Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

COMPUTER-ASSISTED TRAINING SYSTEM FOR YOGA AND PILATES USING COMPUTER VISION METHODS

STEVE BOBY GEORGE, Dr. JESWIN ROY DCOUTH, SANJAY JAIMY, EDWIN DAJI and AGNEL ANTONY

Albertian Institute of Science and Technology Cochin University P. O., Kochi, Kerala, India.

ABSTRACT:

In recent times after the COVID-19 outbreak. People are subjected to being confined at their homes and they are finding it extremely difficult to practice yoga or pilates without a coach beside them. Yoga and pilates practitioners must pay high fees for a personal trainer who would essentially help them identify their mistakes and correct them in real time. The idea of a coach or trainer monitoring the practitioner would improve his/her's output tremendously. To minimize the need for a trainer, the proposed system would act as a trainer and identify the user's mistakes and correct them on the go. This would solve the problem of health and body ailments due to incorrect postures and would decrease the expense borne by the practitioners drastically.

Keywords: MoveNet, MediaPipe, Keypoints, Feature Vector, TensorFlow.js.

1. INTRODUCTION

Yoga was created to help people enhance their physical, emotional, and spiritual well-being via meditation. Yoga, which focuses on flexibility and wide muscle areas through repetitive movement, has proven to be incredibly helpful in the long run. Many styles of yoga include meditation and deep breathing exercises, which are beneficial for relaxing the body and mind [1]. Invented by a man named Joseph Pilates in the early twentieth century, Pilates is a physical exercise regimen that emphasises core strength and flexibility. Pilates was developed during World War I with the goal of assisting injured troops in regaining their health through the strengthening, stretching, and stabilisation of certain muscular groups. Pilates employs fewer, more exact motions in order to attain the Pilates principles of proper alignment, centering of the body and focus. It also incorporates breathing and flow into the workout [2]. Yoga and Pilates are both primarily performed for the objective of enhancing one's overall health and well-being. This is due to the multiple benefits of these exercises, which include mental wellbeing, physical flexibility, good posture, and a variety of other elements, amongst others. Their popularity has skyrocketed in recent years, and they have been brought to the general public's attention as a result of this. When practising yoga or pilates, it is essential to maintain appropriate body posture at all times. Having perfect alignment keeps your joints in their places and allows you to work at your peak muscular capacity. Furthermore, it aids in the prevention of unintentional accidents, which are frequently caused by incorrect angles of the joints and unequal levelling of

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

the body's symmetry. A computer-assisted training system that can recognize the poses performed by the user and assist him/her in improving their stance by displaying relevant feedback is proposed in this paper. The system can be divided into mainly three modules, learn, practice, and review. Each module has its own relevance. The learn module can be used by the user to learn new poses. When used successfully, the practice module enables the user to perfect newly acquired poses. The review module can be used to view daily analytics of the user. The system is implemented as a web application for ease of access as well as cross platform compatibility.

2. LITERATURE SURVEY

2.1 Kyudo Posture Rectification System [3]

Kyudo is a traditional Japanese sport that gained popularity around 1949. AshiBumi-foot positioning, Dozukuri-torso positioning, Yugamae-body preparation, Uchiokoshibow raising, Hikiwake-bow drawing, Kai-full draw, Hanare-bow releasing, and Zanshin-form maintenance are the eight techniques in Kyudo. Proper posture is required for each of these strategies. Kyudo requires perfect hand and foot alignment before aiming at a target. Self-training is essential in Kyudo. In most situations, these training sessions necessitate the assistance of a trained instructor or coach, which can be rather costly. There is a demand for self-training that gives students effective training at a low cost. As a result, Wardah Farrukh et al. devised a Kyudo posture rectification system to assist archers in perfecting their posture. The proposed system simplifies and enhances Kyudo archer training sessions. The system solely addresses two major techniques: Kai and Hanare.

