ICCIC 2020

Computational Evaluation of Persons' Torso and Waist Size using PoseNet Neural Network.

Ashutosh Joshi[1], Anshul Singh[1], Arham Ali[1] , Ruchi Patel[1]

[1] Medi-Caps University, Indore

Abstract

There is one thing that everyone likes to do and that is shopping, purchasing appliances, attire, accessories, and many more. But the process of shopping from malls requires a lot of time for the consumer. Especially shopping for the wardrobe, the customer starts with selecting dresses and then try it after this he/she evaluates the perfect size for him/her and eventually again try that piece of an outfit. This whole process consumes a huge amount of time of the customer and cumulatively somehow abate the sales of the stores. This paper illustrates a Computer Vision model which can ease the shopping experience and can raise the business of the clothing stores and malls. The model uses an elementary approach to appraising the body sizes of the customers. The model firstly detects the coordinates of the person who comes in front of the camera and then using those coordinates our proposed algorithm determines the torso and waist size of that person. [4] These reported sizes also suggest the customer proceed to the fitting area of attire. The main goal of this model is to lessen the time a customer spends in the shopping mall and upsurge the sales of the stores. Besides this application, there are several other areas where we can use this model. Like tracking the human movement, estimating the structure of the human body in darkness, and many more.

Keywords: *Computer Vision, Evaluating algorithm, PoseNet, shopping, time preserving*

1. Introduction

One of the founders of America, Lt. Benjamin Franklin once wrote in his book that, "Time is Money" taking his words into the mind the idea of this model happened and then we started working on an idea which can lead to multiple advantages. [1] The idea is to build a system that can be easily embraced in the industry and ease the time-consuming tasks of people. In recent years, the lifestyle of people majorly depends upon the clothes and outfits of the person. The vast majority of the population invests their time to style themselves with the best quality of wearables and spends a huge amount of time. [5] The primary goal is to propose a modern solution to this problem of spending time in the shopping malls. Integrating artificial intelligence with clothing stores can easily overcome this problem and can also help the retailers to increase their sales to gain maximum profit.

Using artificial intelligence and computer vision to achieve the purpose not only solved the aforementioned problem but also helped in the conventional engagement of employees to every customer. [2] [3] It is very easy to install this model in any store just by updating the corresponding section of clothes in classified with the sizes.

The whole flow of the model is divided into 3 phases :

(a) PoseNet: This phase evaluates the 17 body parts from the frame of the camera and detects the pose of the multiple bodies.

(b) [2] Extracting coordinates(JSON): In this phase, the model elicits the body(s) coordinates into the JSON format and saves these as a 'keypoints.json' file.

(c) [2] Evaluating sizes: Ultimately, using 'keypoints.json' we select useful features and coordinates to evaluate the torso and waist size of the entering person.

Using human recognition and the PoseNet neural network weights we have had completed this project.

This paper is organised as follows. Section 2 introduces the main ideas and methods of the model. Section 3 sums the algorithms used in the project. Section 4 expresses the current research status in the similar field. Section 5 explains the application of this model in the industry. Section 6 explains the results and discussion of the work and lastly Section 7 concludes the paper and predicts the future of this model in the field.

2. Overview of Model

Under this heading the preparation of the research and development is discussed and explained following considered measures to achieve the final goal of the project.

2.1 Frame Analysis

First and foremost, [10]the model captures the video footage from the webcam using OpenCV{video.capture(-1)}. Then the real-time video is sent to the next level of the model to evaluate the coordinates to the human body pose. The video plays in the web portal/ localhost with a port assigned using Node.JS{yarn watch}. [3] Simultaneously the live video is analysed by the weights of the poseNet neural network.

Figure 1 describes the input to the model in the form of a video and rest of the architecture. This video is broken down into frames and then each frame into a mathematical matrix to proceed in the model.

2.2 Coordinate Evaluation

PoseNet is a modern machine learning and deep learning model that allows for real-time pose estimation of the human body. [1] The poseNet is trained on the millions of images and used to estimate 17 parts/keypoints of the human body that is the nose, leftEye, rightEye, leftEar, rightEar, leftShoulder, rightShoulder, leftElbow, rightElbow, leftWrist, rightWrist, leftHip, rightHip, leftKnee, rightKnee, leftAnkle, and rightAnkle.

