Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

Skeleton and Joint Angle Estimation Based on MobileNet

Halima Tus Sadia
Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
Email: sadializa1998@gmail.com
Lutfur Nahar
Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
Email: nahararju1998@gmail.com
Israt Jahan
Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
Email: israt.cse.19@gmail.com
Md. Khaliluzzaman
Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
Email: khalil@iiuc.ac.bd
Md. Rashedul Islam
Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
Email: rashed_maths@yahoo.com
Md Jiabul Hoque
Department of Computer and Communication Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
Email: jiabul.hoque@iiuc.ac.bd

----------ABSTRACT----------

2D pose estimation is a general problem in computer vision, where the main objective is to detect a person's body key-points and estimate a 2D skeletonized pose of a person.Skeleton estimation is outbound as an essential part of body parts detection in many fields, such as healthcare, rehabilitation, sports and fitness, animation, gaming, augmented reality, robotics. These systems are based on neural network applications and able to give reliable, objective and cost-effective benefits. Various methods are available based on this topic and used to update existing systems. In this regard, in this work, we have proposed a method for skeleton-based angle detection where we have used MobileNet model. This model is developed based on the convolution neural network (CNN). At first, 18 key-points of the human body parts were generated through the model. After that, by using the extracted key-points the skeleton of the human body parts is generated by estimating key-points according to the body part pairs. Furthermore, based on the generated skeletons, different skeleton joint angles at different key-points are estimated. To evaluate the performance of the proposed model at different environmental conditions, a customized dataset was utilized. This approach shows 95.37% accuracy for key-points detection, for joint angle estimation the accuracy is 96.11%, and shows 96.667% accuracy for body part length measurement.

Keywords -**Skeleton, Pose estimation, MobileNet, Heatmap, Convolution Neural Network.**

Date of Submission: April 08,2023 Date of Acceptance: April 26,2023

1. INTRODUCTION

Pose estimation is a fundamental and extensively researched topic in the field of computer vision. It has taken computer vision to another level, just by taking some random images of human beings and estimating their poses. As the time passes by, many different approaches were introduced for efficient body joint detection and estimating these joints in a systematic way to generate human skeleton pose. 2D pose estimation mainly focuses on the problem of localizing body key-points and different body parts.

Detecting key-points of different individuals, especially engaged in different activities, is challenging [5]. A commonly used approach for 2D pose estimation is to use a human detector and estimating their poses individually. This one is called top-down approach which is very time consuming and a very complex process. İn reverse bottom-up approach is more attractive and robust [17].

Skeleton pose estimation is explicitly used in human activity recognition and robotics field. By estimating the angle of a certain body part in certain time duration, machine predicts human activity. As a result, estimating body joint angles correctly is extremely crucial [8]. Skeleton angle estimation can help to detect the athlete's posture, techniques, strengths, weakness and others. It can help the trainers to improve their athletes' performance. Gamers can inject their own poses in the gaming environment, which makes gaming more interactive and enjoyable. Animating a person from 2D field to 3D graphics requires motion tracking [11]. Motion tracking is depended on human's body skeleton angles. Nowadays, famous animation movies are developed by using this technology. Angle, rotation, scale variation and other information about the body parts in a certain time of period helps to predict human activity [8]. It can be extremely helpful for fields like health, security, defense, manufacturing and so on. Detecting the movement of a

Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

human body parts, the exercise is divide into phases of singular and concentric movements to explore different angles of fold and total gesture with the help of estimating the key-points, and giving analytics in the build of graphic experiment.

Main challenges that were faced during the research were suitable dataset collection and camera angle. The testing dataset contains real-time images and all the images were captured from 30-degree angle, to make sure that all the body angles are clearly visible.

The paper is organized as following way. In section II, the present state of the art is presented. The method is explained in section III. The experimental results and discussion is analyzed in section IV. The paper is ended with the conclusion section.

2. RELATED WORK

Previous skeleton pose estimation approaches were proven efficient. One of these efficient methods is MobileNet. Compared to other approaches MobileNet is more vastly used due to its compatibility with embedded mobile applications [1]. Also, it generates more accurate heat-map compared to other approaches. Heat-map is a kind of graphical representation, where, the images having high probability of joint, is represented with higher color intensity values. With the help of heat-map, key-points are detected, then consequently by measuring body part length and using triangular formula skeleton angle is estimated. Howard, et al. [1] introduced an efficient model for mobile applications called MobileNet-V1. Its architecture is built by using depth wise separable convolution filters which builds light weight deep neural network. To introduce a better version of MobileNet, Sinha et al. [2] introduced the Thin MobileNet.

There are also different versions for it. Edelet al. [3] represent comparison between MobileNet-V1 and MobileNet-V2 models which was mainly used for pedestrian recognition. Groos et al. [4] proposed efficient for pose estimation. It offers flexibility while delivering more appropriate key-point score than counterparts in order to spread less parameters and computational costs. Artacho et al. [12] proposed Omni-Pose which is a better version of WASPv2 module. İt increases network performance with high resolution of feature map.

Cao et al. [5] used Part affinity field for OpenPose pose estimation approach, which is used to encode the location and confidence map for orientation of limbs over image input. They applied parsing algorithm for bipartite matching to merge body parts. Hidalgo et al. [6] present multi task learning approach for OpenPose, they also trained a unified model for various key-point detection. Liu et al. [15] represent center point to pose network which predicts center area of a person then regress the whole body pose of each person. It estimates heat-map of each pixel to avoid missing points for joining to different instance. Zhang et al. [7], improved data argumentation via differentiable search algorithm, model architecture by sequential, cascaded, recurrent and adversarial architecture.

By using the skeleton angle joint information İnce et al. [8] proposed a new Human Activity Recognition (HAR) model. Creation of the HAR model was based on rotation and scale variations, complex camera motion, and large interclass variation and data margin issues. Aubry et al. [13] represent the methodology of action recognition from RGB videos by extracting motion as skeleton of people using OpenPose. Sun et al. [18] represented high resolution network, which connects high to low subnet work in parallel without using heat-map.

Patil et al. [9] represents a method for skeleton generation that helps to keep track of the people appearing in the image as a part of scene. It takes RGB image as input and generates a matrices array as output which consists of key-points. Xiao et al. [10] provided a baseline method not only for pose estimation but also for tracking by greedy matching method based on backbone network ResNet. In Chu et al. proposal's [14], convolutional neural systems and a multi-setting consideration tool are combined into a comprehensive structure for estimating human posture.

3. PROPOSED METHOD

In this study, our approach is to develop a skeleton joint angle estimation method. The proposed approach will use a comparatively less complex neural network model and estimate six different angles from the skeleton that was generated through the neural network model. Here, MobileNet-V1 is used for the proposed approach. As it is comparatively less complex and less space consuming, it is contains adjustable hyper-parameters and suitable for both desktop and embedded mobile applications.

3.1 Method

Our approach is to articulate pose estimation and use this pose for skeleton generation and joint angle estimation. The input image of a 224 x 224 x 3 resolution is passed through the neural network model[19] and a set of heat-map is generated. Then using these heat-map a set of key-points are detected and a skeleton is generated by min-max localization. Then the distance formula and cosine formula help to estimate the angles for six joints. The proposed method is presented in Fig. 1.

Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

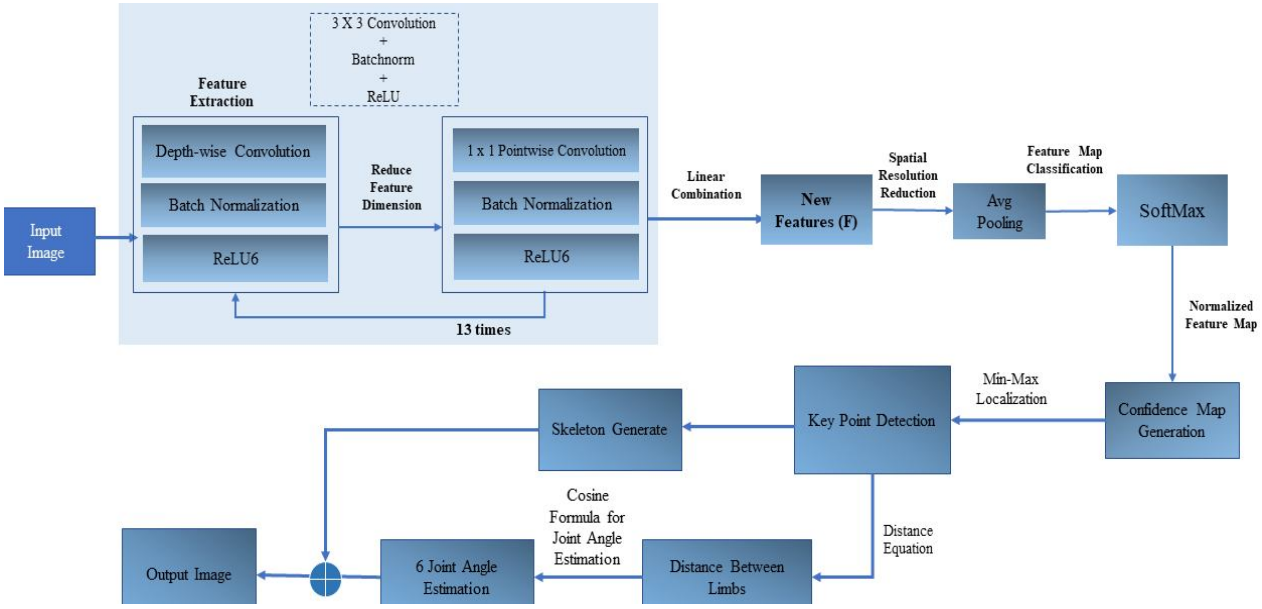


Fig.1. Framework for the proposed approach.

First an RGB image of w x h resolution is taken as input. Neural network model called MobileNet-V1 is used here. It contains 28 layers and has a comparatively simpler architecture rather than other pre-trained models.First layer of MobileNet model is a convolution layer. This layer is followed by batch normalization and ReLU activation function. Core layer of this model is built upon depth-wise separable convolution which is followed by batch normalization and activation function ReLU6. Depth-wise separable convolution layer is a combination of both depth-wise convolution and point-wise convolution. Point-wise convolution combines all of them in a linear format and generates features map which is F. Then for spatial resolution reduction, average pooling is applied on F. For the classification of feature map F, Softmax function is used at the final stage of MobileNet. It will finally result in heat-map. Heat-map is a graphical representation of color intensity values which have high intensity values in certain places, where the probability of key-points is high.

3.2 Skeleton Generation

Using the Heat-map, we get body key-points using the higher values of probability. Then using ellipse shape of red color each key-point is identified in 18 positions. After that, all the 18-body key-points are assembled and joined as per the body pairs declared beforehand. For example, key-point from 'neck' can make pair with 'right shoulder' and 'left shoulder', which will be ['neck', 'right shoulder'] and ['neck', 'left shoulder].

Body Part Length Measurement for Certain Positions

Body part lengths for different positions are different, as the measurement is conducted on 2D surface. Using Euclidian distance formula for two points, body part lengths can be measured. Here, an image of a forearm is taken. Here two key-points P and Q are coordinated at wrist and elbow. Then using eq. (1), length of PQ which is the forearm length, is measured.

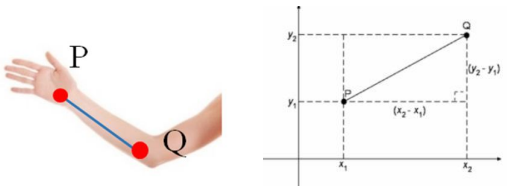


Fig.2. Body part length measurement process.

$$PQ^2 = (x_2 - x_1)^2 - (y_2 - y_1)^2 \quad (1)$$

Here, points P is coordinated at (x_1, y_1) and Q is coordinated at (x_2, y_2). Then Euclidian distance formula i.e., eq. (1) is used for measuring the length of PQ. Thus, it calculates the distance between two real valued objects so as the distance between two key-points P and Q.

3.3 Joint Angle Estimation

After the body part lengths are measured, using Cosine formula, skeleton joint angles are estimated. Here, this approach measures skeleton joint angles for 6 joints. Only, three key-point's co-ordination is necessary to measure joint angles.

$$\cos(p) = \frac{PQ^2 + PR^2 - QR^2}{2 * PQ * PR} \quad (2)$$

$$p = \cos^{-1} \frac{PQ^2 + PR^2 - QR^2}{2 * PQ * PR} \quad (3)$$

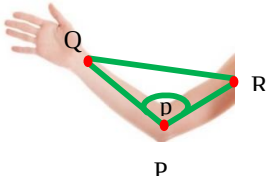


Fig.3. Body joint angle measurement.

Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

Here, three points (P, Q, and R) are coordinates at(x_1, y_1), (x_2, y_2) and (x_3, y_3) respectively. Then using Euclidian Eq. (1), length for PQ, QR and PR are estimated. After that, using Eq. (2) we find, Cos ΔQPR i.e., is p . $\Delta p$is calculated by using Eq. (3). Δpindicates the angle between two body key-points.

3.4 Visual Representation

As the proposed approach is applied to a sample image, the image is passed through MobileNet for key-point extraction. MobileNet is a pose estimation model from TensorFlow. Then skeleton is generated from the key-points and finally the angles are estimation. The complete process is explained step by step with a sample image in Fig. 4.

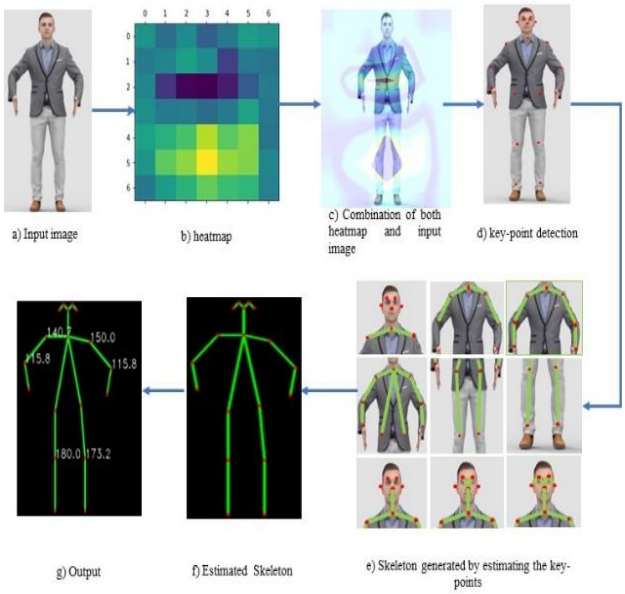


Fig.4. The proposed approach applied on a sample image.

4. RESULTS AND DISCUSSION

For the experiments, different samples under different environmental conditions are considered. Also, these samples were collected in three different types of illumination, which are normal, low, and high. This experiment was conducted on Google Colaboratory and the images were taken in RGB format. All the images were resized into 400 x 600 pixels.

The main purpose of this research is to generate a 2D skeleton from a single input image and to estimate the joint angles of the skeleton. To conduct the research, we have studied several papers on Convolutional Neural Network, Deep Learning and mainly Pose Estimation

.

Here, a neural network model MobileNet was used which detected the body key-point from the human body image. Then by estimating the key-points the skeleton was generated, and by measuring the distance between the key-points the body part length for different poses were measured. Finally, by using body part length and the cosine formula, the joint angles were detected. For testing the model, a customized dataset of 30 images were used. The sample dataset is presented in Fig. 5.

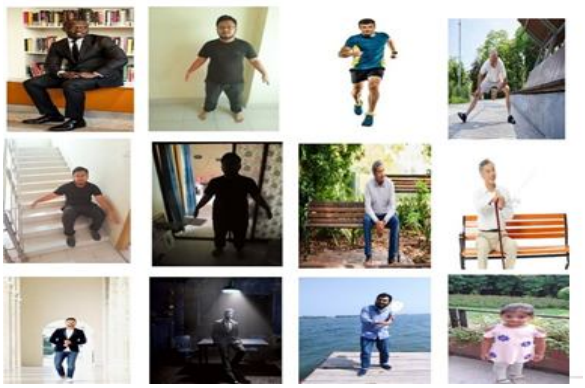


Fig. 5.The sample data used for evaluating the proposed model.

Fig. 6 indicates experimental example for skeleton joint angle estimation for outdoor environments. Images from sample 1, 2 and 3 were captured in outdoor settings. These images were collected from both online resources or captured in real-time. Images from sample 1 are captured from the normal illumination, samples 2 are from low illumination and samples 3 are from high illumination. Each of them had different pose and different types of background which made the experiment more effective and concrete.

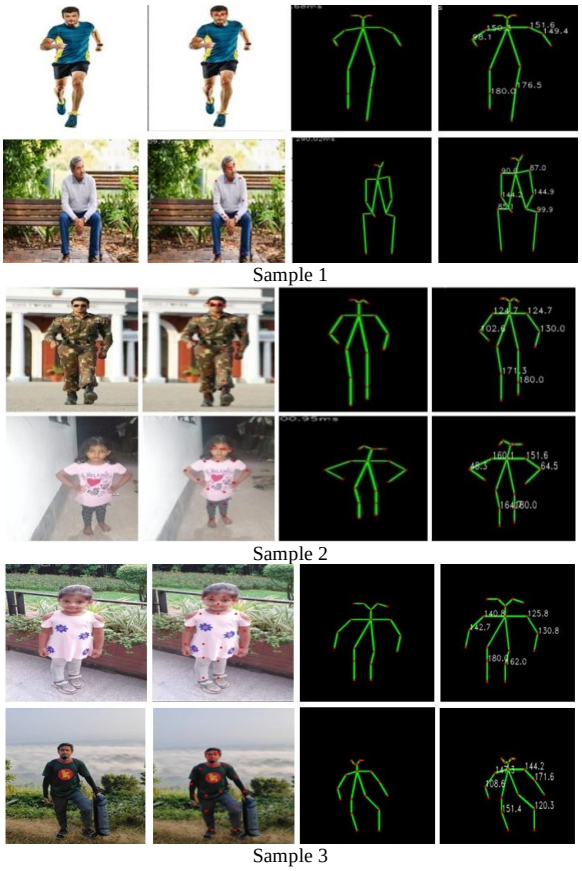
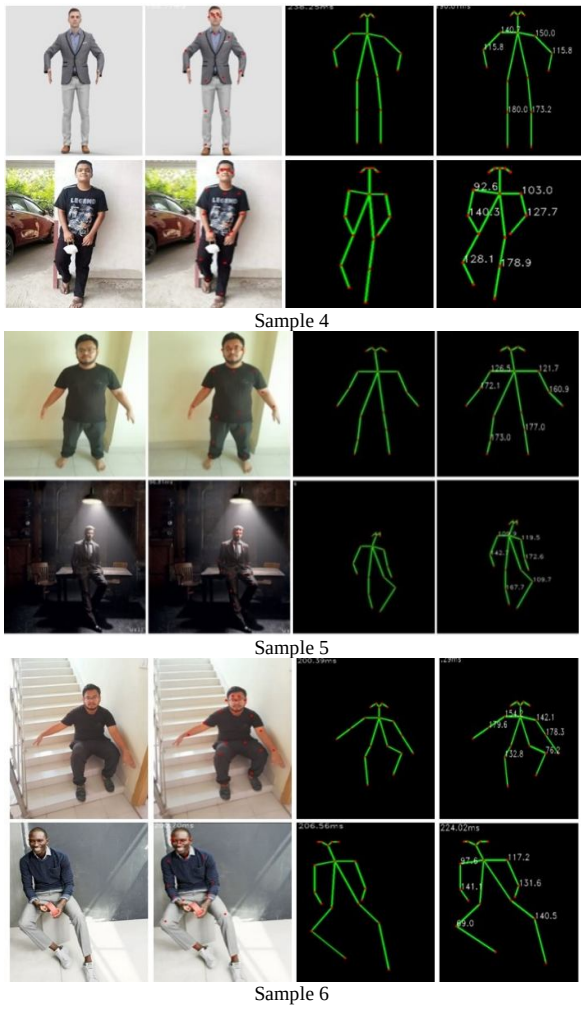


Fig. 6. Experimental examples of skeleton joint angle estimation in outdoor under 3 types of illumination: Sample 1(normal), Sample 2 (low), Sample 3(high) Here, a) Input Image, b) Key-point detection, c) skeleton generation by joining the Key-points, d) Joint Angle Estimation.

Fig. 7 indicates experimental example for skeleton joint angle estimation for indoor setting. Images from sample 4, 5

Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

and 6 were captured from indoor environments and each of them were collected from either online resource or captured in real-time. Sample 4 images are from normal illumination, samples 5 images are from low illumination and samples 6 images are from high illumination.



Sample 4

Sample 5

Sample 6

(a) (b) (c) (d)

Fig. 7. Experimental examples of skeleton joint angle estimation in indoor under 3 types of illumination: Sample 4(normal), Sample 5 (low), Sample 7(high) Here, a) Input Image, b) Key-point detection, c) skeleton generation by joining the Key-points, d) Joint Angle Estimation.

Table 1: Accuracy for no of key point detection and body part length measurement

Settings	Illumination	No of key-point detected (90)	No of body part length measured(50)	Accuracy For key-point	Accuracy For Body-part length
Indoor	Low	83	46	94.81%	95.33%
	Normal	84	47		
	High	89	50		
Outdoor	Low	86	50	95.92%	98%
	Normal	87	49		
	High	86	48		

To measure the accuracy of the approach for body part length measurement, at first the number of correctly detected body part lengths were counted, then it was compared with the number of body part lengths the method should estimate.

Fig. 8 demonstrates graphical representation of accuracy for body part length measurement. Here, the graph is plotted between the no of total body part length measured by the model and the desired amount of body part length each image should measure.

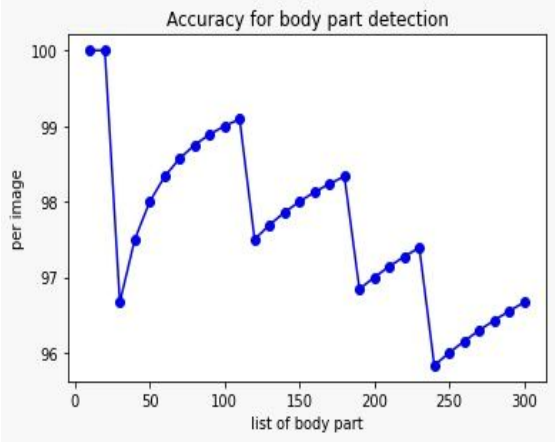


Fig. 8.Accuracy for body part length measurement.

Accuracy for joint angle detection depends on key-point detection. As long as the key points are correctly detected, the joint angles can be measured accurately. Accuracy curve for joint angles detection are generated between the number of correct predictions for each image and the number of total joint angles. Here, number of total joint angle in each image is six.

Fig. 9 shows graphical representation for accuracy curve of joint angle detection. The graph is plotted between the number of total joint angles the approach should estimate per image, which is six per image and the number of angles that were actually detected for each image.

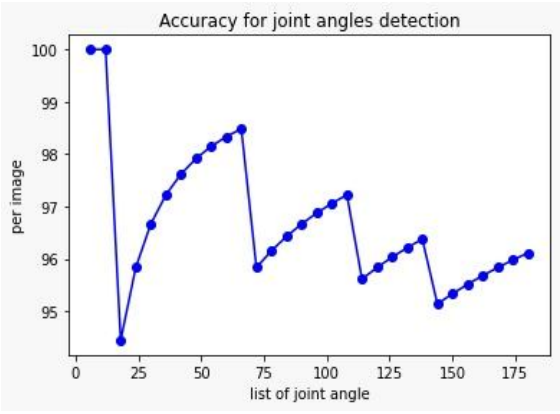


Fig. 9.Accuracy for joint angle detection.

The accuracy of the joint angle estimation from different environmental conditions is presented in Table 2. The accuracy of the body key-point, joint angle, and body length measurement are shown in Table 3.

Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

Table 2: Accuracy for joint angle estimation

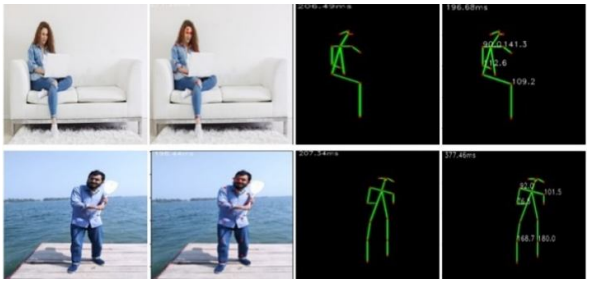
Environment	Illumination	No of Angles (30)	Accuracy for different Settings	Total Accuracy
Indoor	Low	28	95.55%	96.11%
	Normal	28		
	High	30		
Outdoor	Low	30	96.66%	
	Normal	29		
	High	28		

Table 3: Total accuracy for key point detection, body part length measurement and joint angle estimation

Features	Accuracy (%)
For body key-point	95.37
For joint angle	96.11
For body part length measurement	96.67
Average	**96.05**

4.1 Discussion

The proposed approach provides the significant outcome to generate the skeleton and angles from the input person's image. However, the proposed approach does not provide the accuracy for the all orientation input images especially for the image where the body parts are occulted. Some processing example to show the problem is presented in Fig. 10.



(a) (b) (c)(d)

Sample 7

Fig. 10.Some Limitation for the proposed approach.

Fig. 10 shows the limitation of the proposed approach. The limitation is mainly occlusion problem. Here, in the upper image, the left hand is entirely hidden in this image, and the left leg is only partially visible due to the occlusion caused by the presence of another object. As a result, the proposed method is unable to detect the key-points of the left elbow, left wrist, left hip, left upper knee and left ankle of the body.

As a result, the joint angles of the body could not be detected.In the lower image, presences of a different object hid the key-points of the left shoulder and the left elbow, making it impossible for the proposed method to identify them. As a result, it was impossible to determine the angle between the left elbow and left wrist.

The proposed method is compared with [4] and presented in Table 4.

Table 4: Compare with present state-of-the-art

Method	Accuracy (%)
EfficientPose – IV [4]	91.20
MobileNet [Proposed]	**96.05**

5. CONCLUSIOINS

In this work, we have proposed a method that can estimate body pose from a person's image. This approach is estimated that person's body part length for the pose and six angles between the joints for that person. For evaluating the proposed method, we have used our own customized dataset. The method shows satisfactory accuracy for joint angle estimation in different environmental conditions such as 95.55% for indoor and 96.66% for outdoor environmental condition. There are some limitations in this approach. The proposed approach is unable to overcome the occlusion problem. In future, we will focus on this issue. Also, we have a plan to make our method work for video sequences and for automated pain assessment. Again, our proposed approach can estimate 6 joint angles we will update this method for 8 joint angles.

REFERENCES

[1] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., "Mobilenets: Efficient convolutional neural networks for mobile vision applications", arXiv preprint arXiv:1704.04861, (2017).

[2] Sinha, D. and El-Sharkawy, M., "Thin MobileNet: an enhanced mobilenet architecture", 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON), (2019).

[3] Edel, G. and Kapustin, V., "Exploring of the MobileNet V1 and MobileNet V2 models on NVIDIA Jetson Nano microcomputer", In Journal of Physics: Conference Series IOP Publishing, (2022).

[4] Groos, D., Ramampiaro, H. and Ihlen, E.A., "EfficientPose: Scalable single-person pose estimation", Applied intelligence, 51:2518-2533, (2021).

[5] Cao, Z., Simon, T., Wei, S.E. and Sheikh, Y., "Realtime multi-person 2d pose estimation using part affinity fields", In Proceedings of the IEEE conference on computer vision and pattern recognition, 7291-7299, (2017).

[6] Hidalgo, G., Raaj, Y., Idrees, H., Xiang, D., Joo, H., Simon, T. and Sheikh, Y., "Single-network whole-body pose estimation", In Proceedings of the IEEE/CVF International Conference on Computer Vision, 6982-6991, (2019).

[7] Zhang, F., Zhu, X. and Wang, C., "Single person pose estimation: a survey", arXiv preprint arXiv:2109.10056, (2021).

[8] İnce, Ö.F., Ince, I.F., Yıldırım, M.E., Park, J.S., Song, J.K. and Yoon, B.W., "Human activity recognition with analysis of angles between skeletal joints using a RGB-depth sensor", ETRI journal, 42(1):78-89, (2020).

[9] Patil, M.R.R., Chaugule, S.V. and Malemath, V.S., "POSE ESTIMATION FOR SKELETON DETECTION", International Journal of Engineering

Volume: 14 Issue: 06 Pages: 5727 - 5733 (2023) ISSN: 0975-0290

Applied Sciences and Technology, 4(4):192-196, (2019).

[10] Xiao, B., Wu, H. and Wei, Y., “Simple baselines for human pose estimation and tracking”, In Proceedings of the European conference on computer vision (ECCV), 466-481, (2018).

[11] Akhter, I. and Black, M.J.,” Pose-conditioned joint angle limits for 3D human pose reconstruction”, In Proceedings of the IEEE conference on computer vision and pattern recognition, 1446-1455, (2015).

[12] Artacho, B. and Savakis, A., “Omnipose: A multi-scale framework for multi-person pose estimation”, arXiv preprint arXiv:2103.10180, (2021).

[13] Aubry, S., Laraba, S., Tilmanne, J. and Dutoit, T., “Action recognition based on 2D skeletons extracted from RGB videos”, In MATEC Web of Conferences EDP Sciences, (2019).

[14] Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A.L. and Wang, X., “Multi-context attention for human pose estimation”, In Proceedings of the IEEE conference on computer vision and pattern recognition, 1831-1840, (2017).

[15] Liu, H., Wu, J. and He, R., “Center point to pose: Multiple views 3D human pose estimation for multi-person”, Plos one, 17(9):0274450, (2022).

[16] Martinez, J., Hossain, R., Romero, J. and Little, J.J., “A simple yet effective baseline for 3d human pose estimation”, In Proceedings of the IEEE international conference on computer vision 2640-2649, (2017).

[17] Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P.V. and Schiele, B., “Deepcut: Joint subset partition and labeling for multi person pose estimation”, In Proceedings of the IEEE conference on computer vision and pattern recognition, 4929-4937, (2016).

[18] Sun, K., Xiao, B., Liu, D. and Wang, J., “Deep high-resolution representation learning for human pose estimation”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 5693-5703, (2019).

[19] Özbay, E. and Özbay, F.A., "A cnn framework for classification of melanoma and benign lesions on dermatoscopic skin images", International Journal of Advanced Networking and Applications, 13(2):4874-4883,(2021).