2.2 Computer-Assisted Yoga Training System [4]

To succeed in sports, most athletes and practitioners must spend time alone, in addition to the regular training sessions offered by a coach or instructor. Those who self-train without the guidance of a coach or instructor will likely make slow progress and may even injure themselves due to inappropriate postures or training methods. So many athletes and trainers are eagerly awaiting the introduction of computer-assisted training solutions that will help them improve their performance while lowering injury risk. Yoga, an ancient Indian technique, has gained popularity since it not only enhances physical health but also purifies the body, mind, and spirit. Yoga positions can cause catastrophic harm to muscles and ligaments if not performed correctly. So a computer vision-based yoga self-training system was created to help practitioners practise alone. Practitioners receive automatic postural guidance and feedback to help them improve their posture. In some cases, the system may be confused while extracting feature points from specific positions.

2.3 Golf Self-Training System [5]

Since a proficient golf swing is a crucial aspect of golfing performance, golfers consistently exert great effort to improve their swing abilities. However, without the guidance of a coach, a golfer is unable to accurately identify swing flaws and know how to alter swing mechanics. Using intelligent glasses, the proposed computer vision-based golf instruction system, "Improve My Golf Swing," enhances the golfer's swing technique. Using computationally efficient image processing techniques, the golfer's posture is evaluated. Incorporating professional golf training knowledge, the suggested system automatically generates feature lines to assess the golfer's body alignment. During a swing, the golfer can monitor both his or her personal posture and the enhanced feature lines presented on his or her intelligent glasses. Instead of quantitative swing analysis, the technology provides golfers with more intuitive and explicit advice for posture improvement. According to the posture error detected in relation to a pose, the golfer receives the associated audible alerts. More feature lines could be added to the system to improve posture analysis.

2.4 Karate Self-Training System [6]

A Hidden Markov Model (HMM)-based karate technique recognition system was proposed by Hachaj et al. The data was gathered with the use of a Microsoft Kinect camera. Later, individual body joints were used to determine angles-based properties that could be applied to them. This was followed by segmentation of the data signals collected from a recording with the use of a Gesture Description Language (GDL) classifier. The signal that was recognised by the GDL classifier was processed by the HMM classifier. A system for recognising karate techniques has been proposed, which, according to the findings, has shown to be successful. Karate movements are difficult to categorise because of their speed and agility. A misclassification was responsible for the vast majority of errors as a result.

3. EXPERIMENTAL SETUP

The experimental setup is as follows:

3.1 Yoga

The proposed system is designed to recognise the type of pose presented by the user and provide appropriate feedback based on the confidence score calculated. Currently, the existing datasets that are publicly available for yoga poses are aberrant. Thus, the dataset was prepared from scratch.

3.1.1 Implementation Details

TensorFlow.js was used to implement the proposed model. TensorFlow.js is a JavaScript package that allows you to create and run machine learning algorithms. Machine learning models can be run in a web browser as well as in the Node.js framework. The library is a component of the TensorFlow ecosystem, and it provides a set of APIs that are consistent with those in Python, allowing models to be transferred across the Python and JavaScript ecosystems [7].

3.1.2 MoveNet

The 17 key points of the body are detected using MoveNet, which is an extremely quick and accurate model. The Lightning and Thunder models are the two variations of the model that are available. Specifically, Lightning is designed for latency-critical applications, whereas Thunder is designed for applications requiring great precision. When it comes to live fitness, sports, and health applications, both models are significantly faster than real time (30+ frames per second) on the majority of modern desktops/laptops/phones. In order to accomplish this, TensorFlow.js is used to execute the model entirely client-side, in the browser, with no additional server calls required beyond the initial page load and no other dependencies to download and install. The thunder variant is used for the purpose of this study.

3.1.3 MoveNet Architecture

Move Net is a bottom-up estimate system that makes use of heat maps to correctly locate human key points in a given area. Two components make up the architecture: a feature extractor and a set of prediction heads. Prediction techniques such as Center Net have been used to develop this technique, however they have been changed to enhance speed and accuracy. All of the models were trained with the Tensor Flow Object Detection API, which is available on the Tensor Flow website. With a connected feature pyramid network (FPN), MobileNetV2 can be used as MoveNet's feature extractor. This allows for a high-resolution (output stride 4) and semantically dense feature map output in the Move Net environment [8]. The feature extractor has four prediction heads that are responsible for intensively predicting a:

1. The person centre heat map estimates the geometric centre of each individual instance.
2. Key point regression field: predicts a comprehensive set of key points for an individual; used to arrange key points into instances.
3. Person key point heat map: predicts the location of all key points regardless of the number of individual instances.
4. 2D per-key point offset field: forecasts local offsets between each output feature map pixel and the precise sub-pixel location of each key point.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

