Real-time Pilates Posture Recognition System Using Deep Learning Model

Hayoung Kim, Kyeong Teak Oh, Jaesuk Kim, Oyun Kwon, Junhwan Kwon, Jiwon Choi, and Sun K. Yoo[(✉)]

Department of Medical Engineering, Yonsei University, Seoul, South Korea
sunkyoo@yuhs.ac

Abstract. As the pandemic situation continues, many people exercise at home. Mat Pilates is a popular workout and effective core strengthening. Although many researchers have conducted pose recognition studies for exercise posture correction, the study on Pilates exercise is only one case on static images. Therefore, for the purpose of exercise monitoring, we propose a real-time Pilates posture recognition system on a smartphone for exercise monitoring. We aimed to recognize 8 Pilates exercises—Bridge, Head roll-up, Hundred, Roll-up, Teaser, Plank, Thigh stretch, and Swan. First, the Blazepose model is used to extract body joint features. Then, we designed a deep neural network model that recognizes Pilates based on the extracted body features. It also measures the number of workouts, duration, and similarity to experts in video sequences. The precision, recall, and f1-score of the posture recognition model are 0.90, 0.87, and 0.88, respectively. The introduced application is expected to be used for exercise management at home.

Keywords: Pilates · Pose recognition · Pose correction · Deep neural network · Real-time · Smartphone

1 Introduction

The interest in digital home training programs has increased since the outbreak of COVID-19 as they are accessible and cost-effective to exercise [1]. Among the home exercises, Pilates is a popular exercise that has recently become increasingly widespread in rehabilitation therapy [2]. Pilates was developed by Joseph Pilates of Germany and is a full-body exercise made for rehabilitating patients during World War I [3]. The Pilates mat exercise program is effective in treating chronic low back pain by evaluating core muscle thickness [4]. In addition, Pilates exercise without using any equipment is useful for improving respiratory functions and disease-related symptoms [5]. However, an injury may occur by repeating the wrong motion without experts. Therefore, a Pilates posture recognition system for home training is required.

In computer vision, exercise training systems through human posture estimation are being developed to increase exercise effectiveness and prevent injury in various sports fields [6]. Dittakavi, B. [7] conducted a study to classify postures from fixed images using

probability techniques for Yoga, Pilates, and Kung Fu exercises and to explain which joint motion was incorrect. Wu, Y. [8] classified 45 yoga movements and created a model that gave points to the exercise movements without an expert. They extracted image features using a convolutional neural network (CNN) and trained the model using a unique contrastive loss that combines the L2 norm with cosine similarity. Li, J. [9] presented a model for classifying and evaluating 117 yoga movements. Yoga movements from 22 subjects were measured with an RGB-D camera. At this time, 3D coordinates were corrected by taking pictures from both the front and side. In addition, a new Cascade 2S-AGGN (Cascade graph convolutional neural network for yoga pose classification and assessment) model was presented by constructing a graph convolutional neural network (GCN) in a hierarchical structure. Zhao, Z [10] also used GCN to classify three types of motion and presented a model for correcting posture. The expert's motion and the subject's motion were trained separately and evaluated the subject's motion. Dynamic time warping is used to compensate for the difference in each person's movement speed. In previous studies, there is one case of Pilates exercise [7] correction of static images. However, Pilates pose correction studies on video sequences are necessary since Pilates involves segmental movements of the spine. Therefore, we studied Pilates exercise correction on video sequences.

The purpose of this study is to develop a real-time Pilates mat exercise recognition system on a smartphone for exercise management at home. We developed a Pilates exercise posture classification model that automatically recognizes the 8 Pilates exercises. We even added a parameter measurement function for exercise monitoring. We finally developed a real-time posture recognition system on the smartphone for user convenience.

