International Journal of Technology and Engineering Studies
volume 8 pp. 34-40 doi: https://dx.doi.org/10.20469/ijtes.8.40005

Real-Time 3D Human Pose Estimation Using Deep Learning Model for Ergonomics

Jayabhaduri Radhakrishnan*
Computer Science and Engineering,
Alliance University,
Bengaluru, Karnataka 562106, India.

Aadesh Vijayaraghavan
Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering,
Sriperumbudur Taluk 602 117, India.

Ajay Karthik R
Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering,
Sriperumbudur Taluk 602 117, India.

Ramana Prasath G
Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering,
Sriperumbudur Taluk 602 117, India.

Mohamed Arshath S
Consultant Physiotherapist,
Chennai 600041,
India.

Abstract: In the recent days, most of the people stay in a hunchback position for a long time, due to usage of electronic gadgets like smartphones, personal computers, laptops and tablets, which causes neck and back pain in large numbers, which causes Text neck syndrome, Musculoskeletal disorders, Carpal Tunnel Syndrome and Computer Vision Syndrome. Hence it has become mandatory for people to be mindful of their posture while sitting for long hours. Human pose estimation has gained a lot of attention amongst researchers in a wide range of applications including computer vision, video analytics and motion analysis. To address these risk factors, attempts have been made to develop a 3D Human Pose Estimation (3D-HPE) model for detecting and correcting frontal plane (anterior) sitting postures in computer workstation ergonomics using MediaPipe, and various variants of YOLO, algorithms. The proposed model locates landmarks and analyzes kinematic points from the input video captured through a web camera. From these kinematic points, the 3D-HPE model analyzes whether the human postures are good or bad based on the temporal duration of prolonged time of poor posture. YOLOv8 gives the highest accuracy which is determined by mAP (Mean Average Precision) value found 91.2 in terms of computational time and human pose estimation. Hence, YOLOv8x-pose is the most suited deep learning algorithm for Real-time 3D Human Pose Estimation in Ergonomics. The proposed model notifies the human by sending an alert message to the device for posture correction.

Keywords: *Deep learning, ergonomics, human pose estimation, musculoskeletal disorders, text neck syndrome*

Received: 10 August 2022; ***Accepted:*** 02 September 2022; ***Published:*** 28 December 2022

I. INTRODUCTION

In many facets of our life, ergonomics, the science of planning and arranging spaces and items to maximize human performance and well-being, plays a crucial role. Ergonomics are essential in any setting, including homes, businesses, and even leisure activities which is based on the ideas of improving efficiency, comfort, and safety while lowering the possibility of health problems and

*Correspondence concerning this article should be addressed to Jayabhaduri Radhakrishnan, Computer Science and Engineering, Alliance University, Bengaluru, Karnataka 562106, India. E-mail: jayabhaduri17@gmail.com

discomfort.

A. *Ergonomics*

Ergonomics in the workplace is a major factor in driving productivity and employee satisfaction. It helps to lessen tiredness, injuries, and discomfort by customizing workstations, tools, and equipment to match the individual. Higher productivity, less absenteeism, and fewer healthcare expenses for employers follow from this. The influence of ergonomics goes beyond the workplace. It affects the design of items we use on a daily basis, such as chairs, keyboards, smartphones, personal computers, laptops and tablets. It directs the creation of medical tools and approaches to patient care in the field of healthcare, ensuring that surgeries and treatments are carried out with the least amount of stress on patients and medical staff. Its primary goal is to improve our health, performance, and general quality of life while reducing the risk of physical discomfort, injuries, and health problems.

Detection, recognition and analysis of human actions and behaviors plays an important role in real-time applications such as surveillance [1]; [2]; [3]; [4], humancomputer interaction [5], assistive technologies [6], sign language [7]; [8]; [9], computational behavioral science [10]; [11] and consumer behavior analysis [12]. These applications have paved the way for the researchers in the Computer Vision community to conduct research on action recognition and human pose estimation [13]; [14]; [15]; [16] [17]; [18]. Sitting posture recognition plays a vital role in ergonomics in preventing work-related Musculoskeletal Disorders (MSDs), computer vision syndrome and Text neck [19].