[11] The video of particular width x height is minced into output stride of size width resolution x height resolution x number of keypoints and then into a heat-map. [1] The actuarial of the weights and video extrapolates the coordinates of above-mentioned keypoints. In Figure 2, the flow of this paper's project is explained in a sequential manner and Figure 3 showcases the coordinates are selected by the poseNet architecture and marked over the frame.

Below the heading the sequence diagram of whole process the drafted to give a lucid understanding about the processes.

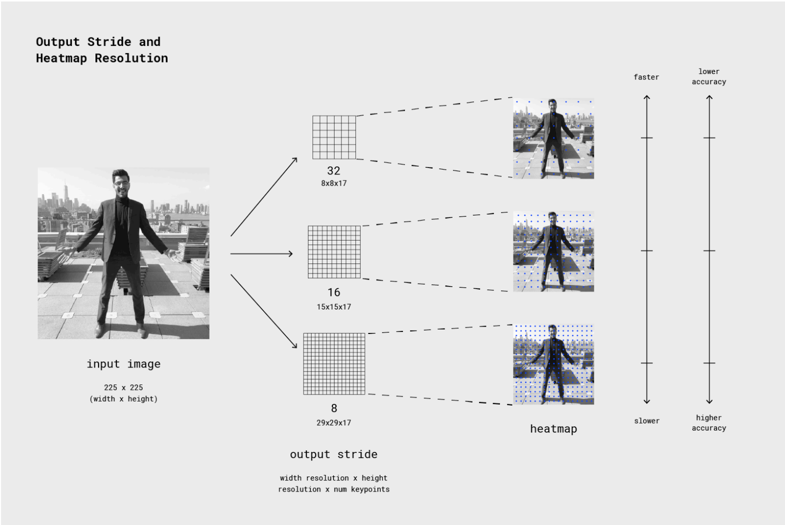

Fig. 1. Flow of PoseNet Model [3]

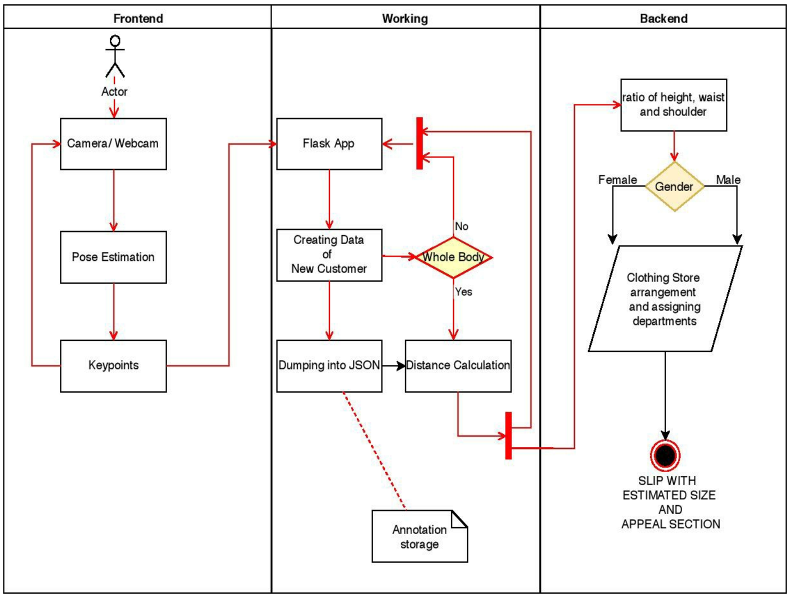

Fig. 2. Sequence diagram of the proposed project

2.3 Dumping into JSON

After getting the coordinates of the human body in the captured real-time video, the coordinates are transferred to the flask app using the port extension of the localhost. [9] Once the live coordinates are sent to the python flask app, we conclude the steam of coordinates into the [4]JSON format{dump.json()} and used the json format as a DataFrame to evaluate the body sizes of the human.

