CLASSIFICATION OF WORK-OUT POSE USING DEEP LEARNING

Raikibul HASAN

Prototyping with Deep-Learning

Master of Data Science, University of Luxembourg

Abstract: Pose estimation is a supervised machine learning task for estimating the human's different types of body poses or classifying and identifying the joints of the human body from images, video clips, or real-time video. Detecting a human's pose automatically from real-time video or image is a difficult task with good accuracy. To build an exercise instructor platform that will help old people meet online and do exercises related to heart disease accurately on their own, we need an artificial model that can properly classify different types of exercises. To get an incredible output for this purpose, a deep learning model can be the best fit. In this paper, we describe a deep learning model to identify all bone joints of a human body, and then the training sets have used to train the model to classify the workout pose of a human body.

Keywords – Human pose estimation, exercise instructor, machine learning, deep learning.

1. INTRODUCTION

Regular physical activity has been shown to reduce morbidity and mortality by decreasing heart disease, diabetes, high blood pressure, colon cancer, feelings of depression/anxiety, and weight while building and maintaining healthy bones, muscles, and joints.[1] It's documented that regular physical activity is essential for healthy aging [2][3]. The American College of Sports Medicine (ACSM) and the American Heart Association (AHA) defines protocols, guidelines, and recommendations regarding the exact type and intensity of elderly exercise regimens [2][3][4][5].

Online workout platform-assisted solutions for elderly people. Pose estimation is one of the most interesting machine learning areas since it is used in different fields, including activity recognition, animation, gaming, augmented reality, etc. An online workout platform for older people patient where the deep learning model classifies their pose and give the accuracy of their pose might be an effective way for their exercise. So, for AI trainer application, realistic gaming fields, in term of health and medical system pose estimation is an interesting field.

Nowadays, pose estimation has achieved a huge gain in performance by using deep learning. Some existing libraries exist like OpenCV, meidaPipe, and TensorFlow to detect human body parts. In this paper, we use the TensorFlow library to find out body part coordinates (17 landmarks) from images. The work has been done before for pose estimation; most of those are in 2D with fewer

yoga poses. The six exercises have been chosen in this paper are related to Cardiac Rehabilitation. The problem is to get better accuracy for exercise and different poses because of background or surroundings, visibility, clothing variations, etc. Therefore, for every image, we take 17 different coordinates values (X, Y) and additionally 17 values for the ground truth of each pose.

To make a classifier using deep learning, we first use the TensorFlow lite model to extract the coordinate of the human joint. Afterward, we build a classifier model with six different types of exercise using CNN. The evaluation of this classification system will be done by using classification scores and a confusion matrix. The model makes predictions of six different exercises from images, and we can examine is the prediction is correct or not. The contributions of this paper are summarized as follows:

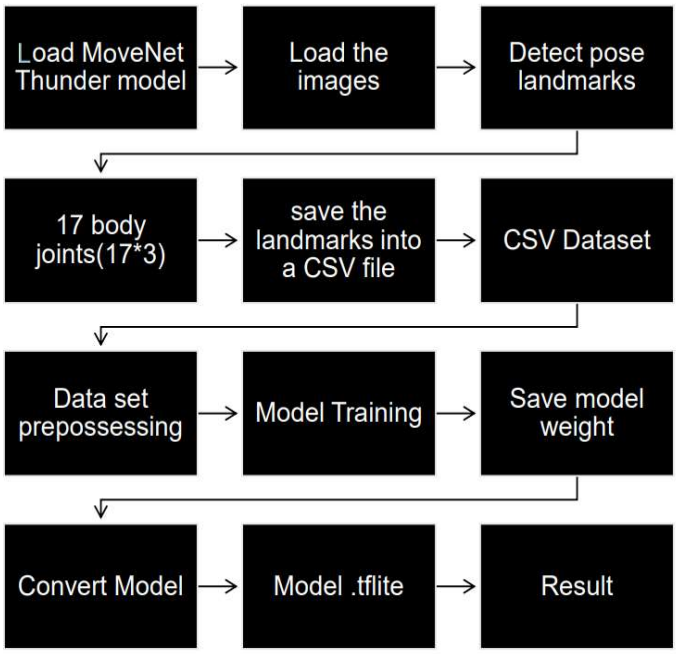

Figure 1: Proposed system model

2. RELATED WORK

The researchers [6] in their paper discuss the issues in human pose estimation and gives the overview of considerable research work in pose estimation, including deep learning approach. They reconstructed a model using convolutional neural network that estimates the poses and demonstrates the potential of CNN's. This paperwork estimate poses in 2 dimensional and take 14 key points of human body. The author of this paper suggested to label the images based on the activity category and to estimate the poses.