1) Text neck syndrome :The phrase "text neck" has become more well-known as a health problem in the current digital era. It describes the forward-leaning head and neck stance that many people take when absorbed on digital devices. Due to the sustained stress it puts on the cervical spine, this position, which at first glance seems harmless, can have serious repercussions [20]. Text neck is a serious issue since it frequently results in musculoskeletal issues. Mild discomfort that first develops over time may become chronic pain and more serious spine problems. Due to their widespread use and integration into our everyday lives, cellphones and other portable gadgets play a key role in this problem. Beyond simple discomfort, the effects of text neck frequently result in headaches, shoulder pain, and back pain. This discomfort may have an effect on your mood, productivity, and general well-being. Children and teenagers who use devices frequently are also at risk for these issues.

Text neck is a representation of how closely technology and health are intertwined. It highlights the value of ergonomics in the digital age and the necessity of thoughtful device use [21]. Text neck can be avoided with simple changes like holding devices at eye level and taking frequent screen breaks. Addressing text neck's effects is essential as our world becomes more connected. It urges us to prioritize our physical health along with the advantages of technology in our digital lives. In essence, text neck serves as a warning about the importance of maintaining our health in the digital age, making sure that technology improves our lives without harming our bodies.

2) Musculoskeletal disorders :In terms of both public health and the world's workforce, musculoskeletal disorder (MSDs) poses a ubiquitous and complex problem. These illnesses cover a broad range of ailments that impact the muscles, tendons, ligaments, joints, bones, and associated anatomical systems, resulting in a variety of unpleasant and frequently incapacitating symptoms. MSDs have become a significant issue in modern life, ranging from the crippling swells of low back pain to the paralyzing grasp of carpal tunnel syndrome. They have a significant negative impact on people, economies, and businesses all over the world. MSDs have a significant impact on workplaces, healthcare systems, and communities because of their frequency across a range of vocations and age groups.

Beyond the physical domain, the effects can have an impact on mental and emotional health and lower overall quality of life. Thus, efforts to study, control, and prevent these illnesses have become crucial. In order to understand the intricacies of MSDs, epidemiological research is crucial since it illuminates the causal causes, risk factors, and interactions between occupational and lifestyle variables. A thorough investigation of this subject is necessary to guide policies, interventions, and practices intended to lessen the burden that these conditions place on people and society as a whole as the globe struggles with the issues brought on by the growing occurrence of MSDs. This multidimensional study examines the epidemiologic data pertaining to work-related neck, upper extremities, and low back musculoskeletal problems, lighting the way forward.

3) Human pose estimation : Human Pose estimation models can be ergonomists to evaluate the productivity and performance of employees by studying their posture [22]. At present, postural monitoring and diagnosis is carried out by means of specific questionnaires, however, these models cannot be suitable to monitor a person on a continuous basis.

2D human pose estimation is a computer vision task

that leverages learning to accurately identify and locate critical points or landmarks, on the human body in a 2D image. These key points often involve body parts such as limbs, joints, well as essential regions like the head and thorax. The process of determining a human body's modular 3D joint locations from an image or video is known as three-dimensional (3D) human pose estimation. 3D human pose estimation is currently receiving more attention in the computer vision community due to its widespread applications in a wide range of fields, including human motion analysis, human-computer interaction, and robots. However, it is a difficult task due to depth ambiguities and the lack of in-the-wild datasets.

B. *Object Detection*

Object detection witnesses great success in computer vision due to the significant developments in neural networks especially deep learning [23]; [24]; [25]. To tackle 2D human pose estimation, neural networks (CNNs) are employed due to their capability to extract intricate features from images effectively. To begin the process, a dataset of photos is needed, that have been annotated with key point positions. The model architecture plays a role in this task as researchers and professionals create networks that can effectively comprehend the intricate spatial interactions and arrangements of these key points. By reducing a predefined loss function often measured using error, which quantifies the difference between predicted and actual key point coordinates the model becomes capable of predicting the (x, y) coordinates of these key points during training.

