# VIRTUAL YOGA INSTRUCTOR USING DEEP CONVOLUTIONAL NEURAL NETWORK

**D.K. Neha Gandhi, R. Sireesha, K. Ramya Sudha, K.Manjunath, M.Suhail**

***Abstract:****The humanbody being the most complex biochemical factory, learning how to engineer it helps enter a sense of modalities at will. The world is caught up in a rat race and keeping oneself fit has not been given much thought that it deserves.This paper's main objective is to motivate society to practice yoga by providing a comfortable environment by building a virtual yoga instructor. With pose estimation, smart assistance has been built to guide and instruct the users throughout their practice period to attain the correct posture. This helps users experience a hassle-free environment and enhances rendition.This can also be used to reduce stress, anxiety, and tension through a series of guided meditative poses. This solution includes human pose estimation using TensorFlow and OpenCV frameworks. Deep Convolutional Neural Network has been used as the architecture, various data collectionstrategiesand processing tasks have been performed and explained in this paper.This concept is simple yet effective as it is cost-efficient and very simple to implement where minimum hardware specifications are required.*

***Keywords: Pose estimation, TensorFlow, OpenCV, Virtual Yoga Instructor, Web Scraping, Data Processing***

## I. INTRODUCTION

The influx of technological advancements provides a chance for Scientists and researchers to explore new ideas for solving day to day problems of the world. Physical fitness plays a very important role in shaping the mind and body to lead a positive way of life. Through the art of yoga, one can achieve inner peace. People in their busy lives find it difficult to visit different yoga studios and places to get the right platform. Various applications include video surveillance, assisted living, advanced driver assistance systems (ADAS) and sports analysis are devised but none of them proved to be effective. The existing yoga practices involve either high-cost machinery or complex software. In order to provide a more optimized experience, yoga is given more significance rather than involving unneeded other fitness exercises.

This paper includes a meticulously generated model that predicts yoga postures and helps an individual to heed the right stance. Computer vision and neural networks are integrated to build a virtual yoga instructor. Section 2 includes related work that discusses views and different takes on the above problem which have beenreferred to, for better approaches and new ideas. Section 3 deals with experimental setup. Section 4 includes the methodology for yoga pose detection mechanism. The results obtained are covered in section 5 and finally the results are concluded in section 6.Images from google and custom images are fed into the model. Better accuracy is rendered through processing the data and converting them into black and white images. Image segmentation through edge detection helps to extract image data and the data obtained is normalized. Neural networks need a lot of data for precise and error-free results. Thus data augmentation aids in removing this complexity. Though model1 has good training accuracy overfitting of training set leads to poor testing accuracy. Thus necessary improvements are provided. Hidden layer count is increased for a more desired accuracy. Model 5 yields needed results with an accuracy of more than 70%. The system built is cost-effective and highly efficient using a methodical approach towards solving the problem. The software by using computer vision and neural networks aids in assisting humans by pose detection. A machine with a visual display unit and camera suffices working on this software.

## II. RELATED WORK

With the introduction of DeepPose by Toshev et al, research on human pose estimation began to shift from classic approaches to Deep Learning. Most of the recent pose estimation systems have universally adopted convolutional networks as their main building block, which has yielded drastic improvements over the existing models.[1]The DeepPose model is taken as an example to create several models of neural networks to make the pose estimation better.

Huang Yi's research on A study on Deep Neural Networks framework[2] described that development of DNN and proved that the deep convolutional neural networks has a better ability to identify scaling, transformation and other forms of distortion-invariant images.

Xin, Mingyuan & Wang, Yongs research on Image classification model describes that the CNN uses feature detection layer to learn from training data implicitly and because of weight sharing mechanism, neurons on the same feature mapping surface have the same weight. The paper also suggested that convolutional networks depth plays a major role in the process of image classification.[3]

Hua-Tsung Chen has proposed a Yoga posture recognition system, which can recognize what Yoga posture the practitioner is performing, and then retrieving Yoga training information from the Internet to remind him/her attention to the posture[4]. The usage of Microsoft Kinect OpenNI library helps in capturing the binary body maps of the user. Body contours are extracted from the body maps which is a good way to represent a position. Star skeleton algorithm is slightly modified for robust results. It connects the centroid and the contours. After the star skeleton computation is

performed, yoga posture isrecognized by defining the normalized feature vectors. This process requires experimental setup cost.