"Methods" describe a data set of Pilates exercises, a posture recognition method, and real-time exercise monitoring capabilities. In the "Results", the results of the Pilates posture recognition model and the real-time monitoring system are mentioned. In the "Discussion", the results are discussed. In "Conclusion", the conclusion is described.

2   Methods

2.1   Data Collection

We have selected 8 Pilates exercises as follows: Bridge, Head roll-up, Hundred, Roll-up, Teaser, Thigh stretch, Plank, and Swan. The example of the selected Pilates exercises is shown in Fig. 1.

The Pilates data were acquired with the camera facing the exercise mat and at a distance of 2.5m. We used the front camera of the Galaxy 22 smartphone. The 8 movements are repeated about 5 times and recorded as videos. The image acquisition sampling rate is 30 fps. A total of 15 subjects of videos were acquired, and the sex ratio was 6 males and 9 females.

We extracted features of 33 joints from the acquired videos using the Blazepose model [11] to train and test the Pilates recognition model. In many posture recognition studies, Blazepose was used because of its high accuracy and performance [12–14]. The 33 joint features are the location of the x, y, and z coordinates and the visibility indicating whether the joint is visible. 9 target postures were manually labeled for each

image of the video sequences. It includes 8 Pilates postures and an "unknown" class to distinguish it from the prescribed exercises. The entire movement from the initial to the final position is trained, and the information of one frame is input into the model. The 15 subjects' videos are split into 11 and 4 for training and test data. The total number of Pilates classification data is 247,203. The training data is 179,446, and the test data is 67,757. The number of samples for each 8 Pilates postures is shown in Table 1.
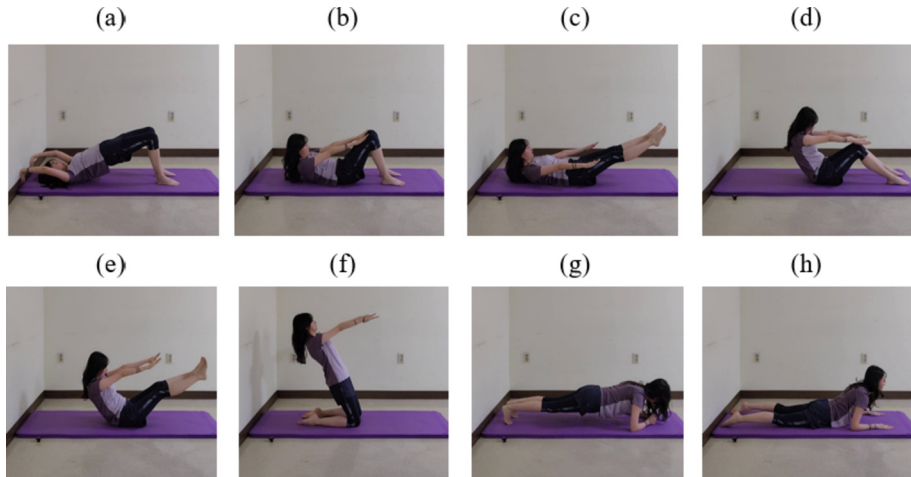

Fig. 1. Example of Pilates poses. (a) Bridge, (b) Head roll up, (c) Hundred, (d) Roll up, (e) Teaser, (f) Thigh stretch, (g) Plank, (h) Swan.

Table 1. The number of Pilates postures recognition samples.

Index	Class	Training samples	Test samples
0	Unknown	87,814	34,251
1	Bridge	11,083	3,460
2	Head roll up	7,596	2,915
3	Hundred	18,763	7,533
4	Roll up	11,566	4,392
5	Teaser	7,574	3,299
6	Thigh stretch	11,364	3,472
7	Plank	12,206	3,826
8	Swan	11,480	4,609

2.2   Pose Recognition Model

The Pilates posture recognition model is designed to automatical classify 8 Pilates and unknown postures. We designed a simple deep neural network to deploy on an Android phone. The input size of the model is 1×132 because the x, y, z, and visibility values of 33 joint points from on frame are entered. Three fully connected layers with 128, 64, and 16 neurons are stacked. A dropout rate of 0.2 was used between each fully connected layer to prevent overfitting. Finally, a soft-max layer for classifying multiple classes was connected. The output of the model is probabilities of 9 poses.

2.3   Real-Time Exercise Monitoring System

We propose a Pilates posture recognition system to operate in real-time on the smartphone. The sequence was designed to extract the number of exercises, exercise duration, and similarity with expert posture to monitor Pilates exercise. To operate the system on the smartphone, Mediapipe [15] framework was used.

Real-Time System Architecture. The architecture of the posture recognition system is shown in Fig. 2. First, in the pose estimation section, the features of one person's 33 joints are estimated by the Blazepose model when the input image comes through the camera. In the next pose recognition section, the pose recognition model classifies 8 Pilates exercises using the pose features. After then, since there is flickering in which the model predicts a different posture while predicting posture, a moving average filter is added that uses the average value of the last 10. Then, the pose counter & time measure section counts the number of exercises and measures the exercise duration. In the pose correction section, the similarity of the expert and current posture is calculated using joint features and recognition results. Finally, these exercise measurement parameters are combined in the draw overlay image section to generate an output image, and the result is displayed on the screen.
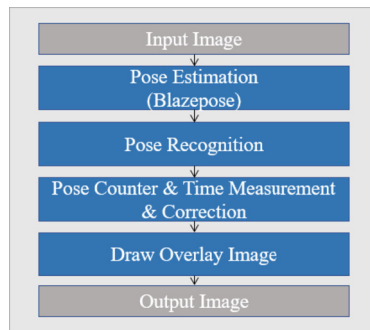

Fig. 2. Architecture of the Real-time Pose Monitoring System

Recognition of the Up-and-down Movement. The up-and-down for one posture was recognized using the Euclidean distance differential between two specific joints. It is used

for the following function, posture counting and correction. In the Bridge pose, the hips move away from the ankles as they lift up and come closer as they lower down, recognized as Bridge-up by ankle-to-hip distance increase. Similarly, Head roll-up, Roll-up, and Teaser are identified by ear-to-knee distance decrease. Thigh stretch-up is recognized by hip-to-foot index distance increase. Plank and Swan up-and-down movements are detected using hip-to-elbow and ear-to-elbow distance differentials, respectively.

Exercise Count and Time Measurement. Exercise parameters for Pilates postures are shown on the screen for users to monitor exercise counts and duration. The number of exercises and duration were determined through pose recognition results and recognition of up-and-down posture. The Hundred Posture, a sustained position, was counted independently without distance comparison. For other postures, counts were measured when transitioning from up to down or vice versa. Every exercise time was measured by starting when each exercise was recognized and ending when the motion was not recognized.

Pose Correction. Joint angles were used as feedback parameters for Pilates postures. To enhance workout effectiveness, it is needed to pay attention to correct posture. The similarity between the current and the expert's posture was assessed by comparing angles. The method for calculating the angle of each joint is as follows. As shown in Fig. 3 (a), when there are three joint points X_1, X_2, and X_3, the radian angle was obtained by Eq. (1). And radian angles were converted to degree units.

$$\theta = cos^{-1}\frac{x_1 - x_2}{y_1 - y_2} - cos^{-1}\frac{x_3 - x_2}{y_3 - y_2}. \tag{1}$$

Angles were compared to reference angles using 4 specific pairs, as depicted in Fig. 3 (b). These pairs include the shoulder angle between the ears and hip, hip angle between shoulder and knee, knee angle between hip and ankle, and ankle angle between knee and foot index. Pilates is primarily a core strengthening exercise, and the arms play an auxiliary role in balancing, so the joint angles around the torso were selected.

