PAPER • OPEN ACCESS

Yoga Asana Identification: A Deep Learning Approach

To cite this article: Josvin Jose and S Shailesh 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1110** 012002

View the article online for updates and enhancements.

Yoga Asana Identification: A Deep Learning Approach

Josvin Jose[1], **Shailesh S**[2,3]
1Carmel College of Engineering and Technology Punnapra , Alappuzha ,Kerala , India

2Sacred Heart College, Thevara, Kochi, Kerala, India

3Department of Computer Applications, Cochin University of Science and Technology, Kalamserry, Kochi, Kerala, India

E-mail: *josvinvalekalam912@gmail.com, shaileshsivan@gmail.com

Abstract. Yoga is a healthy practice that originated from India, to rejuvenate a man in his physical, mental, and spiritual wellness. Moving with the brisk technology advancements, there is a vast opportunity for computational probing in all social domains. But still, the utilization of artificial intelligence and machine learning techniques for applying to an interdisciplinary domain like yoga is quite challenging. In this work, a system that recognizes a yoga posture from an image or a frame of a video has been developed with the help of deep learning techniques like convolutional neural networks (CNN) and transfer learning. We have considered images of 10 different asanas for training the model as well as evaluating the prediction accuracy. The prediction model backed with transfer learning shows promising results with 85% prediction accuracy and this system can be considered as an initial step to build an automated yoga image and video analysis tool.

1.Introduction

The innovations in technology and science are happening in a drastic phase, which makes human life more and more hassle-free. Nowadays everybody is aware of its relevance in day to day life. As in every domain, the influence of computers and computer-powered technologies are well established in health care and related domain. Apart from the usual medical practices, other practices like Yoga, Zumba, martial arts, etc are also widely accepted among society as a way to achieve good health. Yoga [1] [2] is a set of practices sprout out in ancient India which deals with the wellness of physical, mental, and spiritual condition of a man. Yoga has gained a big significance in the medical community. The benefits of yoga are improved health, mental strength, weight loss, etc. But Yoga must be practiced under the supervision of an experienced practitioner since any incorrect or bad posture can lead to health problems such as ankle sprain, stiff neck, muscle pulls, etc. The yoga instructor must correct the individual postures during training. After proper training only, one can practice Yoga on their own. However, in today's situation, peoples are more comfortable in their homes and everyone changes almost everything to online mode. This situation demands a technology-driven Yoga practice [3] [4]. This can be done with the help of mobile applications or with virtual tutoring applications. In both cases, there are vast opportunities to explore computationally for making them more powerful, intelligent, and efficient. We put forward a classification model for recognizing and identifying a yoga asana from an image or from a frame of a video in this work. The classification model [5] is developed using the deep learning techniques [6] backed with image processing and computer vision methods. The entire work focuses on classifying 10 classes of yoga asanas. The postures like Bridge, Child's, Downward dog, Mountain, Plank, seated forward bend, Tree, Triangle pose, Warrior1, Warrior2 were analyzed and recognized by the deep learning model and then used to identify the exact class of the yoga posture Figure 1.1 shows the sample images of these 10 classes of yoga asana:

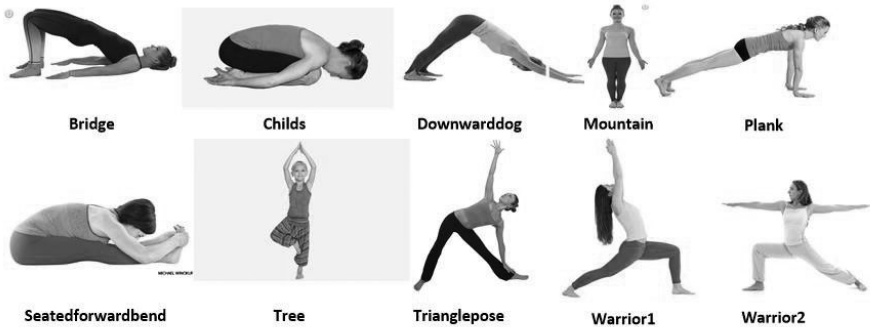


Figure 1.1: Yoga asanas

2.Literature Review

Since the domain of interest of this work, Yoga is not much explored, only a few works were found related to this. So, this section is logically partitioned into two, one which describes the works related to Yoga posture classification, and the other part describes some relevant works done in the general problem of human posture estimation and classification. Sruti Kothari [7] explained a computational method using deep learning, particularly CNN, for classifying the Yoga postures from images. They have considered a dataset containing 1000 images distributed over 6 classes for building the classification model. Nearly 85% accuracy was obtained for this work. Hua- Tsung Chen [8] proposed a yoga posture recognition system that can recognize the yoga posture performed by the trainer. In the first step, he used a kinetic to capture the body map of the user and body contour extraction. Then, the star skeleton which is a fast skeletonization technique by connecting from the centroid to other joint parts is done as the next step. From this technique, accuracy of 96% is acquired. Edwin W.trejo [9] introduced a model for pose correction in yoga. the user will receive real-time instruction about the pose correction made by an expert. for this, they used a recognition algorithm based on the AdaBoost algorithm to create a robust database for estimating 6 yoga poses. Finally, 92% accuracy is obtained.[10] Yoga is a traditional exercise that can bring harmony and peace to both body and mind. But self-learning yoga without the help of a instructor is a hard task. But here a solution is proposed for this, a photo of themselves doing the pose is uploaded then it compares to the pose of the expert and difference in angles of various body joints is calculated.[11] An alternative computationally efficient approach for yoga pose recognition in real world is presented.

A 3D CNN architecture is designed to implement real time recognition of yoga poses, it is a modified version of C3D architecture. For computational efficiency computationally fully connected layers are pruned, and supplementary layers such as the batch normalization and average pooling were introduced.as a result an accuracy of 91.5% is obtained.[12]A dataset consisting of 6 yoga asanas is created. For this work CNN which is used to extract features from key points and LSTM to give temporal prediction is used. As a result a accuracy of 99.04% is obtained. Shailesh s[13][14] proposed a way to annotate dance videos based on foot posture(stanas) in an automatic manner. For this, he used transfer learning. By using transfer learning the features from the images are extracted and a deep neural network is used for image classification. The accuracy obtained was a promising one. Matthias Dantone [15] proposes a unique way to estimate human poses from a 2D image by proposing novel, non-linear joint regression. As a joint regressors, they combined 2 layered random forests. A discriminative independent body part classifier was employed in the first layer, independent body part classifier, and in the second layer, by modeling the interdependence and co-occurrence of the part the joint locations is predicted. Muhammad Usama Islam[16] proposed a method to detect different joint points of a human body and from these points, we calculate the various angle to estimate the poses or asanas and the accuracy of it by Microsoft kinetic. By this method, if a person's angle accuracy falls to below 97% we conclude that the person could not get the pose done. Sadeka Haque [17]proposed a unique way to estimate human postures present in a two-dimensional human exercise image by using CNN, for this they proposed a dataset that contains 2000 images of 5 classes of exercise by conducting various experiments they finally achieved 82.68% accuracy. Sven Kreiss [18] introduced a multi-person pose estimation by using bottom-up strategy with the help of two-methods Part Intensity Field (PIF) and Part Association Field (PAF). The PIF method is used to localize body parts and PAF method is used to associate body parts to each other to form a full human pose, this method is suited for delivery robots and self-driving cars. An accuracy of 96% is obtained from this method. Dushyant Mehta [19] created a technique for 3D motion capture by using an RGB camera. the first step of this technique includes CNN that estimate 2D and 3 D pose features along with identifying all visible joints of the individuals. In this technique, he used a new architecture called SelecSLS Net to improve the information flow without decreasing the accuracy. In the second stage, the pose features of each individual are turned into a complete 3D pose estimate by using a fully connected neural network. In the third stage, to the predicted model a space-time skeletal model is fitted, and for each subject, a full skeletal pose in joint angles is returned. Dongyue Lv [20] created a model to train a golf player to make a perfect swing by capturing and remodel the swing movement in a portable way. To increase the capture accuracy a Dynamic Bayesian Network (DBN) model- based golf swing reconstruction algorithm is proposed, a smart motion reconstruction system for Golf swing (SMRG) is used based on the DBN model and kinetic as capturing device.