In Yoga posture recognition, by detecting human joint points in real time using Microsoft Kinect [5], Muhammad Usama Islam and co-authorshave proposed a system which monitors body movement and monitor the accuracy of different yoga postures. They have used Microsoft Kinect to detect 10 joint points of the human body in real time and these joint points determine the posture. Reference structure is built for every yoga position by data acquisition and then by comparing it with candidate reference model.

Human posture recognition [6]-[12] is one of the main research areas in computer vision. The reason for its importance is the abundance of applications that can benefit from such a technology.

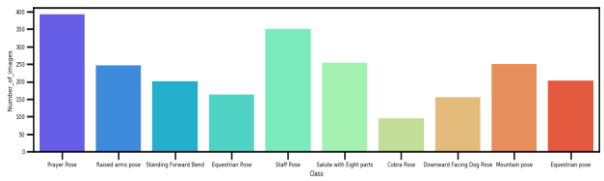
## III. EXPERIMENTAL SETUP

Our experimental setup consists of the latest 8th Gen Intel® Core™ processors and NVIDIA® GeForce® GTX 10-Series graphics, ROG Strix GL503.

The customized data as mentioned in the methodology section was collected on a phone with a camera of a standard 48-megapixel one, a wide-angle one, and a 2X optical zoom one - along with the brand new Qualcomm Snapdragon 855 processor, a 90Hz screen, and frosted glass rear panel.

In our experiment, the models are trained on images of 15 professional yoga instructors and the collected data is visualized in figure 1. On an average, each pose has almost 200 images without augmentation.

Figure 1: Number of images according to the yoga poses are




summarised

## IV. METHODOLOGY

The model is used to detect the pose of a person using the Convolutional neural networks. This paper combines two technologies i.e., Deep learning and computer vision. All the computer vision tasks are performed using the OpenCV image pre-processing library. Then with the help of Kera's a higher-level API of TensorFlow, different models are built, and the results are analysed.The Metrics used to measure the model is accuracy. The Adam optimization algorithm is used in the model which is an extension to SGD that has recently seen broader adoption for deep learningCV applications.[13] As this is a classification task, the loss parameter chosen is Sparse Categorical Cross Entropy. As the generalization error began to degrade, due to the involvement of early stop, it required the training to be stopped at 15$^{th}$ epoch and the results yielded well for that.

The model is as good as the data. In order for our model to be better, the data should be reliable and have minimum skew because it is no use to have a lot of bad data as quality matters too. By having balanced data, the model works well and generalizes. Thus there are three data tasks to be performed prior to training and testing.

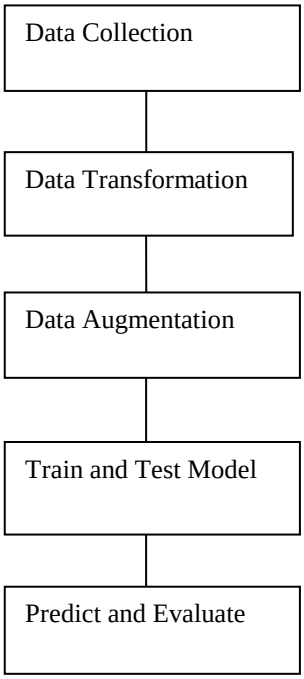



Figure 2: Represents the flow of the process

### A. Data collection

When collecting data for a few classes, unbalanced data is probable. This is avoided by down-sampling and up-weighting as stated in the A Survey on Methods to Handle Imbalance Dataset.[14] Data collection took place in two stages. They are:

**1. Web Scraping**

The required images of respective classes are scrapped from Google Images using beautiful Soup library in Python. There are many unwanted images collected during this process which may occupy 45% of the data. It is difficult to handpick and distinguish the unwanted images from the required images. By using an image classifier built by simple DNN( Deep Neural Network) Machine Learning Architecture, the unwanted images can be separated from the wanted images.

**2. Creation of customized data**

Collecting images of people in different postures is quite demanding. The images should have unbiased data on genders, environments, and backgrounds.

### B. Data transformation

Collected data is processed for error free input and is then fed into the model for accurate results. Transformation is performed prior to training which helps in performing the computation only once and it can look at the entire dataset to determine the transformation. Transformations are mandatory for data compatibility. Quality transformation techniques enable the model to give better results. Data transformation took place in 4 stages.

**1.Resizing input**

Resizing input to a fixed size of 128 X 128. Feed-forward neural networks and DNN have a fixed number of input nodes, so input data must always have the same size. The original image can be seen in figure 3.a.

**2.Grayscale Conversion**

Its computationally cheaper to input grayscale images than RGB/BGR images because the dimensionality decreases by three times while using a grayscale image (that has only one layer). It can be visualized in figure 3.b.

**3.Edge Detection**

Canny edge detection plays an important role in image edge detection.[15] It can be visualized in figure 3.c.

**4.Normalization**

Transforming numeric data to the same scale. Converting the scale of 0-255 to 0-1. Normalization can be performed by dividing the entire NumPy array by 255.[16]

**5.Data split**

After data processing, the whole data is split into 60% Training sets, 20% Testing sets, and 20% Validations sets.

Figure 3.aOriginal image

Figure 3.bGrayscale image

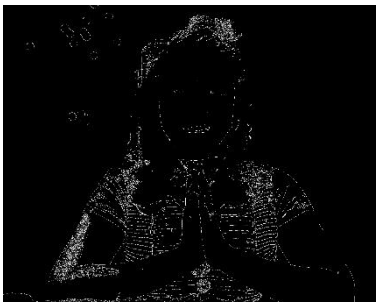


Figure 3.c Canny EdgeDetection

### C. Data augmentation

Networks are heavily reliant on big data to avoid overfitting.[17] As there is no more data available to avoid overfitting data augmentation is followed in this computer vision application. Data augmentation improved the performance of the model and helped in avoiding overfitting. Generally, Overfitting occurs when a model memorizes the data with very high variance such as to perfectly model the training data.[18]

In our model, few of the parameters are taken to rotate, resize, shear the data.

The width shift range of 0.2 is performed to shift the images by moving the pixels horizontally by keeping the same image dimensions. The height shift range of 0.2 is performed to shift image pixels vertically. Through this, some of the pixels are changed to new pixel values. The shear range of 0.2 is set to fix one axis and stretch the image at a certain angle. The images are flipped horizontally means reversing the rows or columns of pixels. The flip augmentation is specified by a Boolean horizontal flipwhich is True. The brightness of the image can be augmented by either randomly darkening images, brightening images, or both.The intent is to allow a model to generalize across images trained on different lighting levels.

### D. Training and testing

As there are 12 poses in Surya-Namaskars, there should be 12 classes. There are two repeating poses and thus to avoid redundancy,10 classes are considered and labelled accordingly. Usage of different models.

The Metric used to measure the model is accuracy. The Adam optimization algorithm is used in the model which is an extension to SGD that has recently seen broader adoption for deep learningCV applications.[19] As this is a classification task, the loss parameter is Sparse Categorical Cross Entropy. There was an early stop and the results yielded well for epochs at 15.

**Algorithm: The process to build a virtual Yoga Instructor**

```
class PreProcessing:
     def resize(image):
          #resize the image which fits to
          the neural network
     def grayscale(image):
          #convert the RGB image to
          grayScale image
     def edgeDetection(image):
          #Use canny edge detection to
          detect the edges in the image
     def augmentation(image_path):
          # As there is no more data
          available to avoid overfitting
          data augmentation is followed

class ModelTraining:
     def models(trainSet,testSet):
          #use of different models
     def split(data):
          #data is split into 80%
          training and 20% testing
          datasets
          return trainSet, testSet
     def evaluate(model):
          #measure the accuracy

process = PreProcessing( )
MT = ModelTraining( )

#After the data collection phrase, store
the path of images image_path
```

```
for image in image_path:
     process.resize(image)
     process.grayscale(image)
     process.edgeDetection(image)

     #Let all the images are stored in a
     variable called data

#normalize the data
data=data/255.0
process.augmentation(data)
trainSet,testSet = MT.split(data)
#use of model() and evaluate() yields
the results of the performance of the
models
```

## V. RESULTS

Model 1 has the simple ANN architecture (Artificial Neural Network). From table 1 it is evident that data is overfitting the training set. When three more dense layers are added to the Model 1, Model 2 is still observed to overfit the training data. Thus CNN(Convolutional Neural Network) architecture is used for better results. In Model 3, three Convolution layers along with three Max Pooling layers are added to Model 1 architecture but these changes did not yield any better results than the other models. Model 4 is built by combination of Model 2 and Model 3. Although the train accuracy is observed to perform better than the other models, the test accuracy is not improved when compared to the other models. Model 5 is a slight improvement of Model 4, addition of dense layers eliminated overfitting and improved train and test accuracy.

Table 1: The train-set accuracy and the test-set accuracy of different models

| Accuracy | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Train | 96% | 94% | 88% | 93% | 84% |
| Test | 61% | 62% | 63% | 60% | 71% |

## VI. CONCLUSION AND FUTURE SCOPE

In this paper, five different models have been applied to the data and the results are analysed. During the image pre-processing tasks it has been observed that image conversion to grayscale and image resizing steps have consumed low training time and also made the application computationally inexpensive. By imagedata augmentation, it can be concluded thatit has resulted in building an accurate model with good performance.Availing this paper significantly reduces the time and cost of getting to a yoga studio. Different deep convolutional architectures have been explored for the development of Virtual yoga instructor. Based on the experiments it is evident that model 5 provides the best accuracy. This can be further extended to develop a real time detection on various video frames.

## REFERENCES


[1]A. Toshev and C. Szegedy, "*DeepPose: Human Pose Estimation via Deep Neural Networks*," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1653-1660.

[2]Huang Yi, Sun Shiyu, Duan Xiusheng and Chen Zhigang, "*A study on Deep Neural Networks framework*," 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, 2016, pp. 1519-1522.

[3]Xin, Mingyuan & Wang, Yong. (2019). "*Research on image classification model based on deep convolutional neural network*". EURASIP Journal on Image and Video Processing. 2019. 10.1186/s13640-019-0417-8.

[4]Chen HT., He YZ., Hsu CC., Chou CL., Lee SY., Lin BS.P. (2014)" *Yoga Posture Recognition for Self-training*". In: Gurrin C., Hopfgartner F., Hurst W., Johansen H., Lee H., O'Connor N. (eds) MultiMedia Modeling. MMM 2014. Lecture Notes in Computer Science, vol 8325. Springer, Cham

[5]M. U. Islam, H. Mahmud, F. B. Ashraf, I. Hossain and M. K. Hasan, "Y*oga posture recognition by detecting human joint points in real time using microsoft kinect*," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, 2017, pp. 668-673.

[6] O. Patsadu, C. Nukoolkit, and B. Watanapa, "*Human gesture recognition using kinect camera*," in Computer Science and Software Engineering(JCSSE), 2012 International Joint Conference on.IEEE, 2012, pp.28–32.

[7] S. Obdr˘z´alek, G. Kurillo, J. Han, T. Abresch, R. Bajcsyet al., "*Real-time human pose detection and tracking for tele-rehabilitation in virtual reality,*"Studies in health technology and informatics, vol. 173, pp. 320–324, 2012.

[8] K. Rector, C. L. Bennett, and J. A. Kientz, "*Eyes-free yoga: annexure game using depth cameras for blind & low vision exercise*," in Proceedings of the 15th International ACM SIGACCESS Conferenceon Computers and Accessibility. ACM, 2013, p. 12.

[9] T.-L. Le, M.-Q. Nguyenet al., "*Human posture recognition using human skeleton provided by kinect,*"in Computing, Management andTelecommunications (ComManTel), 2013 International Conference on.IEEE, 2013, pp. 340–345.

[10] Z. Zhang, Y. Liu, A. Li, and M. Wang, "*A novel method for user-defined human posture recognition using kinect,*" in Image and SignalProcessing (CISP), 2014 7th International Congress on. IEEE, 2014,pp. 736–740.

[11] J. Charles and M. Everingham, "*Learning shape models for monocular human pose estimation from the microsoft xbox kinect,*"in ComputerVision Workshops (ICCV Workshops), 2011 IEEE International Confer-ence on. IEEE, 2011, pp. 1202–1208.

[12] F. Pedersoli, S. Benini, N. Adami, and R. Leonardi, "*Xkin: an open source framework for hand pose and gesture recognition using kinect,*"The Visual Computer, vol. 30, no. 10, pp.
1107–1122, 2014.

[13]Kingma, Diederik & Ba, Jimmy. (2014). "*Adam: A Method for Stochastic Optimization. International Conference on Learning Representations*,"Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

[14]Sonak, Apurva and R. A. Patankar. "*A Survey on Methods to Handle Imbalance Dataset*." (2015).IJCSMC, Vol. 4, Issue. 11, November 2015,International Journal of Computer Science and Mobile Computing, pp.338 – 343

[15]Li J., Ding S. (2011) "*A Research on Improved Canny Edge Detection Algorithm. In: Zhang J. (eds),*" Applied Informatics and Communication. ICAIC 2011. Communications in Computer and Information Science, vol 228. Springer, Berlin, Heidelberg

[16]S. van der Walt, S. C. Colbert and G. Varoquaux, "*The NumPy Array: A Structure for Efficient Numerical Computation*," in Computing in Science & Engineering, vol. 13, no. 2, pp. 22-30.

[17]Shorten, Connor and Taghi M. Khoshgoftaar. "*A survey on Image Data Augmentation for Deep Learning.*" Journal of Big Data 6 (2019),pp. 1-48.

*[18]* Allamy, Haider. (2014).*" Methods to avoid over-fitting and under-fitting in supervised machine learning(comparative study)."*

[19]Robbins, Herbert; Monro, Sutton. "*A Stochastic Approximation Method*". Ann. Math. Statist. 22 (1951), no. 3,pp. 400--407.

## AUTHORS PROFILE

 **Sireesha Rodda** is currently working as Professor at the department of Computer Science and Engineering, at GITAM(Deemed to be) University. She has completed her Ph.D. Her research project is Network Intrusion Detection System for Sensitive Data funded by Science and Engineering Research Board(SERB).She has 15 publications in the field of Database systems and Wireless Sensor Networks.

 **D.K. Neha Gandhi** is currently pursuing Bachelor of Technology from GITAM(Deemed to be ) University,Andhra Pradesh,India. She has good experience in the field of machine learning and computer vision. She has 3 months industrial experience in the field of computer vision. Her research interests include Deep Learning and Computer Vision.

 **K. Ramya Sudha** is currently pursuing Bachelor of Technology from GITAM(Deemed to be)University,Andhra Pradesh,India. She has experience in the field of computer vision. She has 2 months industrial experience in the field of computer vision. Her research interests include Computer Vision.

 **Manjunath Kurapati** is currently pursuing Bachelor of Technology from Gitam University, Andhra Pradesh, India. He has 2 months of industrial experience in the field of RPA and his main research work focuses on machine learning.

 **Mohammad suhail** is currently pursuing bachelor of technology from gitam University,Andhra Pradesh I am good with writing content for the project and I have 2 month of experience working with the tech firm