OUTLINING OF CLOTHES USING POSENET POINTS

P.Laharika[1], T. Niharika[2], Sk. Md.Saad[3], Md. Irfan[4], Mrs.S.Naga Sindhu[5]

[1,2,3,4]Bachelor of Technology, Dept. of Computer Science and Engineering, Dhanekula Institute of Engineering and Technology, Andhra Pradesh, India

[5]Assistant Professor, Dept of Computer Science and Engineering, Dhanekula Institute of Engineering and Technology, Andhra Pradesh, India

Abstract: Even in the pandemic situations, online shopping has become the most appropriate manner for people to buy various products, particularly clothing. However, the main disadvantage of buying clothes online is that we have the benefit of doubt about whether it will be the right choice for our individual personalities based on a variety of factors (i.e., height, colour etc.). To address this issue, we allow buyers to directly inspect the outfit and select their preferred apparel. Posture estimation is a computer vision task that detects an individual's pose, which is defined as the orientation and position of the individual. It operates by detecting a number of keypoints, allowing us to understand the individual's primary components and estimate its present orientation. We will be able to build the shape of the object in 2D based on such keypoints, and it estimates the human position in isolated RGB photos using the pre-trained Tensorflow PoseNet model. The model estimates the position of 17 key points on the human body, such as the elbow, right and left shoulders, left and right hips, and other body parts. We can outline the outfits using this model and the posenet keypoints mappings.

***Keywords*: Posture estimation, Keypoints, TensorFlow, PoseNet model, Outlining Outfit.**

I. INTRODUCTION

With the emergence of e-commerce, there are an escalating number of online stores worldwide. Different services and modes of interaction have been introduced into online fashion websites to increase user excitement and draw their interests. Customers can use the features of online fashion websites to freely preview, vote, leave comments, and share their preferred collections with friends. Furthermore, one of the unique aspects of an online fashion store is the ability to provide consumers with virtual dressing rooms where they may virtually try on as many different clothing items as they want. [1] The objective of the project is to exhibit how a person transforms as they dress.

Virtual dressing based on human pose estimation, like so many other computer vision tasks, is a difficult challenge to solve. The nature of the data involved in training these strategies is the root cause of this inconsistency incomplexity. While generalized individual pose point prediction activities may have reference to dozens of thousands or millions of frames for learning, datasets with/without human labels are always in the thousands or, at best, hundreds of millions of frames. The reason for this is because we need ways to forecast pose points of people in such photographs in order to have labels for them. These are the key motivations for preferring to use convolutional neural networks to achieve a relatively shallow architecture for detecting pose points.

An active image of a human is used as the project's input, which is cut and then fed into the PoseNet architecture, which then integrates and overlays essential elements of the human with dress and display.

II. RELATED WORKS

PoseNet is a deep learning model for estimating human posture in photos or videos. Its model is used to predict the key postures of human body joints. Because stance motions are frequently driven by unique human actions, knowing a human's body pose is essential for action recognition. There has been a rich history of research on 3D skeleton-based action recognition [2][3][4][5] due to the emergence of affordable depth sensors [6][7] like Kinect and its potential for skeletal tracing. We focus on 2D skeleton-based action identification in this study, where human poses are approximated using standard RGB photos rather than depth images. Many applications exist for the Pose Estimation model, including Gesture Control and Action Recognition. We review the performance of contemporary action detection techniques that employ colour photos [8][9] as input to estimate stance.

We propose a system that predicts the pose points of human body parts [10][11][12] in the camera and overlaps the person's selected outfit to pose points using a convolution neural network (CNN) model based on the above existing system using the notion of Human pose estimation.

III. METHODOLOGY

3.1 DEEP LEARNING

The next emergence of deep learning is profound learning. Deep learning models are approximately inspired by the characteristics of social brain data processing. Just as we use our perceptions to look for patterns and classify different types of information, significant learning algorithms for equipment can be taught to perform the same tasks. Our 1st system proposed is developed using the techniques of deep learning. Deep learning is an additional part of machine learning algorithms, and therefore classified as artificial intelligence in wider areas. Deep learning deals with deep neural networks, recurring neural networks [13].

