Yoga Pose Classification Using LSTM Architecture

1st Min Khant Soe
DSAI, School of Engineering
Asian Institute of Technology
Bangkok, Thailand
st122277@ait.asia

2nd Pyae Sone
DSAI, School of Engineering
Asian Institute of Technology
Bangkok, Thailand
st122645@ait.asia

3rd Win Win Phyo
DSAI, School of Engineering
Asian Institute of Technology
Bangkok, Thailand
st122314@ait.asia

***Abstract*—Human Pose Estimation aims to determine the localization of human joints, also known as key points, such as wrists, shoulders, elbows, knees, etc., from an image or video. In this century, people prefer to do exercises at home or online workouts, but they need the guidance of an instructor to train for these exercises, especially beginners. In this research paper, we use OpenCV and deep learning mechanism to estimate the human yoga pose.**

***Index Terms*—human pose, yoga, human pose recognition, OpenCV, Mediapipe, LSTM, deep learning**

I. Introduction

Since the early days of the computer vision field, Human Action Recognition has been a challenging problem since it comes with many difficulties regarding the technology. However, due to the rapid technological development in recent years, there has been a lot of research regarding HAR in real-world applications such as education, healthcare, video surveillance, abnormal activity detection, sports, and entertainment. [1].

Among them, the fields of fitness and exercise have attracted many researchers. One form of exercise that requires perfect posture is yoga practice. Nowadays, overweight adults make up roughly 39% of the world's population [2]. The significance and need for playing yoga is apparent from the earlier material. Playing yoga not only helps us shed pounds, but it also helps us stay active and maintain a healthy weight, a nice body, and a peaceful mind if we do it on a routine basis. Since most people don't have the luxury of having a personal instructor or have the time to practice in a decent gym, an application that has a deep learning mechanism that is capable of determining a person's yoga posture and providing feedback based on their activity might be useful for modern-day people [3].

This project uses five yoga poses: Cobra, Down Dog, Tree, Warrior1, and Stranding Forward Bend, with a combination of OpenCV Mediapipe for pose recognition and the LSTM algorithm to estimate the posture classification. In this project, we expect to achieve that our deep learning model has the accuracy of 90% or above, and classify the yoga poses accurately.

II. Methods

A. Data Acquisition

A data set consisting of the yoga videos has been collected from two yoga practitioners and some students from the Asian Institute of Technology (AIT). In the data set, there are 30 videos of each of the five yoga poses chosen by us for this project. These selected 5 yoga poses are described with explanation below.

1) Down Dog: The head is lowered till it touches the floor, and the body's weight is supported by the palms and feet. The arms are extended out straight in front of you, shoulder width apart; the feet are a foot apart, the legs are straight, and the hips are as high as possible.

2) Cobra: Place your palms beneath your shoulders and press down until your hips lift slightly. The backs of the feet are on the ground, the legs are spread out, and the gaze is directed forward in the preparation stance. The back is arched until the arms are straight, and the gaze is pointed straight upwards or slightly backwards in the full stance.

3) Tree: The entire sole of the foot remains in contact with the floor. In the half lotus position, the right knee is bent and the right foot is placed on the left inner thigh.

4) Warrior1: The ankles and calves will be stretched, the quadriceps and back will be strengthened, the psoas will be lengthened, and the upper body and arms will be stretched.

5) Forward Standing Bend (Uttanasana): The posture is entered by bending forward at the hips until the palms may be put on the floor, ending behind the heels, from a standing position.

B. Feature Extraction

For feature extraction, the cross-platform framework Mediapipe Holistic pipeline for yoga pose tracking proposed by Google was used as a tool. Our project is available for working not only with the photo and video files but also with a live video stream from a webcam. The Mediapipe python library can be made a detection of human poses with a list of coordinates in the X, Y and Z axis and visibility of each landmark [4]. In the pose detection of Mediapipe, there are 33 specific landmarks, which are the first 11 key-points from 0 to 10 for facial landmarking, the next 11 key-points from 11 to 22 for upper body detection, and the last 11 key-points from 23 to 32 for the detection of the lower body, as shown in figure 1. Hence, the total key points detected is 132 since each of the coordinates and visibility of landmarks has 33 key points.