The authors [7] in their paper proposed a method for human pose estimation which extends common unary and pairwise terms of graphical models with a global foreground term. They present a branch and bound based algorithm to retrieve the globally optimal solution to the pose estimation problem. The model performs better only When the foreground part is available.

The author [8] has proposed a human pose estimation model with spatial context relationships based on graph convolution. Their graph convolutional network with spatial context relationships (GCN-SCR) method correctly joints by graph convolution.

This paper [9] focuses on the state-of-art progress of 2-D human pose estimation methods based on deep learning. 2 3 Basically, it is a review paper. They summarize and analyze different deep learning method attributes and performance which was used for human pose estimation.

The authors [10,11] in their paper proposed a DNN regressors that results in high precision pose estimates. The pose estimation is formulated as a DNN-based regression problem towards body joints. The method has a powerful formulation and has the benefit of reasoning about pose in a holistic fashion.

The proposed work objective is classification and posed estimation. For the former task, the aim is to identify x-y coordinates and ground truth value for body joints. For the latter, the aim of this paper to make a classifier using Convolution Neural Network (CNN) which can estimate pose.

3. DATASET

The dataset collected from Kaggle datasets is a publicly available and open-source collection. The datasets consist of a variety of pose images. This paper is mainly focused on cardiac rehabilitation exercise; therefore, we choose six yoga poses that can fit heart disease patients. The poses are Tadasana (Mountain pose) (Figure 2) ,Warrior 1((Figure 3), Vrikshasana (Tree pose) (Figure 4), Bitilasana (Figure 5), Dandasana (Figure 6), warrior (Figure 7). The total number of the image is 2523.

Images have been taken in indoor and outdoor environments at different angles and distances from the camera. Individual images have been performed with many variations to build a robust pose recognition model. Three different image files are taken to build the dataset, namely, jpg, png, and bmp. The size of the image dataset is 525 MB. The image shows the variation of different ages, people, and gender.

Figure 2: Mountain Pose

Figure 3: Warrior 1

Figure 4: Tree Pose

Figure 5: Bitilasana

Figure 6: Dandasana

Figure 7: Warrior 2

4. MATERIALS & METHOD

First, need to capture a set of coordinates for each joint and then connect those with the edge to estimate the pose. The model first identifies the body part localization as input and outputs a low-resolution per-pixel heatmap. This heatmap shows the probability of a joint occurring at each spatial location in the image.[12]

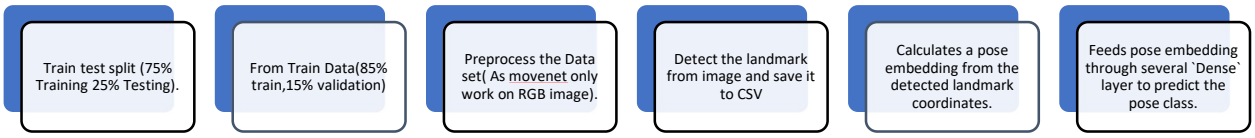

Figure 8: Model summary

There are three different approaches to modeling the human body: Skeleton-based, Contour based, and Volume-based model. In this work Contour type of approach is chosen to model the human body. Model will:

1. Detect the pose of a single person (**3ft ~ 6ft)**
2. Detect the pose of the person who is closest to the image center and ignore the other people who are in the image frame.
3. The model predicts **17 human key points** of the full body.

MoveNet Lighting is smaller, faster, and can run in real-time on browsers and modern smartphones. The PoseNet model's performance varies depending on the device and output stride [13]. Therefore, the MoveNet Lighting version has been used for this paper to estimate the keypoint of the human body. For this project, we are using MoveNet, which is the state-of-the-art pose estimation model that can detect these 17 key points:

1. Nose
2. Left eye
3. Right eye
4. Left ear
5. Right ear
6. Left shoulder
7. Right shoulder
8. Left elbow
9. Right elbow
10. Left wrist
11. Right wrist
12. Left hip
13. Right hip
14. Left Knee
15. Right knee
16. Left ankle
17. Right ankl

In order to train the model, we need to detect the human body joints from the image dataset. (Figure 9) The moveNet lite model detects the landmark data (x and y) and grounds truth labels into a CSV file. Save this value into a CSV file for six different exercise /yoga classes. Then we convert these values into a feature vector. Next, we use these vector values to train our neural network based on the pose classifier.