A weighted joint angle difference method was used to compare postures with experts. The expert posture that the reference for posture comparison was acquired from YouTube videos. One cycle of exercise motions to be corrected is each of the previously recognized up posture and down posture. The one-cycle of exercise motion was compared with the expert by calculating each section's average joint angle difference divided into 10 sections. Then the angle difference was multiplied by each joint's confidence score. Calculate the weighted joint angle difference for all 10 cycles and 8 joints as Eq. (2). Joints with high confidence scores were weighted to have more influence on the similarity measurement, the idea of the weighted distance method [16]. Finally, the angle difference was converted to a similarity score by normalizing 180 degrees as Eq. (3).

$$A_{diff} = \frac{1}{P} \cdot \frac{1}{J} \sum_{i=1}^{P} \sum_{j=1}^{J} V_{i,j}\left(A_{i,j}^s - A_{i,j}^r\right), i \in (1, 10), j \in (1, 8). \tag{2}$$

$$Score = \frac{180 - A_{diff}}{180}. \tag{3}$$

where A_{diff} denotes a weighted joint angle difference. The i and j denote period and joint index, respectively. $V_{i,j}$ is average visibility of the i-th period and j-th joint. $A_{i,j}^s$ denotes

subject's average angle of the i-th period and j-th joint. $A^r_{i,j}$ denotes reference average angle of the i-th period and j-th joint.
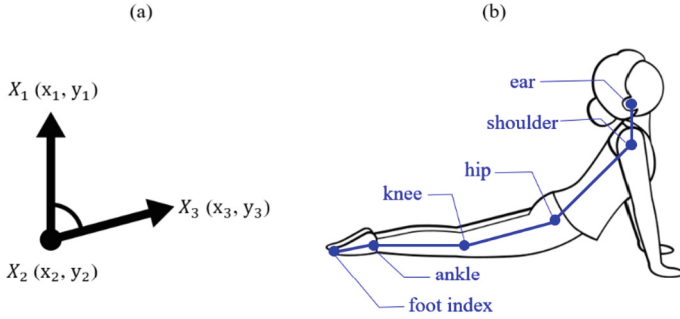

Fig. 3. Pose Angle. (a) Joint angle calculation using two coordinates, (b) main angle used for posture comparison.

3   Results

3.1   Result of Pose Recognition Model

We conducted an experiment with 67,757 test samples from 4 subjects. Performance metrics for evaluating posture classification models used Precision, Recall, and F1-score [17], which are widely used in the performance evaluation of classification models. The formulas are (4), (5), and (6), respectively.

$$Precision = \frac{TP}{TP + FP}. \tag{4}$$

$$Recall = \frac{TP}{TP + FN}. \tag{5}$$

$$F1 - score = 2 \cdot \frac{recall \cdot precision}{recall + precision}. \tag{6}$$

The precision, recall, and f1-score of the recognition model are 0.90, 0.87, and 0.84, respectively. The results of three metrics for 9 classes are shown in Table 2.

3.2   Results of Real-Time Exercise Monitoring System

Results of Exercise Monitoring System on Test Data.   The exercise monitoring system was tested with videos of 4 subjects on the desktop CPU. These videos are the same data used to test the pose recognition model. Table 3 shows the exercise counts for each subject in the 4 test videos. The last image of the Pilates monitoring system for each subject video is shown in Fig. 4. In Fig. 4, the first line indicates the current posture's

Table 2. Result of Pilates postures recognition model.

Index	Class	Precision	Recall	F1-score
0	Unknown	0.93	0.92	0.93
1	Bridge	0.97	0.88	0.92
2	Head roll-up	0.86	0.91	0.88
3	Hundred	0.95	0.97	0.96
4	Roll-up	0.92	0.84	0.88
5	Teaser	0.86	0.75	0.80
6	Thigh stretch	0.92	0.93	0.92
7	Plank	0.83	0.99	0.90
8	Swan	0.87	0.96	0.92

count, duration, and the previous posture's similarity score. Every 8 exercise's count, duration, and similarity score were also displayed on the screen. Figure 5 shows an example of similarity score comparison in the Roll-up posture. In the Roll-up posture, subject 2's score, who stood with his legs stable on the floor (Fig. 5 (a)), was 0.81 (Fig. 5 (b)), but subject 3, who came up using the recoil of bending his knees (Fig. 5 (c)), scored 0.77 (Fig. 5 (d)).