3. Methodology

The overall architecture of yoga asana classification is illustrated in the figure 3.1. Each step included the pipeline are explained below.

3.1 Dataset

The dataset for our work is created by Anastasia Marchenkova. In the dataset, there are 10 classes of different yoga poses. They are bridge, children, downward dog, mountain, plank, seated forward bend, tree, triangle pose, warrior1, warrior2. There are around 700 images in the dataset evenly distributed among 10 classes.

3.2 Preprocessing

The structural features of the data in the dataset of yoga asanas are particular for each class of asanas. To make it more compatible for further phases of the deep learning pipeline, the data have to be allied. In the pre-processing stage, we are passing the data through 3 different pre-processing steps as shown in figure. The data are collected from different sources so they will be different in their dimensions. In the first step of preprocessing the data of different dimensions are resized to 100x100 pixels. In the next step, we are using a Gaussian filter to filter out the noises in the data while keeping the edges sharp. And in the third and final step, the Histogram Equalization is used to bring the distribution of the image intensities to a normal fashion and increased contrast of the image is obtained.

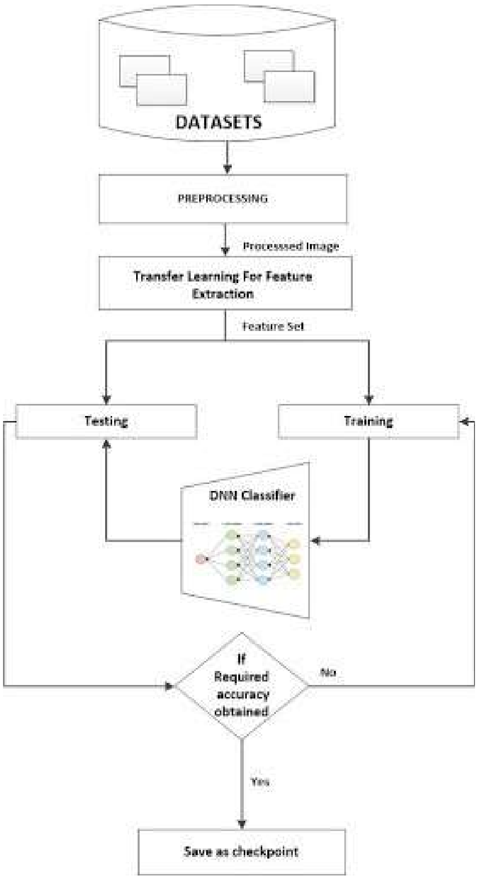


Figure 3.1. Architecture of Yoga Asana Classifier

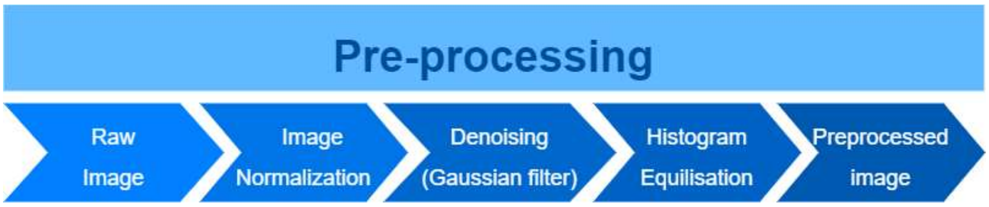


Figure 3.2. Image preprocessing steps

3.3 Transfer Learning (VGG16) for Feature Extraction

Convolutional neural network (CNN) [21] is a popular deep neural network widely used for visual image analysis. The general architecture of CNN is shown below. There are 4 central building blocks on CNN are Convolutional layer, activation layer, pooling layer, and fully connected layer. The Convolutional layer is a filter that passes over a picture and operating some matrix of pixels at a time. The output from this layer will be a layer which speaks about all the pixel the channel observed. The activation layer is a lattice that the convolution layer creates with less estimate than the first picture. This framework is run through an enactment layer, which presents non- linearity to permit the arrangement to prepare itself through back-propagation.