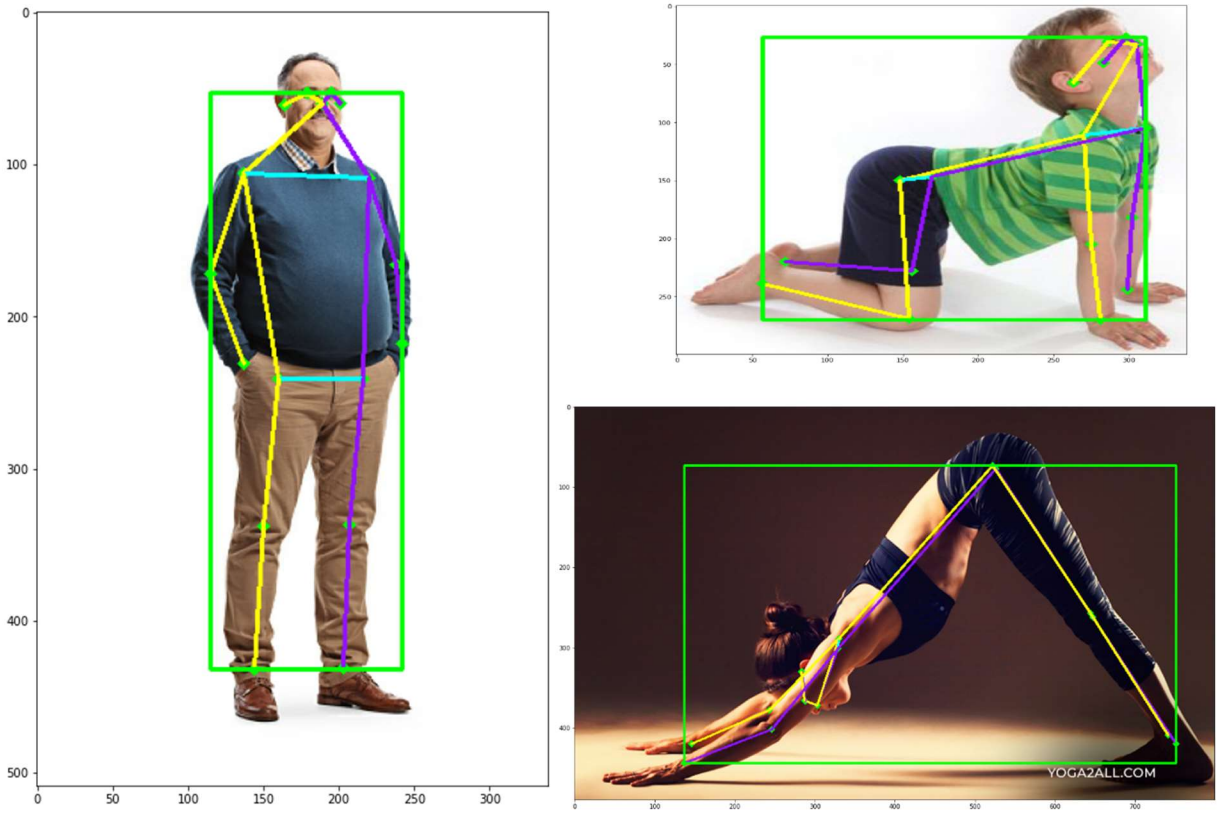

Figure 9: Estimate body point

Convolutional neural networks consist of multiple layers of artificial neurons and are widely used for image classification. In the case of key points, CNN extracts the feature (x, y) value and the ground truth value from the CSV file. Based on the filter size, The convolutional filter slides to the next set of input. After the convolution, an activation function Rectified Linear Unit (ReLU6) is generally applied to add nonlinearity in the CNN since the real-world data is mostly nonlinear and the convolution operation is linear [14]. Tanh and sigmoid are other activation functions, but ReLU is mostly used because of its better performance [9]. The Keras model takes the detected landmark coordinates to predict the pose class.

The loss function used for compiling the model is categorical cross - entropy which is also called Softmax loss. This is used as it allows measuring the performance of the output of the densely connected layer with softmax activation. This loss function is used for multi class classification, and as we have multiple yoga pose classes, it makes sense to use categorical cross entropy. Eventually, we use adam optimizer with an initial learning rate of 0.0001 to manage the learning rate. 300 epochs are used to train our model.