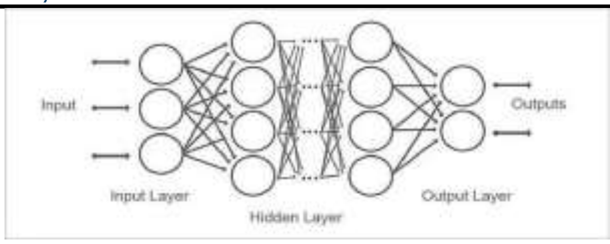


Fig 1: Deep Learning Model

3.2CONVOLUTIONAL NEURAL NETWORK (CNN)

A Deep convolution Naive Bayes (ConvNet/CNN) is an automated system that can consider taking an input image, assign symbolic importance to various aspects/objects of the image (learnable weights and distinctions) and recognise the difference them from each other. In comparison with other feature selection, the well before the processing necessary in a ConvNet is much smaller. While hand-designed filters with only enough training in primitive methods, ConvNet can learn the characteristics of the convolution filters. The ConvNet's model is similar to the connectivity pattern of biological neurons and was captivated by a cerebral cortex organizational structure [14][15][16].

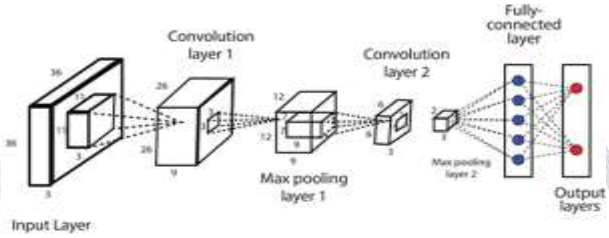


Fig 2: CNN Architecture

3.3 POSENET MODEL

By evaluating the spatial position of the main body ligaments, an ML model can be used to estimate a person's position in an image or video. Pose a question (keypoints). A computer vision technology is used to detect human figures in photos and movies, for example, to determine where the human figure is located [18][20].

An individual's elbow emerges in a shot. It is critical to understand that only the location of the essential body joints is estimated when evaluating posesnet be able to tell who is in a photo or video. TensorFlow asks for frameworks to help with a range of tasks community. These models can assist us in iterating through the design process research [19].

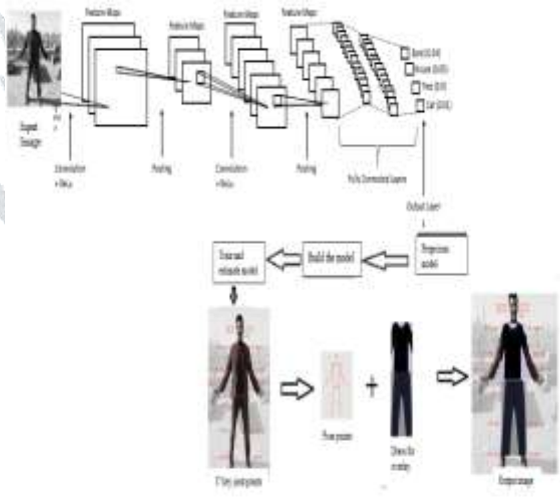


Fig 3: Model Block Diagram

IV. TRAINING AND TESTING

The parameters in all layers are configured using COCO data set's pre-trained models and labels. The test data is compared to the models that have been pre-trained and the pose points that have been generated from the dataset [21].