YOLO, or You Only Look Once, is a game-changer in good posture estimation. It excels in real-time object detection, making it ideal for continuous posture monitoring. YOLO's architecture divides images into a grid and predicts object bounding boxes and class probabilities, which translates well to keypoint estimation tasks like identifying posture landmarks. By training YOLO on labeled datasets, it can quickly and accurately locate these critical body landmarks. What sets YOLO apart is its adaptability to varying conditions and camera angles, handling lighting changes, background clutter, and occlusions. It delivers instant feedback on posture, making it effective for correction and habit-building [26]; [27]. This real-time capability is invaluable for healthcare, individuals, and tech developers aiming to leverage AI for promoting and maintaining optimal posture, contributing to overall well-being.

1) Challenges The challenges addressed in this research work are as follows:

- Variability in Poses (Occlusion): Real-world scenarios often involve humans in varying poses and levels of occlusion. Object detection models must contend with these challenges to ensure accurate identification, even when objects are partially obscured or situated differently than during training.
- Real-World Environment: Real-world environments introduce complexities such as dynamic lighting conditions, unpredictable backgrounds, and varying camera angles. Adapting object detection models to these conditions is vital for practical applications.
- Limited Data and Annotated Datasets: Annotating data for object detection tasks can be labor-intensive and time-consuming. As a result, datasets may be limited in size or diversity. Object detection models must contend with the constraints imposed by the availability of annotated data.
- Time Constraint: In ergonomics, object detection models often have limited time to make accurate predictions.

C. *Contribution*

The contribution in this research work is as follows:

- Developed a 3D Human Pose Estimation (3D-HPE) model for detecting and correcting frontal plane (anterior) sitting postures in computer workstation ergonomics using deep learning models
- Proposed model analyzes whether the human postures are good or bad based on temporal duration of prolonged time of poor posture.
- Model generates higher accurate results in terms of mean Average Precision (mAP).
- Deployed for Real-time work environments.
- Notifies the human by sending an alert message to the device for posture correction.

II. LITERATURE REVIEW

[28] made a thorough review of 3D pose estimation based on existing deep learning models and stated its advantages and disadvantages of each of the models extensively. The benchmark datasets used for comparison and analysis are explored and the study on the current state of development in 3d human pose estimation with the insights that can facilitate future improvements in design and algorithm of such models are deeply discussed. [29] surveyed the human action classifications based on 3D skeleton models. Their survey studies the technologies and approaches used for 3D skeleton based classification and points out the motivations and challenges in that domain. They introduce a categorization of most recent

works in 3D-skeleton based action classification based on the comparative study between different types data pre-processing techniques, publicly available benchmarks and commonly used accuracy measurements.

[30] presents a general framework of human pose estimation in which they explore the limitations and challenges of a few existing work and discuss the scope for future improvements in the field. [31] developed a feasible and reliable RGB-D scene healthy human sitting posture estimation framework using 15 skeletal joints which was extracted from Kinect as the initial input. The health-constrained spatial and relationships between objects and human skeletal joints in RGB-D scene was calculated using the Naive Bayes Classifier. The skeleton joints distribution of a healthy individual was produced and a framework was tested on a dataset with RGB-D scenes. [32] introduces YOLO-pose and 2D multi-person pose estimation in a framework that allows to train a model end-to-end and optimize the object key similarity. All the persons are localized along with their pose in a single inference. Test time augmentations are not used in all experiments unlike the traditional approaches that use flip-test and multi-scale to boost performance.

[21] proposed a deep learning model fusing multi model data and sitting posture recognition containing modality-specific backbones, a cross-modal self-attention module, and multi-task learning-based classification. Their model showed high-performance results indicating that the proposed model is promising for sitting posture-related applications.

III. METHODOLOGY

The proposed research work is for desk job people across the world to provide them a healthy lifestyle and ensure that they dont fall for problems like neck pain, shoulder pain, hunchback. The model constitutes Data Collection, Object Detection, Keypoint Estimation and Pose Estimation using YOLO modules whose architectural diagram is shown in Figure 1.

Fig. 1. Architecture Diagram For 3D-HPE Using YOLO

A. *Data Collection*

A user-friendly interface for recording real-time video in a 2D format inside an ergonomic framework is provided by the Data Collection module, which is created to interact smoothly with the user's built-in webcam. The ability to capture videos continuously ensures unbroken data gathering for lengthy periods of time. The module integrates calibration tools to preserve measurement accuracy and may be configured for use in a variety of research environments. It is also compatible with a number of operating systems. It enables accurate information gathering.