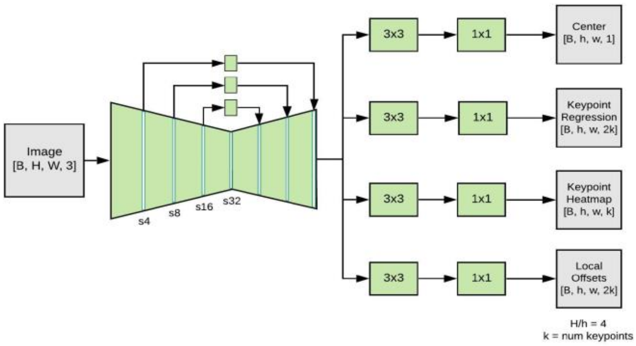

Figure 1 MoveNet Architecture

3.1.4 Dataset Preparation and Processing

Various images were gathered to create a dataset of five poses for beginner pose, they are Camel (Ustrasana), Plank (Phalakasana), Upward Dog (Urdhva Mukha Svanasana), Warrior I (Virabhadrasana I), Warrior II (Virabhadrasana II). A total of 1,562 images were captured for beginner and intermediate poses. Each image was cropped to bounding box coordinates before being inserted into its corresponding training and testing sets, removing everything in the image except the person performing the pose. After cropping the images, the next step was to move them to their respective directories. The training dataset is built by running labelled photos through MoveNet and then recording all landmark data and ground truth labels into a CSV (Comma-separated values) file because the MoveNet model's output landmarks are the input for the pose classifier. The model's training and testing split was set at 85% for training and 15% for testing. Both the train and test dataset are pre-processed using the MoveNet preprocessor.

3.1.5 Building Pose Prediction Model

The model can be broken down into two submodels:

The first sub model generates a feature vector based on the detected landmark coordinates (also known as a pose embedding). A feature vector is a vector containing information about an object in the form of many elements. The features may represent a single pixel or the complete image.

The second submodel predicts the pose class by feeding pose embedding through many thick layers. For the purpose of generating the feature vector, the landmark coordinates are transformed by relocating the pose's centre to the origin, scaling the pose such that the pose size equals one, and flattening these coordinates into a feature vector. This feature vector is used to train a neural network-based pose classifier. The neural network is specified using the Keras framework. Before predicting the pose class, the Keras model computes the feature vector using the detected pose landmarks. Figure 2 illustrates the model's configuration.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

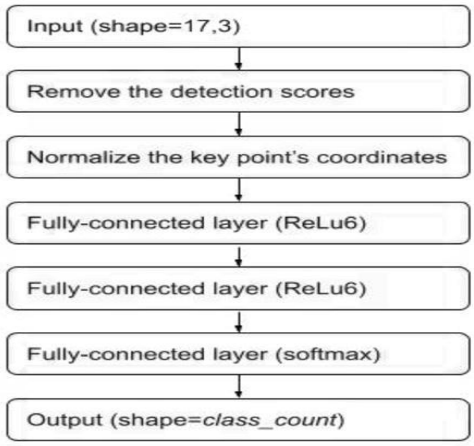

Figure 2 Neural Network Configuration

3.2 Pilates

The setup was designed to count the number of repetitions performed by the user.

3.2.1 Implementation Details

Implementation was done using Google's MediaPipe and classification was done using angle heuristics since pilates includes exercises that require movement.

3.2.2 Mediapipe

Using Google's MediaPipe open-source technology, it is possible to create Machine Learning pipelines that are multimodal and cross-platform. MediaPipe Pose is a machine learning approach for high-fidelity body pose tracking that uses BlazePose, a pose detection model, to infer 33 3D landmarks and a background segmentation mask on the complete body from RGB video frames [9]. BlazePose is a machine learning framework composed of two models: a Detector and an Estimator. The Detector extracts the human region from the input image, while the Estimator accepts as input a 256x256 resolution image of the discovered person and returns the key points [10].