5. RESULT & DISCUSSION

The model is built-in Anaconda environment 3.9.7 using Python libraries like cv2, TensorFlow - Keras, NumPy, Pandas, and Scikit Learn on a system Ryzen 5 with 8GB RAM.

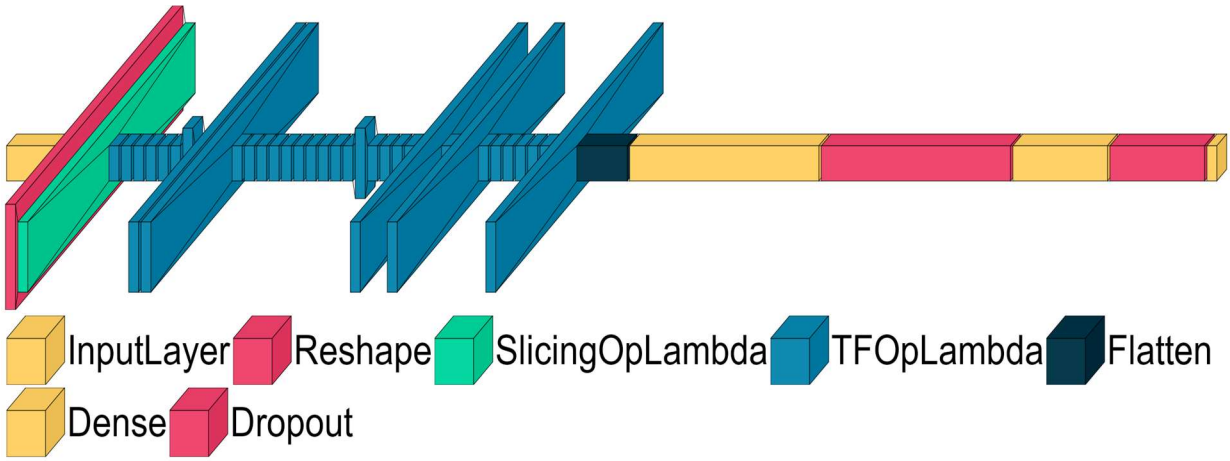

Figure 10: Model Layer

The training loss assesses the error on the training data of the model. It means how a model fits on the training set. On the other hand, validation loss assesses the error on the validation set of the model where the validation set is a part of the dataset to validate the model's performance.

In terms of training accuracy means how accurate our model can predict on the training dataset. On the other hand, validation accuracy means the performance of our model on validation data set. The figure:11 curves show that our model performance on the validation data set is more accurate, almost 99%.

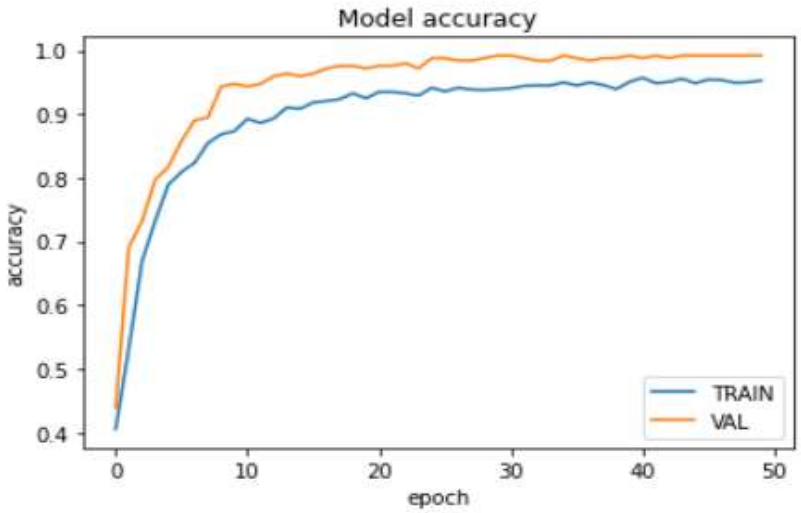

Figure 11: Model accuracy curve

The Train accuracy of our model is 0.9527. The validation loss of our model: 0.0646. The validation accuracy is 0.9919, and the Test accuracy is 0.9474. However, the model accuracy curve illustrates an increase in the training accuracy and a decrease in the validation accuracy, which means some underfitting.