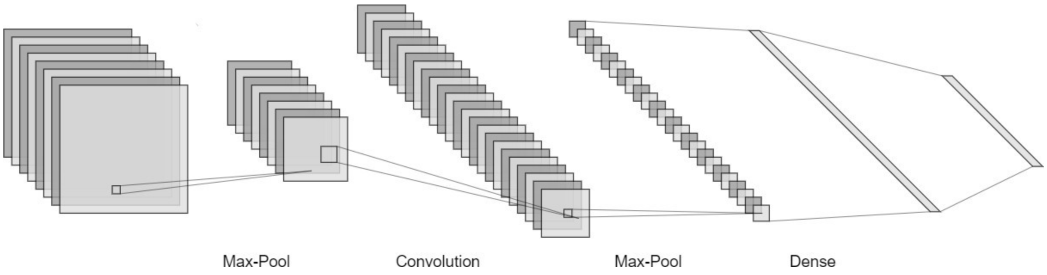


Figure 3.3. Convolutional Neural Network

work is ordinarily (Rectified Linear Unit) ReLu. The pooling layer uses the method ”pooling” to encourage down sampling and diminish the measure of the framework. A channel is passed over that comes out of the past layer and chooses one number out of each bunch of values (the most maximum value is selected and this is called max pooling). The fully connected layers are the traditional multi-layer perception structure. Its input is a one-dimensional vector representing the output. The label with the highest probability of acceptance is the classification decision.

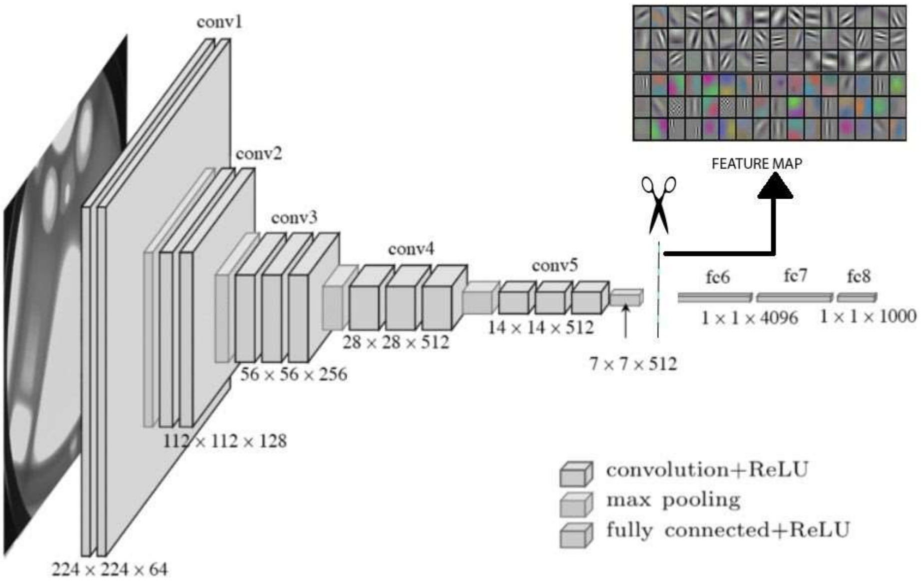


Figure 3.4. Transfer Learning

The Convolutional neuron network was the first attempt to make a CNN model but due to an insufficient amount of data, the performance obtained by the model was not satisfying. To overcome this, instead of manually extracting the features, we tried to obtain the features using transfer learning [22]. Transfer learning involves a method of transferring acquired information to one or more source tasks and is used to enhance the learning of similar target tasks. Although most machine learning algorithms are designed to solve a single task, a subject of ongoing interest in machine learning is to develop algorithms that promote transfer learning. It assumed that fully connected layers collect information relevant to solving a particular problem. For example, the fully connected layers of Alex Net will show the features that are essential for classifying an image into one of 1000 categories of artifacts. For a new problem, The last layer of features of the CNN state pre-trained on ImageNet and use the extracted features to create a new model. In practice, in order to ensure that the system will not fail to learn the previously acquired knowledge, we either maintain a fixed pre-training parameter or adjust it with a small learning rate. The figure below shows a transfer learning model in which a pre-trained deep learning model is used as a feature extractor.

