

Yoga Pose Detection and Classification Using Deep Learning

Project

CSE 4238

Soft Computing Lab

Submitted by

Sabikun Nahar	170204093
Md. Latifur Rahman	170204102
Md. Abir Hossain	170204106

Submitted To

Nibir Chandra Mandal
Mr. H M Zabir Haque



Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

Dhaka, Bangladesh

March 15, 2022

ABSTRACT

An approach to accurately recognize various Yoga asanas using deep learning algorithms has been presented in this work. A dataset of six Yoga asanas (i.e. Bhujangasana, Padmasana, Shavasana, Tadasana, Trikonasana, and Vrikshasana) has been created using 15 individuals (ten males and five females) with a normal RGB webcam and is made publicly available. A hybrid deep learning model is proposed using convolutional neural network (CNN) and Support Vector Machine(SVM) for Yoga recognition on real-time videos, where CNN layer is used to extract features from keypoints of each frame obtained from OpenPose and is followed by LSTM to give temporal predictions. To the best of our knowledge, this is the first study using an end-to-end deep learning pipeline to detect Yoga from videos. Experimental results provide a qualitative assessment of the method as well as a comparison to the state-of-the-art.

Contents

ABSTRACT	i
List of Figures	iii
List of Tables	iv
1 Introduction	1
2 Motivation	2
3 Methodology	4
3.1 SVM	4
3.2 CNN	4
4 Experiments	5
4.1 Dataset Collection	5
4.2 Dataset Description	5
4.3 Statistic of your dataset	6
5 Results/Evaluation	7
5.1 Performance of your model.	7
5.2 Comparison between our result and actual paper	7
5.3 Discussion about these two results	9
6 Conclusion And Future Work	10
6.1 Conclusion	10
6.2 Future Work	11

List of Figures

5.1 CNN Loss VS Iteration in our model	8
5.2 CNN Loss VS Iteration in the paper	8

List of Tables

4.1	Dataset details	5
4.2	Key Features	6
5.1	Accuracy	7

Chapter 1

Introduction

An asana is a body posture, originally and still a general term for a sitting meditation pose, and later extended in hatha yoga and modern yoga as exercise, to any type of position, adding reclining, standing, inverted, twisting, and balancing poses. Human activity recognition is a well-established computer vision problem that has imposed several challenges over the years. It is the problem of locating keypoints and the posture of a human body from the sensor data. Activity recognition is useful in many domains including biometrics, video-surveillance, human-computer interaction, assisted living, sports arbitration, in-home health monitoring, etc.. Human posture assessment is a difficult issue in the control of PC vision. It manages confinement of human joints in a picture or video to shape a skeletal portrayal. The issue with yoga anyway is that, much the same as some other exercise, it is of most extreme significance to rehearse it accurately as any erroneous stance during a yoga meeting can be ineffective and conceivably inconvenient. This prompts the need of having a teacher to manage the meeting and right the person's stance. utilized profound figuring out how to distinguish 5 exercise presents: pull up, swiss ball hamstring twist, push up, cycling and strolling. Utilizing this technique for yoga presents is a moderately more up to date application. Diverse posture extraction strategies are talked about alongside profound learning based models - Convolutional Neural Organizations (CNNs) and Recurrent Neural Networks (RNNs). For the dataset, we use 2 models. These are Convolutional Neural network and Support Vector Machine.

Chapter 2

Motivation

Yoga is an ancient science and discipline originated in India 5000 years ago. It is used to bring harmony to both body and mind with the help of asana, meditation and various other breathing techniques. It brings peace to the mind. Due to the increase of stress in the modern lifestyle, yoga has become popular throughout the world. There are various ways through which one can learn yoga. Yoga can be learnt by attending classes at a yoga centre or through home tutoring. It can also be self-learned with the help of books and videos. Most people prefer self-learning but it is hard for them to find incorrect parts of their yoga poses by themselves. Using the system, the user can select the pose that he/she wishes to practice. He/she can then upload a photo of themselves doing the pose. The pose of the user is compared with the pose of the expert and the difference in angles of various body joints is calculated. Based on this difference of angles, feedback is provided to the user so that he/she can improve the pose. Yoga is very important for our regular life.

- In case of a physical injury or illness, holding back from doing Yoga is, in fact, a good idea. Certain Yoga poses can make a physical injury worse. However, targeted Yoga poses can also accelerate the healing of an injury.
- Lack of fitness, strength, or flexibility is not an impediment to doing Yoga. In fact, these issues are the very reason for doing Yoga! The key is to choose the right class or online video that will be respectful of your level of fitness, strength, or flexibility and allows you to ease into Yoga with grace and joy.
- Feeling tired can be an overwhelming reason for avoiding Yoga – it feels like it will take too much energy. On the other hand, Yoga, and especially Yoga in the morning, can be instrumental in reducing tiredness and bring you back to a greater level of energy.
- Feeling uncomfortable in Yoga class can be a surprisingly big deterrent to doing Yoga.

It could be because we feel a bit shy and embarrassed about our Yoga skills by comparison to the others in the class. We may also feel uncomfortable with your Yoga teacher.

- This is probably one of the most common reasons for giving Yoga a miss. Everything just feels too much. However, Yoga and the associated breathing and meditation exercises enable us to connect with ourselves. Through this, we can move to a place of greater calmness and resourcefulness. This, in turn, creates more space for activities such as Yoga.
- A lovely home Yoga space can make all the difference to wanting practice Yoga. Often this is not readily available in small, busy or cluttered homes. However, we really only need quite a small space to roll out a Yoga mat. That can be literally anywhere in a house – even in a corner of a room.

Chapter 3

Methodology

3.1 SVM

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text. Compared to newer algorithms like neural networks, they have two main advantages: higher speed and better performance with a limited number of samples (in the thousands). This makes the algorithm very suitable for text classification problems, where it's common to have access to a dataset of at most a couple of thousands of tagged samples.

3.2 CNN

A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be. The human brain processes a huge amount of information the second we see an image. Each neuron works in its own receptive field and is connected to other neurons in a way that they cover the entire visual field. Just as each neuron responds to stimuli only in the restricted region of the visual field called the receptive field in the biological vision system, each neuron in a CNN processes data only in its receptive field as well. The layers are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns (faces, objects, etc.) further along. By using a CNN, one can enable sight to computers. A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

Chapter 4

Experiments

4.1 Dataset Collection

This dataset is publicly available at <https://archive.org/details/YogaVidCollected>. It comprises of recordings of 6 yoga presents performed by 15 distinct people (5 females and 10 guys). The 6 yoga presents specifically are – Bhujangasana (Cobra present), Padmasana (Lotus present), Shavasana (Corpse present), Tadasana (Mountain present), Trikonasana (Triangle posture) and Vrikshasana (Tree present). The all out number of recordings is 88 with a term of 1 hour 6 minutes and 5 seconds. The rate at which the recordings have been recorded is 30 FPS.

4.2 Dataset Description

Serial No.	Yoga Pose	No. of Videos
1	Bhujangasana	16
2	Padmasana	14
3	Shavasana	15
4	Tadasana	15
5	Trikonasana	13
6	Vrikshasana	15
	Total Video	88

Table 4.1: Dataset details

4.3 Statistic of your dataset

We used 45 frames of each video starting the first one third of the video for simplicity. Then we extracted key features from the selected frame using OPENPOSE library. The key features are given in the table. We collected total 3960 frames from 88 videos. After that we reshaped our data as (3960, 18, 2) for CNN model and (3960, 1, 36) for SVM model.

No.	Keypoint	No.	Keypoint
0	Nose	9	Right knee
1	Neck	10	Right foot
2	Right Shoulder	11	Left hip
3	Right elbow	12	Left knee
4	Right wrist	13	Left foot
5	Left Shoulder	14	Right eye
6	Left elbow	15	Left eye
7	Left wrist	16	Right ear
8	Right hip	17	Left ear

Table 4.2: Key Features

Chapter 5

Results/Evaluation

5.1 Performance of your model.

Model	Train Accuracy	Test Accuracy
SVM	99.8%	72.9%
CNN	92.0%	89.0%

Table 5.1: Accuracy

5.2 Comparison between our result and actual paper

Paper Model	CNN	SVM
	Training accuracy: 98.7%	Training accuracy: 99.5%
	Testing accuracy: 98.5%	Testing accuracy: 93.1%
Our Model	Training accuracy: 92.0%	Training accuracy: 99.8
	Testing accuracy: 89.0%	Testing accuracy: 72.9%