Classification Report:

	precision	recall	f1-score	support
bitilasana	1.00	1.00	1.00	18
dandasana	1.00	1.00	1.00	13
mountain	0.96	0.94	0.95	83
tree	0.93	0.98	0.95	144
warrior1	0.89	0.91	0.90	94
warrior2	0.98	0.93	0.96	180
accuracy			0.95	532
macro avg	0.96	0.96	0.96	532
weighted avg	0.95	0.95	0.95	532

Figure 12: Model classification report

The classification report Figure 12 represent the precision, recall score and F1 score also known balanced F-score. F1- score metric is used to measure the performance of classification machine learning models. This F1-score metric provides robust results for both balanced and imbalanced datasets because it can evaluate a model's recall precision and ability is due to the way it is derived, which is as follows:

$$f1 = 2 * (precision * recall) / (precision + recall)$$

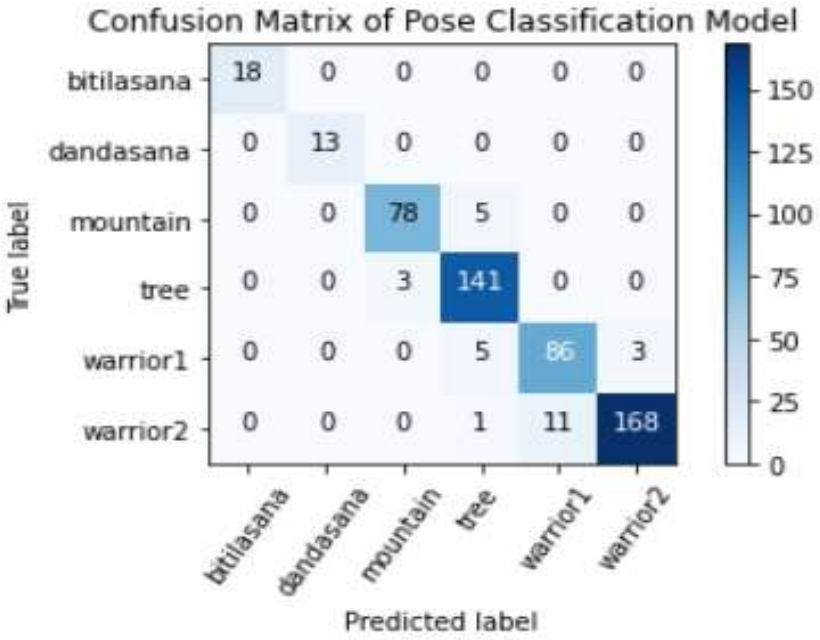

Figure 13: Confusion Matrix of model

The confusion matrix (figure 13) further represents that except for warrior2, the model predicts other poses with reasonable accuracy. The model misclassified 11 warrior2 poses as warrior1 and 5 warrior1 as tree pose.