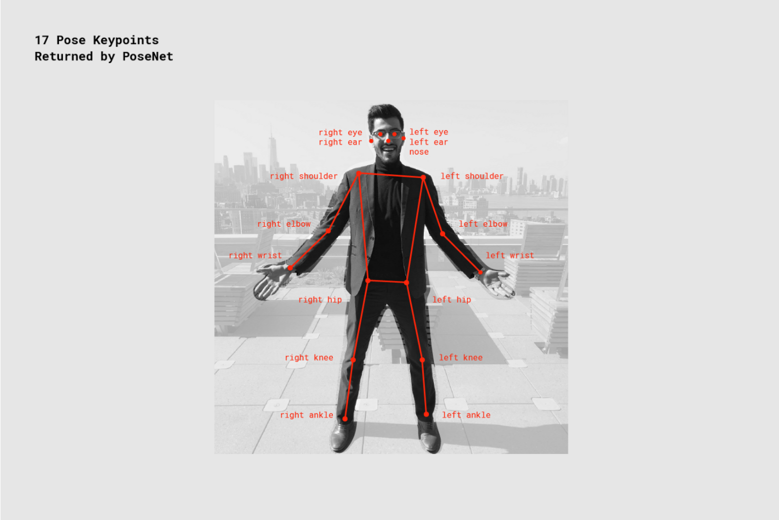

Fig. 3. Coordinates selection by PoseNet and key points. [3] [4]

2.4 Evaluating Algorithm

[4] 'keypoints.json' is the file that stores the x and y coordinates of the human body. importing this json file to the algorithm, the in-house algorithm of the model detects the torso and waist size of the individual entering into the shopping mall. Majorly the algorithm uses a few important keypoints like shoulder, waist, and ankle to evaluate the body sizes of the customer.

3. Algorithm and Methodology

Building an algorithm and method was the most excellent task during the research. It took brainstorming of all the researchers of this paper to build a technique that can estimate the sizes of a human body using only steam of numbers ranging between (0,0) to (1920,1200). [6][8] Out of 17 key points, we majorly worked on Shoulders, Waist, and Ankle to evaluate the sizes.[1]

The method is as follow:-

lets LS and RS are points of left shoulder and right shoulder respectively, and similarly, LW and RW are points of the left waist and right waist respectively and LA and RA are corresponding points for ankles.[2]
The initial stage of the algorithm evaluates a raw height using the distance formula.

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

The different lengths we used in the estimation of the sizes are length between shoulder to ankle that is L_SA, a length between shoulder to the waist that is L_SW and a length between the waist and the ankle that is L_WA.

Along with these lengths we also measured some width of the human body like W_S: width between the left and right shoulder and W_W: width between the left and right waist. [2]

Now the final stage of the algorithm is to compute the actual size of the human body.
We classified the sizes in the following manner:

Torso size :

- XS or less = Extra Small
- S = Small
- M = Medium
- L = Large
- XL or more = Extra Large

Waist Size:

- 28 or less
- 30
- 32
- 34
- 36
- 38 or more

Now the using the above-measured values we build an evaluating variable (OHL) which is equal to the ratio of L_SA and addendum of L_SW and L_WA.

$$ohl = \frac{L_S A}{L_S W + L_W A} \tag{2}$$

Now the torso and waist sizes are calculated by using this variable ʒ. Torso variable is Ť(*T*) and for waist it is ɰ(*w*). The formulae to evaluate these two are:

$$T = ohl \times \frac{W_S + W_W}{W_W} \tag{3}$$

$$w = ohl \times \frac{W_W}{W_S} \tag{4}$$

The ranges of ɰ and Ť lies between 0 - 2.5 and 0 - 4 respectively. Figure 4 gives the lucid detail of measurements we are taking.

Using the final value of Ť and ɰ the model reports the size of the person entering into.