The mathematical definition of transfer learning [23] is given in terms of domain and task. The domain consists of a feature space and a marginal probability distribution P (X) where X = {x1, ..., xn} ∈ χ is a specific domain, = {χ, P (X)} is a task that consists of two components: a label space Y and an objective predictive function{ f (o)}(denoted by T = {Y, f (o)}. f (o) is acquired from the training data consisting of pairs xi, yi , where xi X and yi Y . The function f(o) can be used to predict the corresponding label while f (x) uses a new instance, x. Given a source domain S and learning task TS , a target domain DT and learning task TT, transfer learning aims to help improve the learning of the target predictive function fT (o) in

DT using the knowledge in DS and TS where DS

DT or TS *f*= TT

3.4 Deep Neural Network Classifier

After the feature extraction process, the next phase is to train a deep neural network [24]. It is a fully connected network with one input layer, eight hidden dense layers, and an output layer. For regularization dropout layers are also added between some pairs of hidden layers. Except for the output layer, the output of all dense layers passes the ReLU activation function. The Softmax function is used in the last dense layer to normalize the network output to the probability distribution on the predicted output classes. Then ADAM optimizer and classification cross- entropy are used as the loss function to compile the network. Th figure x shows the structure of the entire DNN network.

4. Implementation and Result Analysis

	precision	recall	fl-score	support
0	0.60	0.75	0.67	16
1	0.67	0.55	0.60	11
2	0.96	0.89	0.92	27
3	0.98	1.00	0.99	43
4	0.42	0.62	0.50	8
5	0.81	0.92	0.86	24
6	0.80	0.67	0.73	12
7	0.88	0.70	0.78	10
8	0.60	0.50	0.55	12
9	0.88	0.79	0.83	19
Accuracy			0.81	182
macro avg	0.76	0.74	0.74	182
weighted avg	0.82	0.81	0.81	182

As a deep learning application, the implementation of this project was a bit challenging. The work was implemented using python with the help of libraries like TensorFlow, Keras, Panda