The pose is represented as a key point with semantic information to reflect the target values for training. The target label vectors for each training image are human parts. There are 17 significant points that have been identified and are indexed by Part ID. 0 denotes the section of the nose, 1 the left eye, 2 the right eye, 3 the left ear, 4 the right ear, 5 the left shoulder, 6 the right shoulder, 7 the left elbow, 8 the right elbow, 9 the left wrist, 10 the right wrist, 11 the left hip, 12 the right hip, 13 the left knee, 14 the right knee, 15 the left ankle, 16 the right ankle [17]

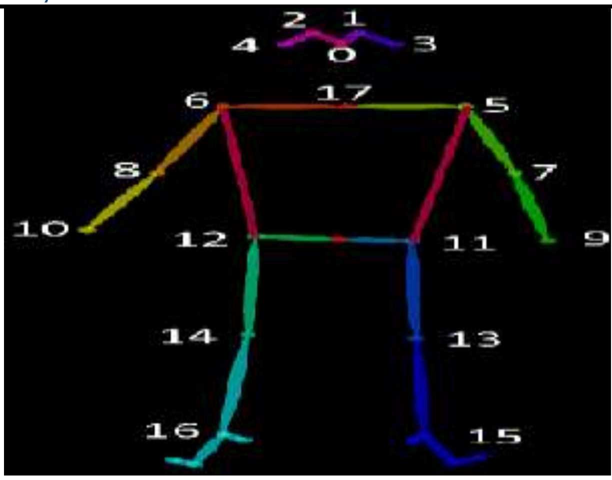


Fig 4: Key points (17) detected in human

V. COCO DATASET

COCO (Common Objects in Context) is a large-scale universal dataset that may be used for a variety of computer vision applications such as object detection, segmentation, and captioning. COCO has a number of characteristics:

- Segmentation of objects
- In-context recognition
- Segmentation of super pixel items
- 330K images (>200K of which are labelled)
- 1.5 million instances of objects
- There are 80 different types of objects.
- There are 91 different types of stuff.
- There are five captions per image.
- 250,000 persons with important information
- Image files and annotation files make up the dataset.

The annotation file is a JSON file that contains all of the information on a particular person (or some other categories). Bounding boxes, area, names of source photos, and other information can be found here.[22]

Method	Backbone	Input Size	AP	AP_{50}	AP_{75}	AP_m	AP_l	AR
CMU-Pose [5]	-	-	61.8	84.9	67.5	57.1	68.2	66.5
Mask-RCNN [12]	ResNet-50-FPN	-	63.1	87.3	68.7	57.8	71.4	-
G-RMI [24]	ResNet-101	353 × 257	64.9	85.5	71.3	62.3	70.0	69.7
CPN [6]	ResNet-Inception	384 × 288	72.1	91.4	80.0	68.7	77.2	78.5
FAIR* [9]	ResNeXt-101-FPN	-	69.2	90.4	77.0	64.9	76.3	75.2
G-RMI* [9]	ResNet-152	353 × 257	71.0	87.9	77.7	69.0	75.2	75.8
oks* [9]	-	-	72.0	90.3	79.7	67.6	78.4	77.1
bangbangren*+ [9]	ResNet-101	-	72.8	89.4	79.6	68.6	**80.0**	78.7
CPN+ [6,9]	ResNet-Inception	384 × 288	73.0	**91.7**	80.9	69.5	78.1	**79.0**
Ours	ResNet-152	384 × 288	**73.7**	**91.9**	**81.1**	**70.3**	**80.0**	**79.0**

Fig 4: Comparisons of COCO dataset

VI. TECHNICAL DETAILS

Detailed technical description of the single-pose estimation algorithm [18][20]:

It's worth noting that the researchers used both a ResNet and a MobileNet model of PoseNet in their research. While the ResNet model is more accurate, its large size and many layers would make page load and inference times less than desirable for any real-time applications. We chose the MobileNet model since it is optimised for mobile devices.

At a high level, the procedure is as follows:

We do a few calculations once the image is run through the model to estimate the pose from the outputs. For example, the single-pose estimation algorithm returns a pose confidence score, which is made up of an array of keypoints (indexed by part ID) each with a confidence score and an x, y position.