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

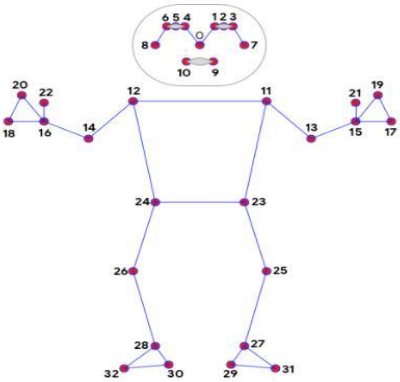

Figure 3 BlazePose 33 keypoints

3.2.3 Classification of Poses Using Angle Heuristics

A function for calculating the angles between three landmarks is defined. This procedure is equivalent to calculating the angle between two lines. The first point (landmark) represents the beginning of the first line, the second point (landmark) represents the end of the first line as well as the beginning of the second line, and the third point (landmark) represents the end of the second line. For example, if the exercise is a lateral rise, which targets the practitioners' deltoid muscles, the angle formed by the right and left shoulders (including hip, shoulder, and elbow landmarks) should be 77 degrees (approx.) and 10 degrees (approx.) for one repetition. Similarly, the minimum and maximum angles for a variety of pilates movements can be computed.

4. Results

4.1 Yoga

The proposed system's results demonstrate how well the model can recognise the pose provided by the user. The precision, recall, accuracy, and F1 ratings, according to Goutte et al, provide a comprehensive view of the system's performance [11]. The results imply that the model performs exceptionally well.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

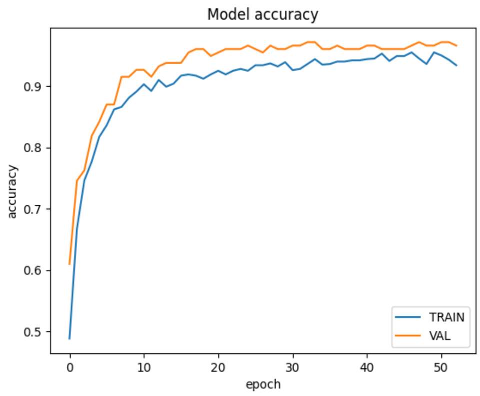

Figure 4 Accuracy Graph

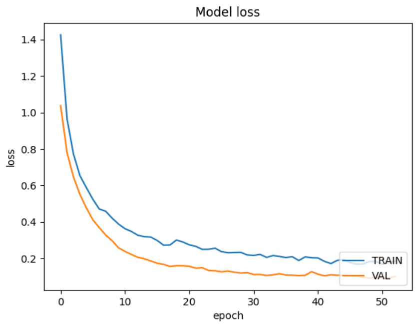

Figure 5 Loss Graph

The model accuracy and loss graphs are illustrated in Figs. 4 and 5. The accuracy of the model is approximately 0.97

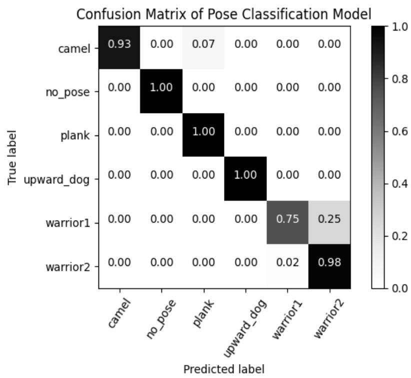

Figure 6 Confusion Matrix with normalization

The confusion matrix further shows that the model does an excellent job of classifying three samples correctly. Some samples of Camel (Ustrasana) pose were misclassified as Plank (Phalakasana) pose and similarly some samples of Warrior I (Virabhadrasana I) and Warrior II (Virabhadrasana II) were misclassified.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

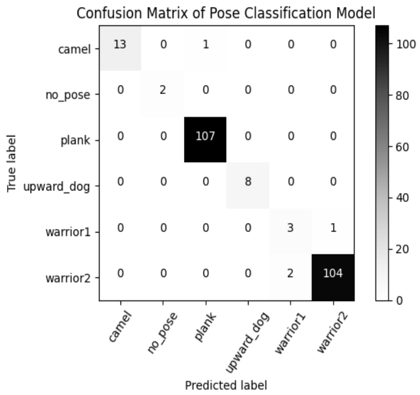

Figure 7 Confusion Matrix without normalization

```
Classification Report:
              precision    recall  f1-score   support

       camel       1.00      0.93      0.96        14
     no_pose       1.00      1.00      1.00         2
       plank       0.99      1.00      1.00       107
  upward_dog       1.00      1.00      1.00         8
    warrior1       0.60      0.75      0.67         4
    warrior2       0.99      0.98      0.99       106

    accuracy                           0.98       241
   macro avg       0.93      0.94      0.94       241
weighted avg       0.99      0.98      0.98       241
```

