–5395).

Contributors, M. (2020). Openmmlab pose estimation toolbox and benchmark. https://github.com/open-mmlab/mmpose.

Díaz-Pereira, M. P., Gomez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic recognition and scoring of olympic rhythmic gymnastic movements. Human movement science, 34 , 63–80.

Di Gironimo, G., Caporaso, T., Amodeo, G., Del Giudice, D. M., Lanzotti, A., & Odenwald, S. (2016). Outdoor tests for the validation of an inertial system able to detect illegal steps in race-walking. Procedia engineering, 147 , 544–549.

Gomez-Ezeiza, J., Torres-Unda, J., Tam, N., Irazusta, J., Granados, C., & Santos-Concejero, J. (2018). Race walking gait and its influence on race walking economy in world-class race walkers. Journal of sports sciences, 36 (19), 2235–2241.

Hanley, B., Tucker, C. B., & Bissas, A. (2019). Assessment of iaaf racewalk judges' ability to detect legal and non-legal technique. Frontiers in sports and active living, 9.

Hoga-Miura, K., Hirokawa, R., & Sugita, M. (2017). Reconstruction of walking motion without flight phase by using computer simulation on the world elite 20 km race walkers during official races. Slovak Journal of Sport Science, 2 (1), 59–75.

Knicker, A., & Loch, M. (1990). Race walking technique and judging-the final report to the international athletic foundation research project. New Studies in Athletics, 5 (3), 25–38.

Lee, J. B., Mellifont, R. B., Burkett, B. J., & James, D. A. (2013). Detection of illegal race walking: a tool to assist coaching and judging. Sensors, 13 (12), 16065–16074.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740–755).

Ludwig, K., Scherer, S., Einfalt, M., & Lienhart, R. (2021). Self-supervised learning for human pose estimation in sports. In 2021 ieee international conference on multimedia & expo workshops (icmew) (pp. 1–6).

Menting, S. G. P., Hanley, B., Elferink-Gemser, M. T., & Hettinga, F. J. (2022). Pacing behaviour of middle-long distance running & race-walking athletes at the iaaf u18 and u20 world championship finals. European Journal of Sport Science, 22 (6), 780–789.

Pavei, G., & La Torre, A. (2016). The effects of speed and performance level on race walking kinematics. Sport Sciences for Health, 12 (1), 35–47.

Santoso, D. R., & Setyanto, T. A. (2013). Development of precession instrumentation system for differentiate walking from running in race walking by using piezoelectric sensor. Sensors & Transducers, 155 (8), 120.

Suzuki, T., Takeda, K., & Fujii, K. (2022). Automatic fault detection in race walking from a smartphone camera via fine-tuning pose estimation. In 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE) (pp. 631-632).

Taborri, J., Palermo, E., & Rossi, S. (2019). Automatic detection of faults in race walking: A comparative analysis of machine-learning algorithms fed with inertial sensor data. Sensors, 19 (6), 1461.

Uchida, I., Scott, A., Shishido, H., & Kameda, Y. (2021). Automated offside detection by spatio-temporal analysis of football videos. In Proceedings of the 4th international workshop on multimedia content analysis in sports (pp. 17–24).

Vapnik, V. (1999). The nature of statistical learning theory. In (p. 138). Springer science & business media.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. https://github.com/facebookresearch/detectron2.

Xu, C., Fu, Y., Zhang, B., Chen, Z., Jiang, Y.-G., & Xue, X. (2019). Learning to score figure skating sport videos. IEEE transactions on circuits and systems for video technology, 30 (12), 4578–4590.