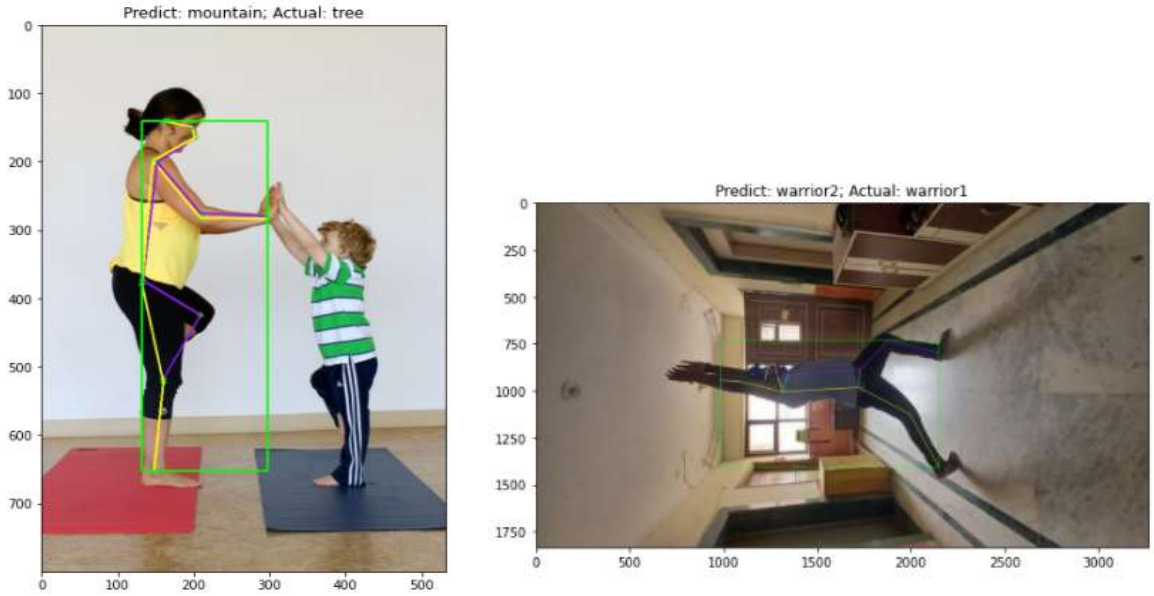

Figure 14: Incorrect predictions

6. CONCLUSION & FUTURE WORK

The model can classify only six yoga poses for a single person, which may upgrade multi-person pose estimation. A pose estimation model that can predict all yoga exercise-related cardiac rehabilitation is a challenging task. In this paper we introduce a large dataset of images that covers a wide variety of human poses and clothing types and includes people interacting with various objects and environments. The model's accuracy depends upon the quality of pose estimation of the TensorFlow moveNet lite model. Calculating the angle of every body part joint from the coordinate value might give a good accuracy for a complicated yoga pose. However, there is still a massive amount of work that we can continue to examine. Future work will focus on a real-time exercise classifier and calculate the angle of the body part to improve the model's accuracy.

ACKNOWLEDGEMENT

I would like to thank Professor Luis LEIVA, Professor Aldo ROMERO, Dr. Bereket YILMA for this amazing course - Prototyping with Deep-Learning where we learned the grammar of deep learning.

REFERENCES

[1] U.S. Department of Health and Human Services A report from the Surgeon General: physical activity and health, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, President's Council on Physical Fitness and Sports, Atlanta GA (1996)

[2] W. J. Chodzko-Zajko, D. N. Proctor, M. A. Fiatarone Singh, C. T. Minson, C. R. Nigg, G. J. Salem, and J. S. Skinner,:"American College of Sports Medicine position stand. Exercise and physical activity for older adults.:" Med. Sci. Sports Exerc., vol. 41, no. 7, pp. 1510–30, Jul. 2009.

[3] M. E. Nelson, W. J. Rejeski, S. N. Blair, P. W. Duncan, J. O. Judge, A. C. King, C. A. Macera, and C. Castaneda-Sceppa, "Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association.," Circulation, vol. 116, no. 9, pp. 1094–105, Aug. 2007.

[4] A. A. McDermott and H. Mernitz,:"Exercise and the Elderly: Guidelines and Practical Prescription Applications for the Clinician," J. Clin. Outcome Manag., vol. 11, no. 2, pp. 117 – 127, 2004.

[5] Bushman, Barbara Ann, editor. | American College of Sports Medicine, "A. C. of S. Medicine, ACSM's Complete Guide to Fitness & Health. Human Kinetics, 2011".

[6] Anubhav Singh1 , Shruti Agarwal2 , Preeti Nagrath3 , Anmol Saxena4 , Narina Thakur5,:"Human Pose Estimation Using Convolutional Neural Networks"

[7] Jens Puwein1, Luca Ballan1, Remo Ziegler2, and Marc Pollefeys1,: "Foreground Consistent Human Pose Estimation Using Branch and Bound"

[8] Na Han School of Computer and Information, Hefei University of Technology,: "Human pose estimation with spatial context relationships based on graph convolutional network"

[9] Yi Liu; Ying Xu; Shao-bin Li: "2-D Human Pose Estimation from Images Based on Deep Learning: A Review".

[10] Chen, X., Yuille, A.L.: Articulated pose estimation by a graphical model with image dependent pairwise relations. In: NIPS. (2014)

[11] Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: NIPS. (2014)

[12] S. Patil, A. Pawar, and A. Peshave, "Yoga tutor: visualization and analysis using SURF algorithm",Proc. IEEE Control Syst. Graduate Research Colloq.,pp. 43-46, 2011.

[13] A. Kendall, M. Grimes, R. Cipolla, "PoseNet: a convolutional network for real-time 6- DOF camera relocalization", IEEE Intl. Conf. Computer Vision, 2015.

[14] Kothari, Shruti, "Yoga Pose Classification Using Deep Learning" (2020).

[15] Efficient Object Localization Using Convolutional Networks (2015) Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, Christoph Bregler.