B. *Object Detection*

3D HPE makes use of YOLO and MediaPipe frameworks to recognize and localize objects in pictures and video frames. MediaPipe object detection model applies inference to each frame of a video stream, searching for objects. Each object detected by MediaPipe generates a bounding box and a label to designate its category such as "person," "car," or "dog." These detections come with confidence scores, which represent the model's level of confidence in the presence of the object. MediaPipe additionally uses non-maximum suppression to assure precise and non-redundant object detection.

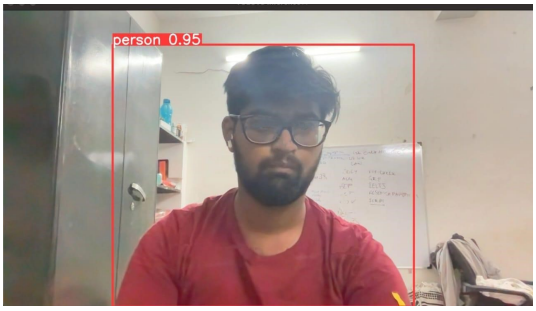

Fig. 2. Object detection

C. *Keypoint Estimation*

The 3D HPE model analyzes real-time video and performs kinematic keypoint estimation to determine the coordinates of specific body joints namely Head, Neck, Eyes, Ears, Left Shoulder and Right Shoulder from a sitting human posture. The inclination of the neck is a critical determinant of posture quality, as the neck bears the full weight of the head and serves as a convergence point for all nerves in the spinal cord.

D. *Pose Estimation Using YOLO*

The model selection process begins with a critical decision, choosing the YOLO architecture best suited for pose estimation. Our research work considers two variants of YOLO namely 'yolov8n-pose.yaml' and 'yolov8x-pose-p6.pt,' each offering unique capabilities. Further enhancing our model's capabilities is the incorporation of pretrained weights, denoted as 'yolov8x-pose-p6.pt.'. These weights bring the knowledge accumulated from a broader dataset, providing a significant boost to our model's understanding of human poses. The model is trained using COCO-Pose dataset which supports 17 keypoints for human figures, facilitating detailed pose estimation with 100 epochs, each epoch consists of the YOLO model to go back and forth through the data. Figure 3 shows the pose estimation using YOLO carried out in our research work.

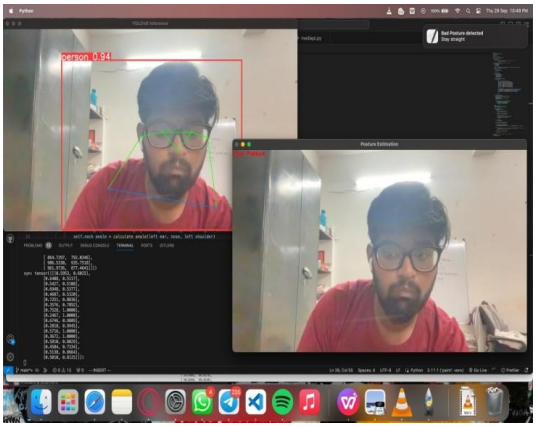

Fig. 3. Pose estimation using YOLO

1) Alert notification :Alert is triggered if bad posture is detected. Notifies users to correct posture with a small suggestion Please Sit Straight to improve posture. Figure 4 shows an alert notification sent to the desktop.

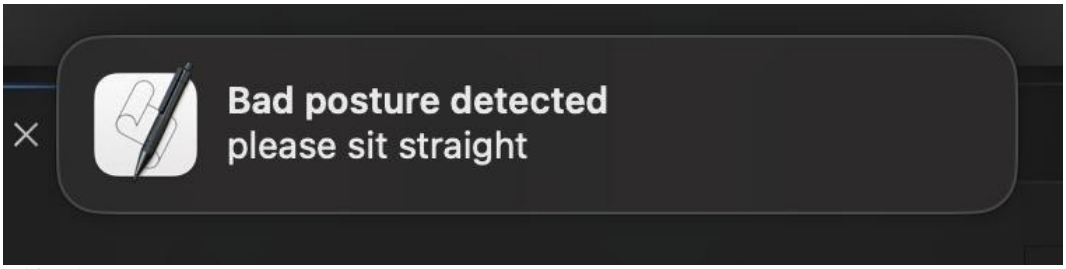

Fig. 4. Alert notification

IV. RESULTS

The developed 3D-HPE model is trained for 100 epochs using Adam optimizer with ReLU activation function. The performance of the model is determined by evaluation metrics namely mAP (mean Average Precision) and Intersection over Union (IoU) as shown in Figure 5. YOLOv8 gives the highest accuracy value found 91.2 in terms of computational time and human pose estimation.

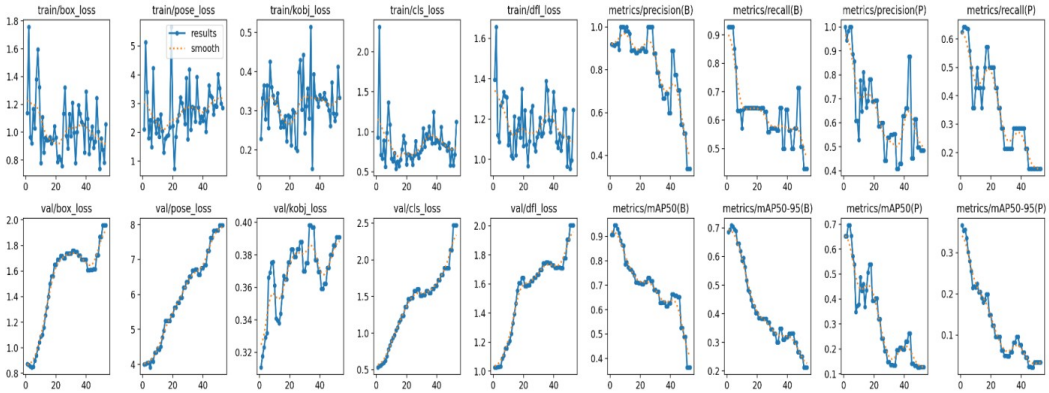

Fig. 5. Evaluation metrics

V. DISCUSSION & CONCLUSION

Our model is developed using YOLOv8x-pose to address health risk factors such as Carpal Tunnel Syndrome, Text Neck Syndrome, Musculoskeletal disorders and Computer Vision Syndrome, to improve their lifestyle using human pose estimation in the context of ergonomics. As the model generates accurate results in terms of Mean Average Precision (mAP), it can be deployed for real-time work environments for desk jobs. This model correctly detects the imbalances in sitting posture and immediately alerts the person to improve his/her posture. In the future, this is aimed to be made into a plug-in like Google extension. This can be extended to be made for multiple persons who become helpful for ergonomical study in professional working spaces.

REFERENCES

[1] S. Kwak, B. Han, and J. H. Han, "Scenario-based video event recognition by constraint flow," in *CVPR 2011*. IEEE, 2011, pp. 3345–3352.

[2] U. Gaur, Y. Zhu, B. Song, and A. Roy-Chowdhury, "A string of feature graphs model for recognition of complex activities in natural videos," in *2011 International conference on computer vision*. IEEE, 2011, pp. 2595–2602.

[3] S. Park and J. Aggarwal, "Recognition of two-person interactions using a hierarchical bayesian network," in *First ACM SIGMM international workshop on Video surveillance*, 2003, pp. 65–76.

[4] I. N. Junejo, E. Dexter, I. Laptev, and P. Perez, "View-independent action recognition from temporal self-similarities," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 1, pp. 172–185, 2010.

[5] Z. Duric, W. D. Gray, R. Heishman, F. Li, A. Rosenfeld, M. J. Schoelles, C. Schunn, and H. Wechsler, "Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1272–1289, 2002.

[6] J.-D. Huang, "Kinerehab: a kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities," in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, 2011, pp. 319–320.

[7] A. Thangali, J. P. Nash, S. Sclaroff, and C. Neidle, "Exploiting phonological constraints for handshape inference in asl video," in *CVPR 2011*. IEEE, 2011, pp. 521–528.

[8] A. T. Varadaraju, "Exploiting phonological constraints for handshape recognition in sign language video," Ph.D. dissertation, Boston University, 2013.

[9] H. Cooper and R. Bowden, "Large lexicon detection of sign language," in *Human–Computer Interaction: IEEE International Workshop, HCI 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings 4*. Springer, 2007, pp. 88–97.

[10] J. Rehg, G. Abowd, A. Rozga, M. Romero, M. Clements, S. Sclaroff, I. Essa, O. Ousley, Y. Li, C. Kim *et al.*, "Decoding children's social behavior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3414–3421.

[11] L. Presti, S. Sclaroff, and A. Rozga, "Joint alignment and modeling of correlated behavior streams," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 730–737.

[12] H. Moon, R. Sharma, and N. Jung, "Method and system for measuring shopper response to products based on behavior and facial expression," Jul. 10 2012, uS Patent 8,219,438.

[13] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 1014–1021.

[14] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in *CVPR 2011*. IEEE, 2011, pp. 1385–1392.

[15] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Strike a pose: Tracking people by finding stylized poses," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 271–278.

[16] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3d human pose annotations," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 1365–1372.

[17] D. Tran and D. Forsyth, "Improved human parsing with a full relational model," in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*. Springer, 2010, pp. 227–240.

[18] C.-H. Chen and D. Ramanan, "3d human pose estimation= 2d pose estimation+ matching," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7035–7043.

[19] X. Zhang, P. Zheng, T. Peng, Q. He, C. K. Lee, and R. Tang, "Promoting employee health in smart office: A survey," *Advanced Engineering Informatics*, vol. 51, p. 101518, 2022.

[20] X. Zhang, J. Fan, T. Peng, P. Zheng, C. K. Lee, and R. Tang, "A privacy-preserving and unobtrusive sitting posture recognition system via pressure array sensor and infrared array sensor for office workers," *Advanced Engineering Informatics*, vol. 53, p. 101690, 2022.

[21] X. Zhang, J. Fan, T. Peng, P. Zheng, X. Zhang, and R. Tang, "Multimodal data-based deep learning model for sitting posture recognition toward office workers health promotion," *Sensors and Actuators A: Physical*, vol. 350, p. 114150, 2023.

[22] W. Kim, J. Sung, D. Saakes, C. Huang, and S. Xiong, "Ergonomic postural assessment using a new open-source human pose estimation technology (openpose)," *International Journal of Industrial Ergonomics*, vol. 84, p. 103164, 2021.

[23] L. Barks, S. L. Luther, L. M. Brown, B. Schulz, M. E. Bowen, and G. Powell-Cope, "Development and initial validation of the seated posture scale," *JRRD: Journal of Rehabilitation Research and Development*, vol. 52, no. 2, p. 201, 2015.

[24] A. R. Pathak, M. Pandey, and S. Rautaray, "Application of deep learning for object detection," *Procedia computer science*, vol. 132, pp. 1706–1717, 2018.

[25] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, p. 103514, 2022.

[26] H. Fu, J. Gao, and H. Liu, "Human pose estimation and action recognition for fitness movements," *Computers & Graphics*, vol. 116, pp. 418–426, 2023.

[27] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3d human pose estimation: A review of the literature and analysis of covariates," *Computer Vision and Image Understanding*, vol. 152, pp. 1–20, 2016.

[28] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, and L. Shao, "Deep 3d human pose estimation: A review," *Computer Vision and Image Understanding*, vol. 210, p. 103225, 2021.

[29] L. L. Presti and M. La Cascia, "3d skeleton-based human action classification: A survey," *Pattern Recognition*, vol. 53, pp. 130–147, 2016.

[30] Z. Liu, J. Zhu, J. Bu, and C. Chen, "A survey of human pose estimation: the body parts parsing based methods," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 10–19, 2015.

[31] B. Liu, Y. Li, S. Zhang, and X. Ye, "Healthy human sitting posture estimation in rgb-d scenes using object context," *Multimedia Tools and Applications*, vol. 76, pp. 10 721–10 739, 2017.

[32] D. Maji, S. Nagori, M. Mathew, and D. Poddar, "Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2637–2646.