Figure 8 Classification report table of the predicted instances

Figure 9 Classification of Warrior II (Virabhadrasana II) Pose by the proposed model

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

4.2 Pilates

The results of the proposed system are satisfactory, and it calculates the number of repetitions performed by the end user with the appropriate precision for the selected pilates exercise.

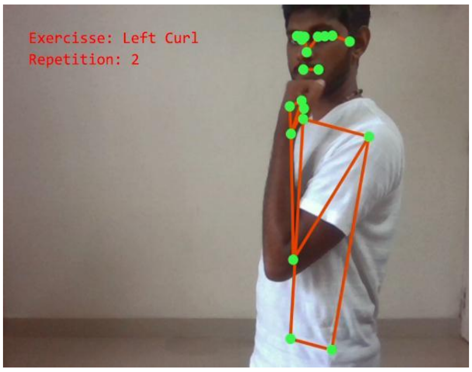

Figure 10 Left Curl counting based on the proposed system

Conclusion and Future Scope

By using computer vision methods, the proposed model can effectively distinguish between the techniques presented by the user and provide appropriate feedback based on calculations. The results indicate that the approach is feasible; nonetheless, effort for enhancement is required in specific areas. In the oncoming alterations, a larger dataset is to be used to increase the performance of both the yoga and pilates classifiers. Another feature to be implemented is to provide users with visual feedback and guided tutorials. Once the improvements have been made, this application can effectively assist the practitioners. The system has been developed with a social mission in mind. It aims to bring into awareness the importance of exercise and well-being in the younger crowd and also providing hassle free yoga and pilates assistance for the older population as they would require time and effort to work the poses. The study in its prime would essentially have both philanthropic and commercial relevance. The system could be implemented for a social cause as well it could be brought into the market for truly commercial purposes. That being said, this study is yet to be explored to its true potential and can be developed to an even bigger scale which could consist of millions of users.

References

[1] Ross, A., & Thomas, S. (2010). The health benefits of yoga and exercise: a review of comparison studies. *The journal of alternative and complementary medicine*, *16*(1), 3-12.
[2] Latey, P. (2001). The Pilates method: history and philosophy. *Journal of bodywork and movement therapies*, *5*(4), 275-282.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:06:2022
DOI 10.17605/OSF.IO/FXSQE

[3] Farrukh, W., & Haar, D. V. D. (2021). Computer-Assisted Self-Training for Kyudo Posture Rectification Using Computer Vision Methods. In *Proceedings of Fifth International Congress on Information and Communication Technology* (pp. 202-213). Springer, Singapore.
[4] Chen, H. T., He, Y. Z., & Hsu, C. C. (2018). Computer-assisted yoga training system. *Multimedia Tools and Applications*, *77*(18), 23969-23991.
[5] Chen, H. T., Huang, T. W., Chou, C. L., Tsai, H. C., & Lee, S. Y. (2015, December). Improving golf swing skills using intelligent glasses. In *2015 Visual Communications and Image Processing (VCIP)* (pp. 1-4). IEEE.
[6] Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods to Oyama karate techniques recognition. *Symmetry*, *7*(4), 1670-1698.
[7] Smilkov, D., Thorat, N., Assogba, Y., Nicholson, C., Kreeger, N., Yu, P., ... & Wattenberg, M. M. (2019). Tensorflow. js: Machine learning for the web and beyond. Proceedings of Machine Learning and Systems, 1, 309-321.
[8] Jo, B., & Kim, S. (2022). Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices. *Traitement du Signal*, *39*(1), 119-124.
[9] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*.
[10] Josyula, R., & Ostadabbas, S. (2021). A Review on Human Pose Estimation. *arXiv preprint arXiv:2110.06877*.
[11] Goutte, C., Gaussier, E.: A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: Lecture Notes in Computer Science, vol. 3408, pp. 345–359 (2005)