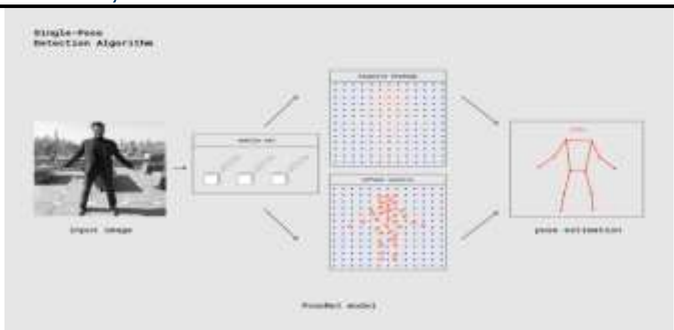


Fig 5: Single pose estimation process

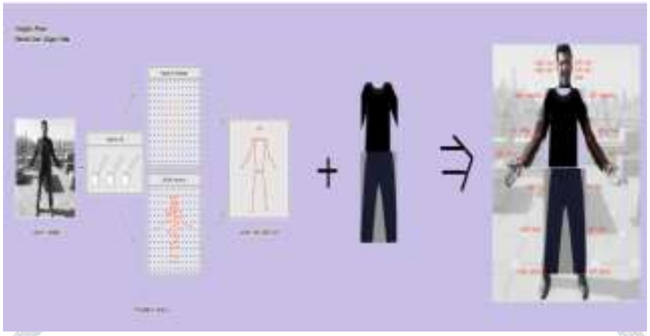
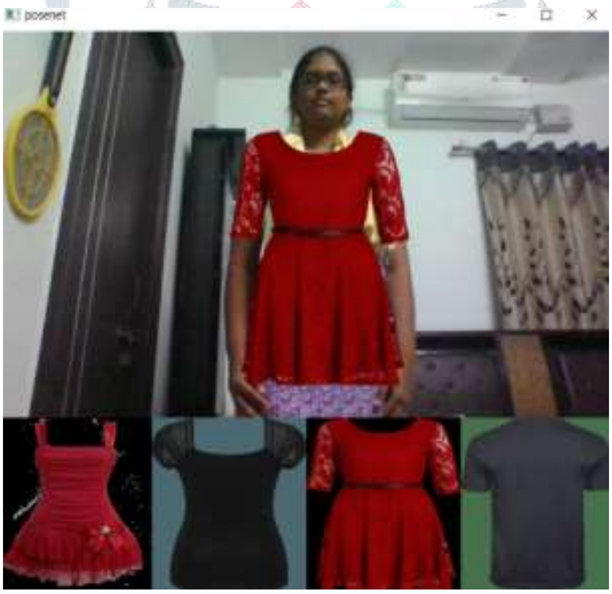


Fig 6: Proposed process to existing process

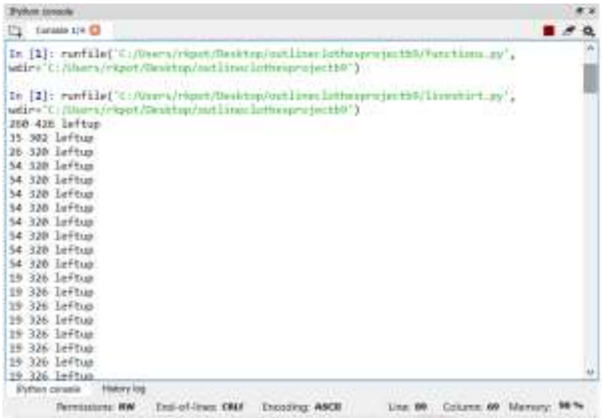
In the input picture space, all keypoint positions contain x and y coordinates and may be mapped directly onto the image. We map our chosen dress onto that image using this keypoint position mapping.

VII. RESULTS

The model shows a webcam capture of a single individual, as well as their choice of clothing morphing on individuals.



As long as the webcam is turned on, the posenet algorithm runs continually. The webcam is reading the x, y, and position adjustments of the dress as follows:



VIII. CONCLUSION

Even when multiple prior methodologies for outlining clothes or overlaying clothing with posture have been developed, these do not produce accurate results. Proposed horizontal, vertical and position-based measurement of rate coefficients immediately connects to superimpose image textures of the high apparel products. Experimental and 16 widely accepted orientations were recognized at 98% especially in comparison with other support poses. Collect, assemble and begin building tissue data in 10 pictures from three models 12 models covering 16 poses in order to try. The obtained results 96,45 percent by validating and running the algorithm without set of data. Tested and engendered well over 50 live footages provided by the user.[1]

IX. REFERENCES

[1] Quang-Song Le and Minh-Triet Tran, "Overlay upper clothing textures to still images based on human pose estimation", Journal of Mobile Multimedia, Vol. 10, No. 1&2 (2014) 043-057

[2] Protopsaltou,D., Luible,C., Arevalo-Poizat,M., and Magnenat-Thalmann,N., A Body and Garment Creation Method for an Internet based Virtual Ftting Room. in Proc. Computer Graphics International 2002 (CGI '02), Springer, 105–122, 2002.

[3] Furukawa, Y. and Ponce, J. , Carved Visual Hulls for Image-based Modeling. in Int. J. Comput.Vision 81 (January), 53–67, 2009.

[4] Bradley, D., Popa, T., Sheffer, A., Heidrich, W., and Boubekeur,T., MarkerlessGarment Capture.in ACM SIGGRAPH 2008, ACM, New York, NY, USA, SIGGRAPH '08, 2008.

[5] Hasler, N., Asbach, M., Rosenhahn, B., Ohm, J.R., and Seidel,H.P., Physically based Tracking of Cloth.in 11th International Fall Workshop on Vision, Modeling, and Visualiza-tion 2006 (VMV 2006), IOS, Aachen, Germany, L. Kobbelt,T. Kuhlen, T. Aach, and R. Westermann, Eds., European As-sociation for Computer Graphics (Eurographics), 49–56, 2006.

[6] Giovanni, S., Choi, Y.C., Huang, J., Khoo, E.T., and Yin,K.K., Virtual Try-On Using Kinect and HD Camera. in Motion in Games - 5th International Conference, MIG 2012, Rennes, France, November 15-17, 2012. Proceedings 2012, 2012.

[7] Isikdogan, F. and Kara, G., A Real Time Virtual Dressing Room Application Using Kinect. 2012.

[8] Hilsmann, A. and Eisert, P., Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications. in In Proc. of Mirage, 94–105, 2009

[9] Hauswiesner, S., Straka, M., and Reitmayr, G., Image-based Clothes Transfer. in Proc. Of International Symposium on Mixed and Augmented Reality (ISMAR), 2011

[10]Ehara, J. and Saito, H., Texture Overlay for Virtual Clothing Based on PCA of Silhouettes. in 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, 139–142, Santa Barbara, USA (2006).

[11]Tanaka, H. and Saito, H., Texture Overlay onto Flexible Object with PCA of Silhouettes and KMeans Method for Search into Database. In Proceedings of the IAPR Conference on Machine Vision Applications, Yokohama, JAPAN, 2009.

[12]Yoon, J.C, Lee, I.K., and Kang, H., Image-based Dress-up System. in Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2011, Seoul, Republic of Korea, February 21 - 23, 2011.

[13]Amit Jnagal, "Image Processing with deep learning – A Quick Start Guide", Infrrd,2018.

[14]https://www.analyticssteps.com/blogs/convolutional-neural-network-cnn-graphical-visualization-code-explanation

[15]https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[16]Tanesh Balodi," Convolutional Neural Network (CNN): Graphical Visualization with python code explanation", analytic steps, 2019.

[17]https://www.tensorflow.org/lite/examples/pose_estimation/overview

[18]https://github.com/tensorflow/tfjs-models/tree/master/posenet

[19]https://towardsdatascience.com/convolutional-neural-network-for-breast-cancer-classification-52f1213dcc9

[20]https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5

[21] https://spyjetson.com/2019/12/jetsonnano-human-pose-estimation-usin.html

[22] https://nanonets.com/blog/human-pose-estimation-2d-guide/