Table 3. The number of Pilates exercises for test data.

Pilates Exercise	Subject 1	Subject 2	Subject 3	Subject 4
Bridge	5	5	5	7
Head roll-up	5	5	6	8
Hundred	2	2	2	2
Roll-up	6	5	5	6
Teaser	5	5	5	6
Thigh stretch	6	6	6	6
Plank	27	15	17	30
Swan	5	5	7	8

Results of Real-time system on a smartphone. We checked the operation of the real-time Pilates exercise monitoring system on the Galaxy 22 smartphone. Model inference ran on the GPU, and the rest of the calculations ran on the CPU. A total of 8 exercises were performed twice, and the screen of the app was recorded. Table 5 shows the number of each Pilates exercises performed for test 1 and test 2. The last image of the application for each test is shown in Fig. 6. The number of exercises, time, and similarity score for
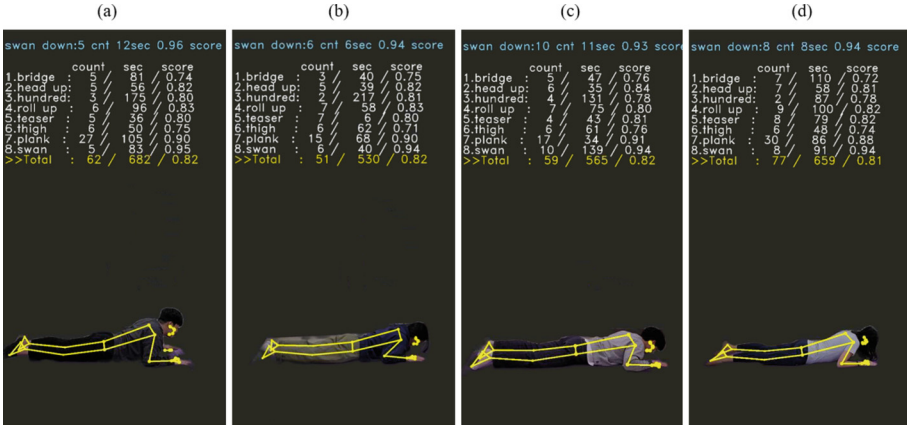
Fig. 4. The result of the Pilates posture monitoring system's count, time, and similarity score on test data. (a) subject 1, (b) subject 2, (c) subject 3, (d) subject 4.
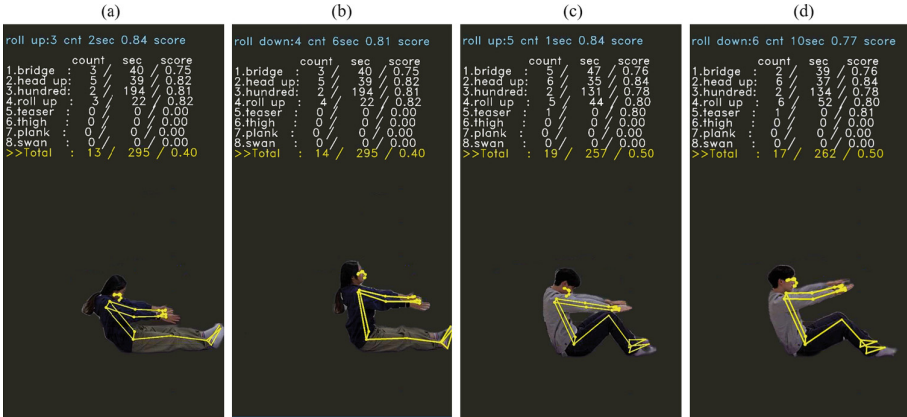