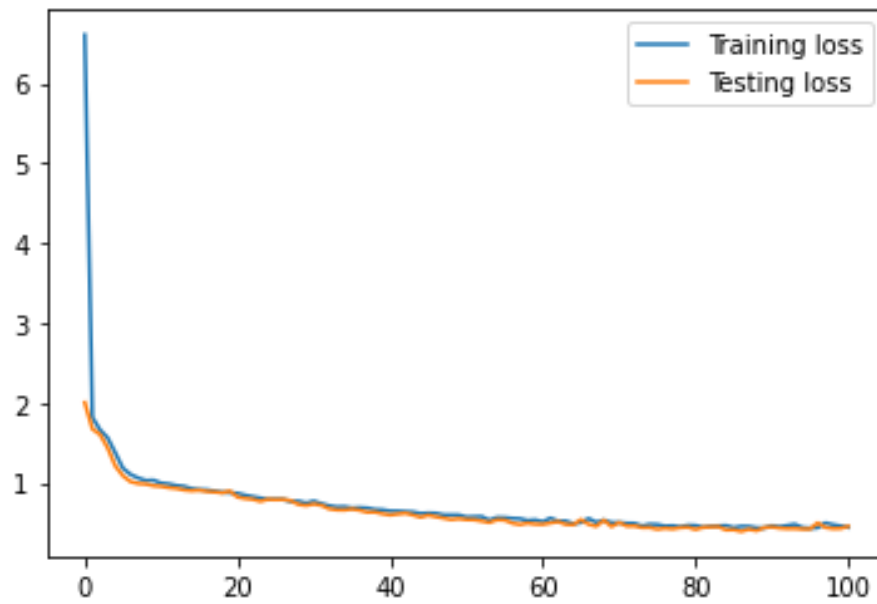


Figure 5.1: CNN Loss VS Iteration in our model

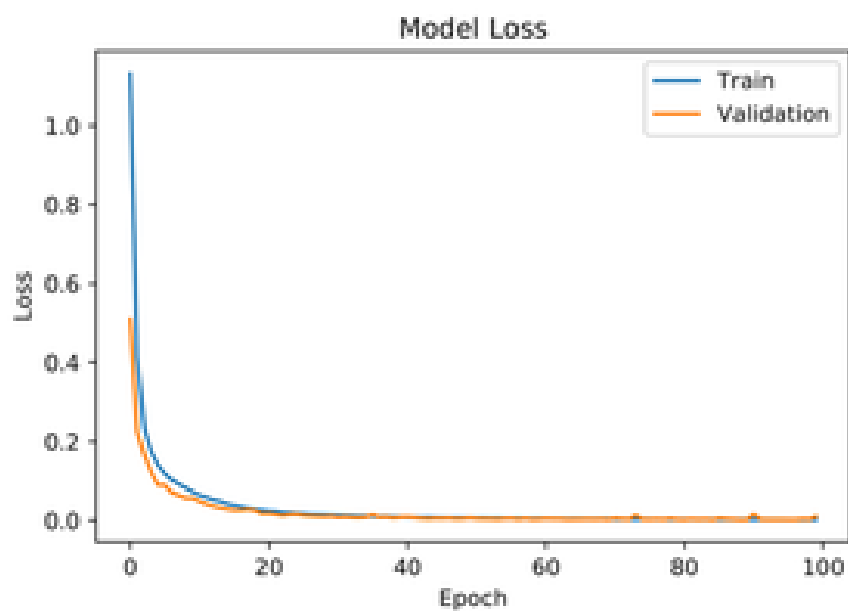


Figure 5.2: CNN Loss VS Iteration in the paper

5.3 Discussion about these two results

The Paper that we are using as reference has more accuracy then our experiment. We think that the main issue is that we used a portion of the total video length (1.5 seconds) as our reference data. And another reason is we used slightly different model from the paper.

Chapter 6

Conclusion And Future Work

6.1 Conclusion

In this paper, we proposed a Yoga identification system using a traditional RGB camera. The dataset is collected using HD 1080p Logitech webcam for 15 individuals (ten males and five females) and made publicly available. OpenPose is used to capture the user and detect keypoints. The end-to-end deep learning-based framework eliminates the need for making handcrafted features allowing for the addition of new asanas by just retraining the model with new data. Human posture assessment has been concentrated widely over the previous years. When contrasted with other PC vision issues, human posture assessment is distinctive as it needs to limit and amass human body parts based on an effectively characterized structure of the human body. The utilization of mixture CNN and SVM model on OpenPose information apparently is profoundly successful and arranges all the 6 yoga. There are various yoga asanas, and subsequently making a posture assessment model that can be effective for all the asanas is a testing issue. The exhibition of the models depends upon the nature of OpenPose present assessment which may not perform well in instances of cover between individuals or cover between body parts. A methodology practically identical to this can be used for present acknowledgment in undertakings for example, sports, reconnaissance, medical services and so forth Multiindividual posture assessment is a totally different issue in itself and has a great deal of degree for research. The methodology of utilizing CNN and SVM on present information got from OpenPose for Yoga pose identification has been discovered to be exceptionally compelling.

6.2 Future Work

In the future, we will deploy more deep learning techniques like Long Short term Memory(LSTM) to improve the performance of the detection system.