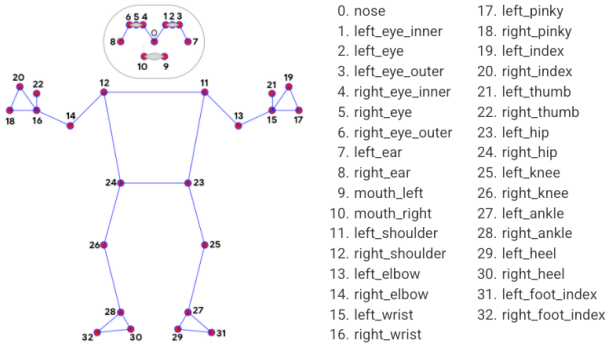

Fig. 1. Mediapipe Pose Landmark Model from mediapipe documentation

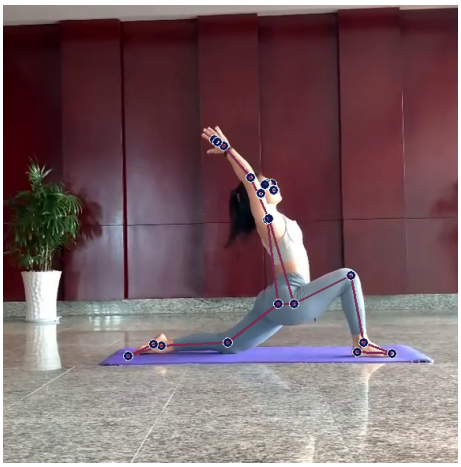

Fig. 2. Yoga pose detection with Mediapipe

C. Deep Learning Model - Long short-term memory

Long short-term memory (LSTM) networks are a sort of recurrent neural network that may learn order dependence in sequence prediction challenges. In this project, LSTM is used to classify the yoga poses since the data set we used is time series based; in a time series, LSTM can deal with unknown latencies between critical occurrences. The LSTM model architecture used for this project is shown in Table 1. In the model, "Adam" is used as an optimizer since it brings together the finest features of the AdaGrad and RMSProp methods to provide an optimization technique that can handle sparse gradients on noisy issues. Besides, "categorical cross-entropy" is used as a loss function because of its ability to minimize the difference between predicted and actual probability distributions. An activation function is used in every layer to provide the model with some form of non-linear feature. The last LSTM layer is followed by dense layers.

D. Model Training and Evaluation

As for the set of training samples, images are captured for five sequences at 10 fps (frames per second) from each video. Since we collected 30 videos for each yoga pose, we got a total of 1500 images for the whole data set. For training and validation, we split the data into 80:20. Then, the model is trained using training data and evaluated with validation data.

TABLE I
LSTM MODEL ARCHITECTURE

No of Layers	*Input Shape*	*Output Shape*	*Activation*
Layer 1: LSTM	(5, 132)	(5, 64)	ReLU
Layer 2: LSTM	(5, 64)	(5, 128)	ReLU
Layer 3: LSTM	(5, 128)	64	ReLU
Layer 4: Dense	64	64	ReLU
Layer 5: Dense	64	32	ReLU
Layer 6: Dense	32	5	Softmax

III. RESULTS AND DISCUSSION

Our LSTM model testing accuracy hit around 94% (figure 4), and the loss function value was 0.07 (figure 3). From the plot figures, the accuracy gradually improves while the loss deteriorates. The accuracy of above 90% is adequate for LSTM. To improve the accuracy, we can add more data, increase the number of epochs, or add more LSTM layers, but even so, our model will not improve significantly since it has already reached 94% accuracy in testing.

In human pose detection and/or classification, using the video data set is more effective since it detects the sequences of human actions as data. However, it can also confuse the model by having some similar data from a similar sequence of actions.

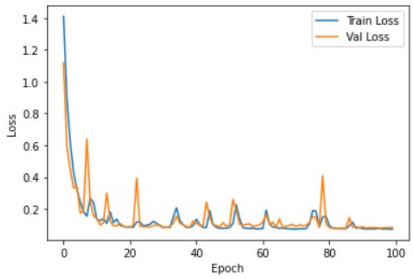

Fig. 3. Loss Curves

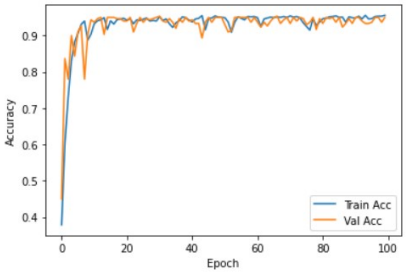

Fig. 4. Accuracy Curves

IV. ACKNOWLEDGMENT

Firstly, we thank Professor Matthew N. Dailey for allowing us do this project and guiding us. Next, we thank two

professional yoga practitioners and some students at AIT for participating in obtaining the video data set for our project.

V. Conclusion

This project can contribute to the use of pose estimation in fitness and sports, which helps people improve their workout performance. Since the implementation of an LSTM deep learning model has highly effective results at recognizing five yoga poses, and an accuracy of 94%, we accomplished this project successfully, achieving our initial goals.

References

[1] C. N. Phyo, T. T. Zin, and P. Tin, "Deep Learning for Recognizing Human Activities Using Motions of Skeletal Joints".

[2] A. Anilkumar, Athulya K.T., S. Sujan and Sreeja K.A., Pose Estimated Yoga Monitoring System.

[3] S. Kothari, "Yoga Pose Classification Using Deep Learning."

[4] M. G. Grif and Y. K. Kondratenko, "Development of a software module for recognizing the fingerspelling of the Russian Sign Language based on LSTM," 2021 J. Phys.: Conf. Ser. 2032 012024.