Fig. 5. Example of Roll-down posture score comparison. (a) Subject 2's roll-up posture. The score is the previous roll-down posture, (b) Subject 2's roll-down posture. The score is the previous roll-up posture, (c) Subject 3's roll-up posture. The score is the previous roll-down posture, (d) Subject 3's roll-down posture. The score is the previous roll-up posture.

each motion are displayed. Figure 7 shows examples of scores and count results for the Teaser and Plank postures in the two tests (Table 4).

Table 4. The repetition of Pilates exercises on the test data.

Pilates Exercise	Test 1	Test 2
Bridge	5	5
Head roll-up	5	5
Hundred	2	2
Roll-up	5	5
Teaser	5	6
Thigh stretch	5	5
Plank	10	10
Swan	5	5
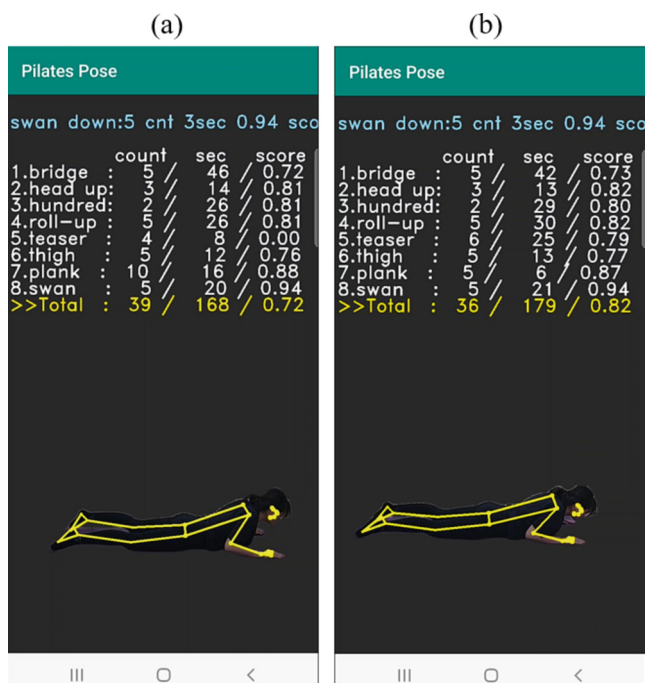
Fig. 6. The result of the real-time Pilates exercise monitoring application on a smartphone. (a) test 1, (b) test 2.