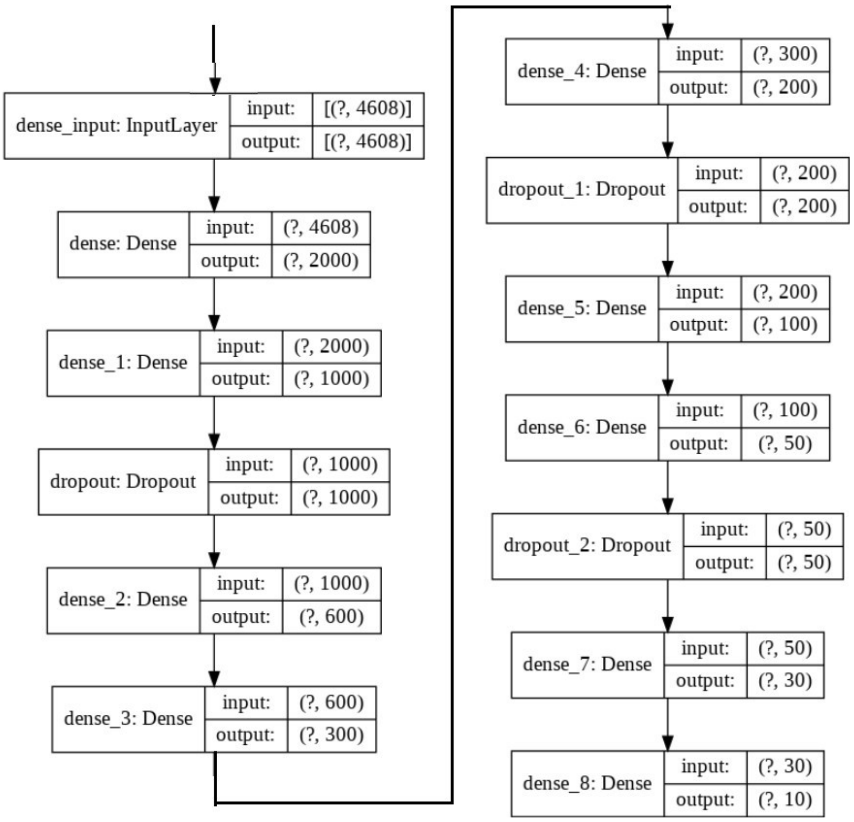
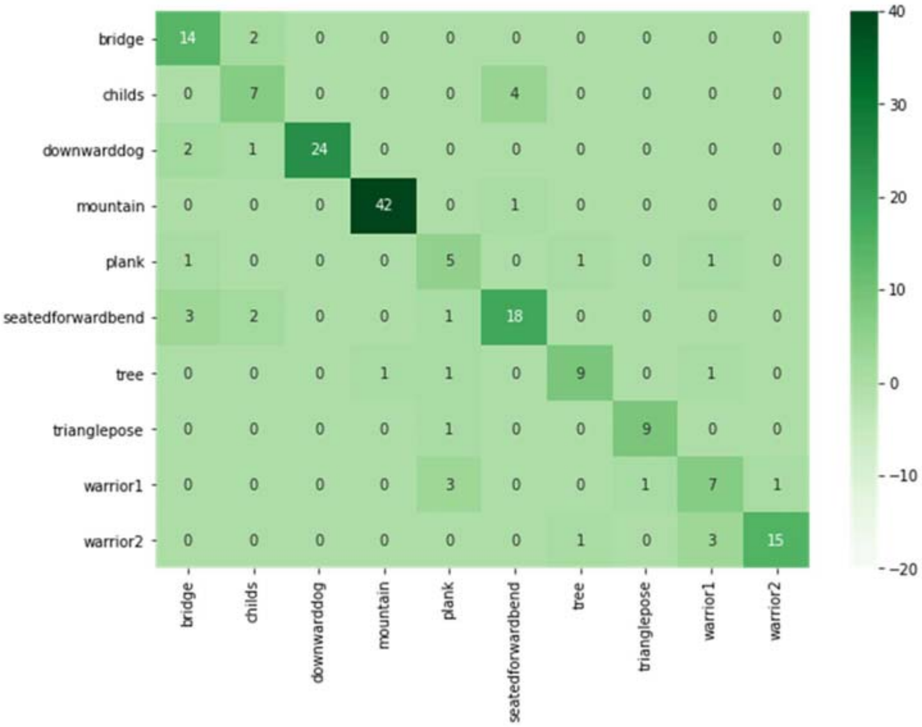


Figure 4.1. Layers of the Architecture

NumPy, OpenCV, PIL. For the transfer learning part, the VGG16 pre-trained model was used with a pre-trained weights magnet. All the images were resized to 100 x 100 and 4608 features were extracted. The entire dataset is split into training and testing sets. The training set is used to build the classifier model, and the test set is used to evaluate the accuracy of the classifier. On testing, we got 85% accuracy the confusion matrix obtained from the actual test labels and predicted test labels were shown in figure the other metrics such as precision, recall, and fi score were also calculated and it is shown in the table 4. Comparing with the traditional CNN architecture and other machine learning models trained with HoG and Hu Moment the proposed transfer learning architecture is more efficient and outperformed others in terms of accuracy and results are given in Table4.

Models	HoG	Hu Moment	CNN	Transfer Learning
Accuracy	70.3%	73.5%	68.3%	85%

Table 1. Accuracy Comparison



	bridge	childs	downwarddog	mountain	plank	seatedforwardbend	tree	trianglepose	warrior1	warrior2
bridge	14	2	0	0	0	0	0	0	0	0
childs	0	7	0	0	0	4	0	0	0	0
downwarddog	2	1	24	0	0	0	0	0	0	0
mountain	0	0	0	42	0	1	0	0	0	0
plank	1	0	0	0	5	0	1	0	1	0
seatedforwardbend	3	2	0	0	1	18	0	0	0	0
tree	0	0	0	1	1	0	9	0	1	0
trianglepose	0	0	0	0	1	0	0	9	0	0
warrior1	0	0	0	0	3	0	0	1	7	1
warrior2	0	0	0	0	0	0	1	0	3	15