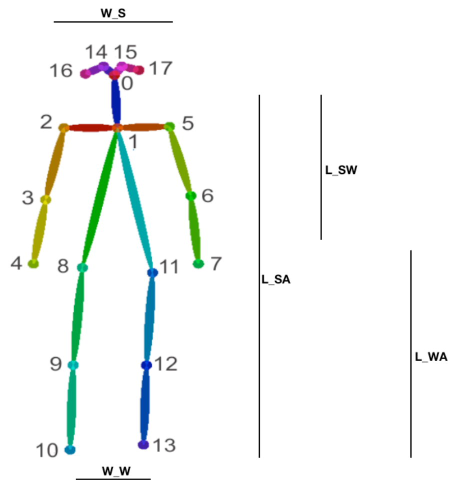

Fig. 4. Useful parameter for evaluation

4. RESEARCH STATUS

The situation of this model is under testing, the accuracy of the project using poseNet neural network is pretty good and the project is ready to market.[5][13] The team is still working on the 2nd version of the system to detect multiple persons more accurately and assign an ID to each individual. Nonetheless, the accuracy and efficiency of the model are better than expected.

5. APPLICATION

Due to modern updation in the working industry, the demand for artificial intelligence is in a rush and the people are adopting this change very smoothly and efficiently. The application of the model is relating in the field of A.I. and M.L. and it has a great impact on the clothing and shopping industry to increase the sales and productivity of the store.

The other applications of this model could be:-

- Smart Fashion Ramp walk.
- Detection of the Human in darkness or behind a wall.
- Size adjustment for labours.
- Kinetic sculpting art.

and many more.

6. RESULT AND DISCUSSION

The outcome of this developed system is the sizes of the human body with sufficient accuracy and abrupter latency. [10] The evaluation of result is clearly explained using figure 5 below.Besides the latest use of this system in the shopping malls, currently, the system has few drawbacks too.[14]

1. It also detects the human body is the inclined angle which leads to giving the wrong measurement of coordinates.
2. Sometimes the customer wears an article of heavy clothing which also reports the faulty measurements of coordinates.

However, these drawbacks are less in the model, the team is advancing this system to give the best and more accurate results to increase the accrue of this system.

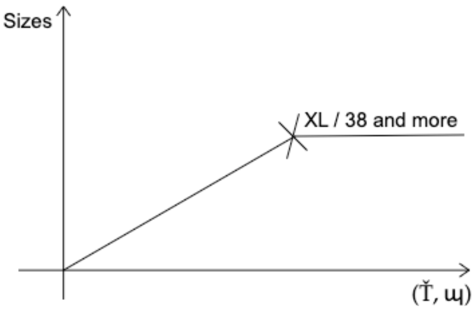

Fig. 5. Result of sizes based on aforementioned evaluation

To compare our proposed research with previous researches, there is a tabular representation which lucidly explains the performance and accuracy of the final product. Fundamentally, the project is working on ResNet-152 and WrnchAI architecture which cumulatively is faster than other heavy networks present.

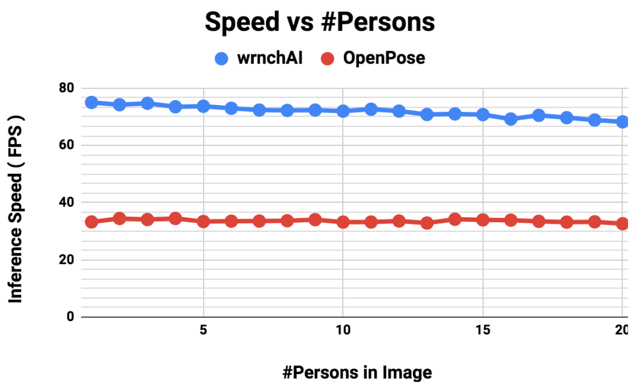

Fig. 6. Comparison between wrnchAI(used) and OpenPose models on the basis of number of persons detected, inference speed and accuracy. [15]

As an epilogue of the result and discussion Fig. 7 numerate the final accuracy of our proposed model using a Residual network.[16]

Method	Backbone	Input size	AP	AP(50)	AP(75)	AP(m)	AP(l)	AR
CMU-Pose	-	-	61.8	84.9	67.5	57.1	68.2	66.5
Mask-RCNN	ResNet-50-FPN	-	63.1	87.3	68.7	57.8	71.4	-
G-RMI	ResNet-101	353x257	64.9	85.5	71.3	62.3	70.0	69.7
CPN	ResNet-Inception	384x288	72.1	91.4	80.0	68.7	77.2	78.5
Bangbangren		-	72.8	89.4	79.6	68.5	80.0	78.7
CPN+	ResNet-101	384x288	73.0	91.7	80.9	69.5	78.1	79.0
Ours	ResNet-152	384x288	73.7	91.9	81.1	70.3	80.0	79.0

Fig. 7. Final summary table of accuracy report of ours model.

7. CONCLUSION

Due to the alarming demand for modern technology, this system could be a revolutionary environment for the shopping experience of the customer. [12] We strongly believe that adopting our research in the industry can increase the businesses and efficiency of the clothing stores.

Broadly, the model is helping the customers to lessen the shopping time to a minimum and also helping the store owners to improve purchases.

I, as the first author of this research, urge the store owner and technical people in the industry to try this approach in their respective stores and help our research to progress in the industry. I would like to achieve my work by quoting an adage that "*Modern problems require modern solutions*" by Chappelle's show.

ACKNOWLEDGMENT

First Author (Ashutosh Joshi):

I would like to express my truthful gratitude to my academic institute Medicaps University, Rau for the constant support of my B.Tech study and mentioned research, for the steadiness, motivation, and extensive expertise. Faculties' supervision helped me in making this project flourishing in my competitive career.

References

1. Shanxin Yuan, Guillermo Garcia-Hernando, Björn Stenger, Gyeongsik Moon, Ju Yong Chang, Kyoung Mu Lee, Pavlo Molchanov, Jan Kautz, Sina Honari, Liuhao Ge, Junsong Yuan, Xinghao Chen, Guijin Wang, Fan Yang, Kai Akiyama, Yang Wu, Qingfu Wan, Meysam Madadi, Sergio Escalera, Shile Li, Dongheui Lee, Iason Oikonomidis, Antonis Argyros, Tae-Kyun Kim : Depth-based 3d hand pose estimation: From current achievements to future goals
2. Pradnya Krishnanath Borkar , Marilyn Mathew Pulinthitha , Mrs. Ashwini Pansare, 2019, Match Pose – A System for Comparing Poses, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 10 (October 2019)
3. https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5
4. https://github.com/tensorflow/tfjs-models/tree/master/posenet
5. https://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html
6. Nozha Jlidi, Ahmed Snoun, Tahani Bouchrika, Olfa Jemai, Mourad Zaied PTLHAR: PoseNet and transfer learning for human activities recognition based on body articulations(2020)
7. George Papandreou, Tyler Zhu, Liang-Chieh Chen, Spyros Gidaris, Jonathan Tompson, Kevin Murphy PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model(2017)
8. Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y.: Fully convolutional instance-aware semantic segmentation. In: CVPR. (2017)
9. X. Qian, Y. Fu, T. Xiang, W. Wang, J.Qiu, Y. Wu, Y.-G. Jiang, and X. Xue,Pose-normalized image generation for person reidentification, in ECCV, 2018.
10. Yang, Y., Ramanan, D.: Articulated pose estimation with flexible mixtures of parts. In: CVPR. (2011)
11. Dantone, M., Gall, J., Leistner, C., Gool., L.V.: Human pose estimation using body parts dependent joint regressors. In: CVPR. (2013)
12. Yasunori Kudo, Keisuke Ogaki, Yusuke Matsui, and Yuri Odagiri Unsupervised Adversarial Learning of 3D Human Pose from 2D Joint Locations(2015)
13. Alex Kendall, Matthew Grimes, Roberto Cipolla PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization(2015)
14. Soroush Seifi, Tinne Tuytelaars How to Improve CNN-Based 6-DoF Camera Pose Estimation(2019)
15. https://www.learnopencv.com/pose-detection-comparison-wrnchai-vs-openpose/
16. https://nanonets.com/blog/human-pose-estimation-2d-guide/