4 Discussion

We conducted a pose recognition and correction on video sequences for home Pilates exercise monitoring. Since there is no open data set for Pilates exercise, datasets were newly acquired. 8 Pilates exercises were selected as easy for beginners to follow and good for back pain prevention, abdominal exercise, and stretching.
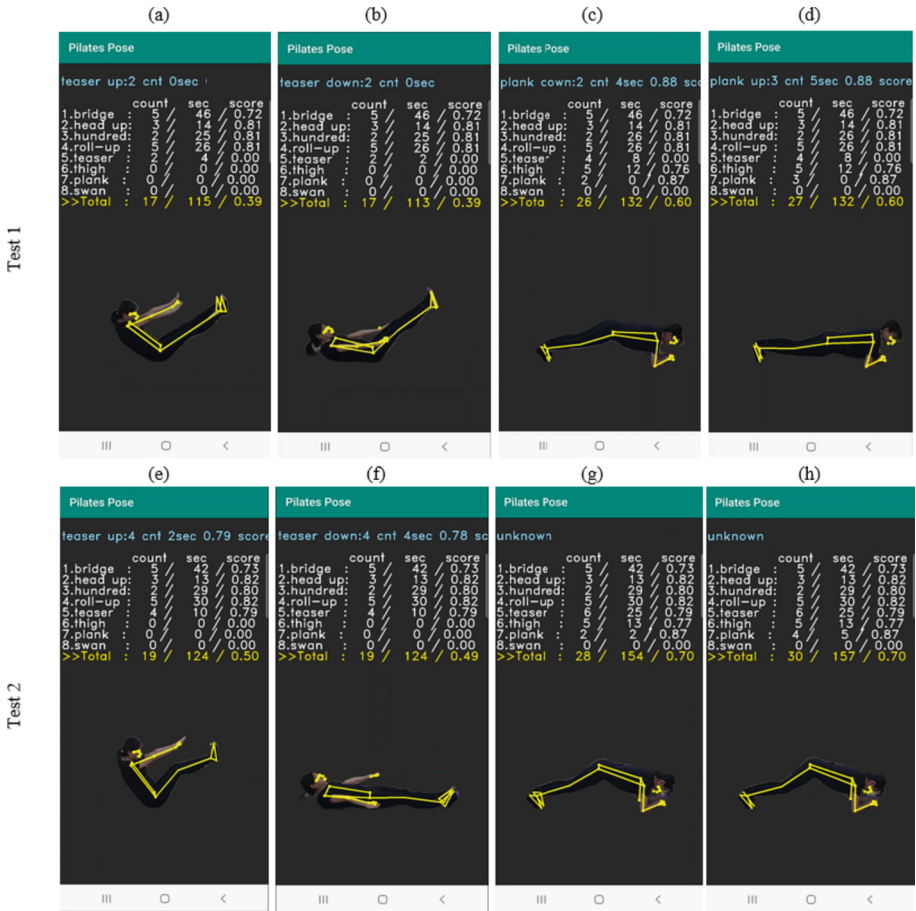
Fig. 7. Example results of Teaser and Plank in Tests 1 and 2 of the real-time Pilates exercise application. (a) Teaser up of test 1, (b) Teaser down of test 1, (c) Plank down of test 1, (d) Plank up of test 2, (e) Teaser up of test 2, (f) Teaser down of test 2, (g) Plank down of test 2, (h) Plank up of test 2.

Using the Blazepose model and a simple neural network, it showed recognition of 8 Pilates exercises. Most errors in the pose recognition models occurred between the 8 target exercises and the unknown class. There were also recognition errors in lying postures such as Hundred, Roll-up, Teaser, and prostrating postures between Swan and Plank since the entire movement was trained. Therefore, a moving average filter was used to prevent the flickering of class prediction in the middle of the video sequences. The number of exercises, exercise duration, and similarity score with experts were calculated using the pose recognition model's prediction. Most count errors in the test videos are due to recognition errors in rollups and teasers (Fig. 4). We confirmed that all exercises counted well except head up, teaser, and plank on the smartphone. In both test1 and test2, the 2 Head-ups were not counted because the head was not raised enough in the

head-up posture (Fig. 6). In test1, 1 Teaser was not counted (Fig. 6 (a)) because the subject fell so quickly as not recognized up-down (Fig. 7 (a), (b)). The remaining 4 Teasers are counted, but the score is 0 because they did not exercise enough to compare the movements (Fig. 6 (a)). However, in test 2, all the Teasers are counted and scored as enough exercise (Fig. 7 (e), (f)). Meanwhile, In Test 1, all the Planks were counted (Fig. 6 (a)) as the postures were well recognized (Fig. 7 (c), (d)). In test 2, 5 Planks were not counted (Fig. 6 (b)) because they were recognized as unknown due to the wrong motion of lifting the hip (Fig. 7 (g), (h)). With the development of our proposed exercise monitoring app, we expect that users will be able to receive Pilates exercise feedback regardless of location [16].

Limitations and Future Work. Although there were some technical limitations, we have plans for improvement. Since the data was acquired on only direction, there is a limitation to posture recognition. It is necessary to acquire data from various angles. The errors in posture recognition not only affect count and time measurements but also pose correction. Therefore, our future plans involve developing recognition models using time series data to enhance performance. Additionally, we aim to implement Devanne, M. [18]'s method to provide detailed feedback on specific body parts and overall posture.

5   Conclusion

In this paper, we propose a Pilates exercises monitoring system on a smartphone in real-time. First, we acquired video sequences for 8 Pilates postures. Pilates postures were recognized using a deep learning model on video sequences. In addition, exercise count and time measurement function for measuring exercise volume were added. We also proposed a weighted joint angle distance method that measures the angle and movement of major joints and compares posture with experts. It is expected that Pilates exercises can be corrected at home without experts.

References

1. Wilke, J., et al.: Restrictercise! preferences regarding digital home training programs during confinements associated with the COVID-19 pandemic. Int. J. Environ. Res. Public Health **17**(18), 6515 (2020)
2. La Touche, R., Escalante, K., Linares, M.T.: Treating non-specific chronic low back pain through the pilates method. J. Bodyw. Mov. Ther. **12**(4), 364–370 (2008)
3. Latey, P.: The pilates method: history and philosophy. J. Bodyw. Mov. Ther. **5**(4), 275–282 (2001)

4. Batıbay, S., Külcü, D.G., Kaleoğlu, Ö., Mesci, N.: Effect of pilates mat exercise and home exercise programs on pain, functional level, and core muscle thickness in women with chronic low back pain. J. Orthop. Sci. **26**(6), 979–985 (2021)

5. Bağlan Yentür, S., et al.: The effects of Pilates training on respiratory muscle strength in patients with ankylosing spondylitis. Physiotherapy Theory and Practice, pp. 1–11 (2022)

6. Chen, H.T., He, Y.Z., Hsu, C.C.: Computer-assisted yoga training system. Multimedia Tools and Appl. **77**(18), 23969–23991 (2018)

7. Dittakavi, B., et al.: Pose tutor: an explainable system for pose correction in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3540–3549 (2022)

8. Wu, Y., et al.: A computer vision-based yoga pose grading approach using contrastive skeleton feature representations. In: Healthcare, **10**(1), p. 36. MDPI (2021)

9. Li, J., Hu, H., Li, J., Zhao, X.: 3D-Yoga: a 3D yoga dataset for visual-based hierarchical sports action analysis. In: Proceedings of the Asian Conference on Computer Vision, pp. 434–450 (2022)

10. Zhao, Z., et al.: 3D pose based feedback for physical exercises. In: Proceedings of the Asian Conference on Computer Vision, pp. 1316–1332 (2022)

11. Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., Grundmann, M.: Blazepose: On-Device Real-Time Body Pose Tracking. arXiv preprint arXiv:2006.10204 (2020)

12. Abolmaali, S., Hafeez, M.M., Mohsin, T., Hassan, M.I., Thakur, A.: Pill ingestion action recognition using mediapipe holistic to monitor elderly patients. International Supply Chain Technology J. **7**(11) (2021)

13. Ohri, A., Agrawal, S., Chaudhary, G.S.: On-device realtime pose estimation & correction. International Journal of Advances in Engineering and Management (IJAEM) (2021)

14. Zhang, Y.: Applications of Google MediaPipe Pose Estimation Using a Single Camera (2022)

15. Lugaresi, C., et al.: Mediapipe: A Framework for Building Perception Pipelines. arXiv preprint arXiv:1906.08172 (2019)

16. Popovici, A.: Analysing Real-Time Vision-Based Pose Estimation Algorithms for RITH Purposes on Embedded Devices (Bachelor's thesis, University of Twente) (2022)

17. Goutte, C., Gaussier, E.: A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21–23, 2005. Proceedings 27, pp. 345–359. Springer Berlin Heidelberg (2005)

18. Devanne, M.: Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation. In: 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), pp. 529–534. IEEE (2017)