Figure 4.2. Confusion Matrix

5. Conclusion and Future Scope

The advancements and innovations in the field of science and technology pave the path to explore different possibilities in multidisciplinary domains. Newest technologies such as artificial intelligence, machine learning, and computer vision are used to implement many real-time applications in our everyday life. Yoga is one of the widely accepted life practices to nurture the body as well as mind. As a seed to develop a system that can assist humans to practice yoga with a virtual trainer, we have proposed an automatic yoga posture identification system from images or videos. While building this system we have experimented with the state-of-art methods in image classification like CNN, but as the amount of data in the dataset is less those methods did not work well. To get better results we have used transfer learning with VGG16 architecture and pretrained ImageNet weights along with a DNN classifier. The results were quite promising; it gave 82% prediction accuracy. The domain still has numerous possibilities to explore. Apart from the images, video analysis can be done to analyze the movement of the yoga asanas for validating the correctness of the movements. The architectures like 3DCNN, Deep - Pose Estimators, LSTM, GRUs are well suited for video-based analysis.

6. References

[1] Catherine Woodyard 2011 International Journal of Yoga 4 49–54

[2] Ross A, Touchton-Leonard K, Yang L and Wallen G 2016 International Journal of Yoga Therapy ISSN 1531-2054

[3] Giacomucci A 2019 Yoga E-learning platform: Practice with frequency and motivation — UX Case Study

[4] Prasanna Mani,Arunkumar Thangavelu A S and Chaudhari N 2017 International Journal of Intelligent Engineering and Systems 10(3):85-93

[5] Rafi U, Kostrikov I, Gall J and Leibe B 2016 British Machine Vision Conference 2016, BMVC 2016 2016-Septe 109.1–109.11

[6] Yadav S K, Singh A, Gupta A and Raheja J L 2019 Neural Computing and Applications 31 9349–9361 ISSN 14333058

[7] Shruti Kothari 2020 Yoga Pose Classification Using Deep Learning Ph.D. thesis SAN JOSE STATE UNIVERSITY

[8] Chen H T, He Y Z, Chou C L, Lee S Y, Lin B S P and Yu J Y 2013 Electronic Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2013

[9] Trejo E W and Yuan P 2018 2018 2nd International Conference on Robotics and Automation Sciences, ICRAS 2018 12–17

[10] Deepak Kumar

[11] Jain S, Rustagi A, Saurav S, Saini R and Singh S 2020 Neural Computing and Applications ISSN 14333058

[12] Yadav S K Yadav, S. K. (n.d.). Real–time Yoga recognition us

[13] Shailesh S and Judy M V 2020 Indian Journal of Computer Science and Engineering 11 89–98 ISSN 22313850

[14] Shailesh S and Judy M V 2020 Intelligent Decision Technologies 14 119–132 ISSN 18758843

[15] Dantone M, Gall J, Leistner C and Van Gool L 2013 Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp 3041–3048 ISSN 10636919

[16] Islam M U, Mahmud H, Bin Ashraf F, Hossain I and Hasan M K 2018 5th IEEE Region 10 Humanitarian Technology Conference 2017, R10-HTC 2017 2018-Janua 668–673

[17] Haque S, Rabby A S A, Laboni M A, Neehal N and Hossain S A 2019 Communications in Computer and Information Science vol 1035 (Springer Verlag) pp 186–193 ISBN 9789811391804 ISSN 18650937

[18] Kreiss S, Bertoni L and Alahi A 2019 Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol 2019-June (IEEE Computer Society) pp 11969–11978 ISBN 9781728132938 ISSN 10636919 (Preprint 1903.06593)

[19] Mehta D, Sotnychenko O, Mueller F, Xu W, Elgharib M, Fua P, Seidel H P, Rhodin H, Pons-Moll G and Theobalt C 2019 CVPR (Preprint 1907.00837)

[20] Lv D, Huang Z, Sun L, Yu N and Wu J 2017 Multimedia Tools and Applications 76 1313–1330 ISSN 15737721

[21] Indolia S, Goswami A K, Mishra S P and Asopa P 2018 Procedia Computer Science 132 679–688 ISSN 18770509

[22] Pratt L and Jennings B 1996 Connection Science 8 163–184 ISSN 09540091

[23] Lin Y P and Jung T P 2017 Frontiers in Human Neuroscience 11 ISSN 16625161

[24] Simonyan K and Zisserman A 2015 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings