



Realization of a Real-Time Decision Support System to Reduce the Risk of Diseases Caused by Posture Disorders Among Computer Users

Enes Gumuskaynak¹, Faruk Toptas¹, Recep Aslantas¹, Fatih Balki¹,
and Salih Sarp²(✉)

¹ Aurora Bilisim, Istanbul, Turkey

{enes.gumuskaynak, faruk.toptas, recep.aslantas,
fatih.balki}@aurorabilisim.com

² Virginia Commonwealth University, Richmond, USA
sarps@vcu.edu

Abstract. Nowadays, diseases caused by posture disorders are becoming more common. This situation reduces the working efficiency of people, especially computer users. This study aims to provide prevention of diseases caused by posture disorders faced by computer users and realize an application software to reduce disease risks. With this realized application, computer users' movements are monitored through the camera, and the situations that may pose a risk of disease for the users are determined. Realized application software is a decision support system. This decision support system provides users suggestions to change their position according to their instant postures and supports them to work more efficiently. The user data is collected by processing the images taken from a camera using the developed computer vision algorithm. Two-dimensional (2D) human exposure estimation is performed with the obtained data. The situations that can decrease the working efficiency are specified with the data obtained from exposure estimation using the developed model. As a result of these findings, increasing the working efficiency is provided by informing the user in real-time about the situation that may decrease the working efficiency.

Keywords: Computer vision with deep learning · Computer vision with pose estimation · Decision support system

1 Introduction

At the point brought by technology, the computer has become one of the indispensable parts of our lives. The time spent in front of the computer increased with the change of both remote working and education [1]. Depending on this situation, diseases such as tendinitis, myofascial pain syndrome, neck hernia, pulmonary embolism, eye strain, and humpback caused by postural disorders continue to increase and decrease the quality of life and work efficiency [2]. Recent advances in Artificial Intelligence (AI) provide new approaches in many industries [3, 4]. In this study, pose estimation is made by processing

the images taken from the computer's camera to prevent the mentioned diseases. The pose estimation was developed with reference to the OpenPose [5] study, which was developed using deep neural networks and shared as open-source in python language. This study uses a bottom-up approach. Although this approach works slower than the top-down approach, it has been preferred. After all, it gives more accurate results because it can capture spatial dependencies between different people that require global inference. Posture analysis is performed using values obtained from the pose estimation study by the authors in [6]. The decision support system works by sending notifications to the user which also appears in the notification center in the case of situations that pose a risk.

In order to increase the efficiency of the pose estimation process, it is sufficient to determine only the eyes, ears, nose, shoulders, and neck to be used in the study. For this reason, a new data set containing the necessary body parts from the MPII Human Pose [7] dataset is curated. This study's customized pose estimation network is obtained by finetuning the model in [5] with the dataset, using the transfer learning method. Real-time posture analysis is performed using the obtained model. As a result of the analysis, situations that may decrease working efficiency are notified to the user in real-time to increase the user's well-being. In order for the study to be easily used by everyone, a user interface has been designed so that the user can customize it in every aspect. Improvements are made by making the designed interface used by people of different ages and professions.

2 Related Works

Object detection algorithms improved significantly with the help of AI [8]. Pose estimation is the detection of body parts in images or videos. Most studies [9–16] use a top-down approach for detection and estimation. As the first step in this approach, the person's part is detected, and then the body parts are estimated. Finally, pose estimation is made for each person. Although this approach applies to single-person detection, it cannot capture spatial dependencies between different people that require global inference. Studies [2] and [17] use a bottom-up approach in pose estimation. In this approach, all parts of each individual in the image are identified. Then, parts of different individuals are combined and grouped. Although this approach works slower than the top-down approach, it gives more accurate results because it can capture spatial dependencies between different people that require global inference [2]. Figure 1 presents an example of these approaches. For this reason, the bottom-up approach was preferred in this study.

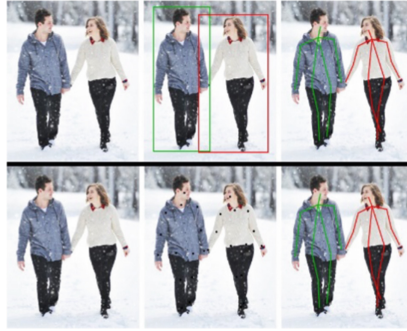


Fig. 1. Top: typical top-down approach. Down: typical bottom-up approach [26, 27].

3 Methodology

A real-time decision support system application is developed for the well-being of computer users. In the OpenPose [5] study, which has proven its success for the purpose, the pose estimation model was retrained using the transfer learning technique utilizing the customized MPII Human Pose [18] dataset.

3.1 Environment

The camera used for image transfer is the bisoncam nb pro, which is one of the models that comes internally in personal computers. The reason for choosing this model is to show that the system can work without any extra financial cost. This is a standard camera with a fixed focus, field of view of 60°, and maximum resolution of 720 p/30 fps. Nvidia Tesla T4 graphics card specially designed for AI applications was used to shorten the time for the model training phase. The high-end graphical processor has Turing architecture, 16 GB GDDR6 memory capacity, 320 Turing cores, and 2560 Cuda cores. For the testing phase of the pose estimation model, a system with similar characteristics to the average computer user was chosen. This system has Nvidia GTX 970 m graphics card, Intel I7-6700HQ 2.60 GHz central processor, and 16 GB DDR4L short-term memory.

3.2 Dataset

The pose detection model utilized the “MPII Human Pose” dataset via transfer learning. This dataset contains approximately 25000 images containing more than 40000 people with annotated body joints. Images were systematically collected using an established taxonomy of human activities each day. Overall, the dataset covers 410 human activities, and each image is provided with an activity tag [18]. In this study, only human exposure estimation was made without distinguishing activity.

3.3 Model Development

As with many bottom-up approaches, the proposed model first detects each person’s parts (key points) in the image, then assigns the parts to different people. As in Fig. 2,

the network is first used to extract features from the image using one of the models in studies [19–22] with its first few layers. A set of feature maps named F is created. Then F is fed into a network of two parallel convolutional layers. Stage-1 predicts 18 confidence maps, each representing a specific part of the human pose skeleton. At this stage, the network generates several Part Affinity Fields (PAFs) (1) $L^1 = \phi^1(F)$, where ϕ^1 refers to the CNNs for inference in Stage-1. The predictions from the previous stage and the original image features F are combined and used to produce refined predictions at each subsequent stage.

Stage-2 predicts a value of 38 PAFs, which represents the degree of relationship between the parts. The successive stages are used to correct the forecasts made by each branch. Two-sided graphs were created between pairs using part confidence maps. By using PAFs values, weak links in bilateral charts are trimmed, and pose skeletons are predicted.

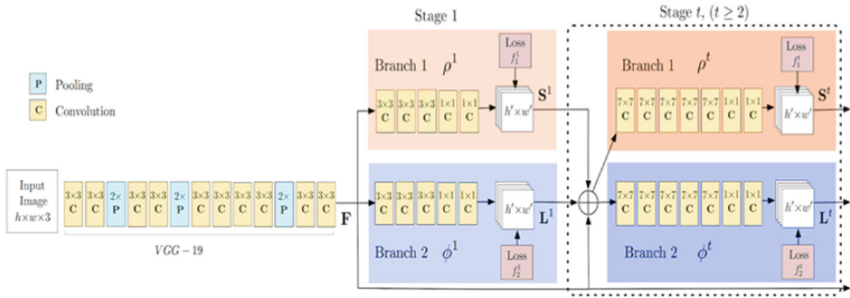


Fig. 2. Flowchart of OpenPose architecture [5].

$$L^t = \phi^\tau(F, L^{t-1}), \quad \forall 2 \leq \tau \leq T_P \quad (1)$$

In Eq. (1), ϕ^τ represent the CNNs for inference at stage t , and T_P is the total number of PAF stages. After T_P iterations, the process is repeated to detect confidence maps, starting with the most recent PAF estimate.

$$S^{T_P} = \rho^\tau(F, L^{T_P}), \quad \forall \tau = T_P \quad (2)$$

$$S^\tau = \rho^\tau(F, L^{T_P}, S^{\tau-1}), \quad \forall T_P < \tau \leq T_P + T_C \quad (3)$$

In Eq. (3), ρ^τ represents the number of CNNs for inference at stage t , and T_C is the number of total confidence map stages.

Confidence map results are estimated above the most recent processed PAF estimates, making a barely noticeable difference in confidence map stages. A loss function is applied at the end of each stage to direct the network to iteratively estimate the PAF values of the body parts in the first stage and the confidence maps in the second stage. L2 loss is used between the expected forecasts and the baseline information maps and fields.

$$f_L^{t_i} = \sum_{c=1}^C \sum_p W(p) \cdot \|L_c^t(p) - L_c(p)\|_2^2 \quad (4)$$

$$f_S^{tk} = \sum_{j=1}^J \sum_p W(p) \cdot \|S_c^{tk}(p) - S_j(p)\|_2^2, \quad (5)$$

Equation (5) has spatially weighted the loss functions to address a practical problem where some datasets do not thoroughly label all people. Where Eq. (4) L_c is the PAF actual reference value (ground truth), S_j is the true confidence map, and is a binary mask with $W(p) = 0$ when p is missing annotation on the pixel. The mask is used during training to avoid penalizing true positive guesses. Interim inspection at each stage fills the gradient periodically, eliminating the vanishing gradient problem.

To increase the efficiency of the study, it was foreseen that it would be sufficient to determine only the eyes, ears, nose, shoulders, and neck ending to be used in the study. In order to detect only these body parts, a new data set consisting of the necessary body parts in the data set was created. With this dataset, the model was retrained using the transfer learning method to benefit the pre-trained weights of the MobilNetV2 [22].

For conditions that may pose a risk of disease, the angle values between the limbs were taken as a reference [6]. The reference values can be changed by $\pm 30\%$ with the slide bar in the program's interface and by $\pm 250\%$ in the advanced settings menu. When the values calculated due to exposure estimation went out of the reference value limits, the system took 50 samples at equal intervals for 10 s, and the average was calculated. The calculated value and the reference values are compared. When there is a situation that may pose a disease risk, the user is notified in real-time with notifications as in Fig. 3.

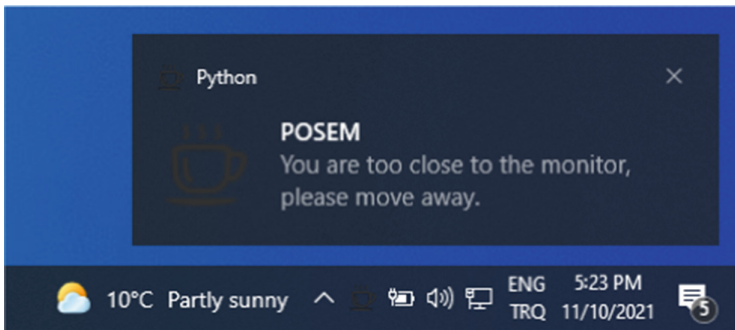


Fig. 3. Notification example of work in Windows 10 environment

3.3.1 Creation of the User Interface

The target audience of this study is computer users, and the user interface has been designed so that everyone can use it easily. In addition, the work has been made saved as a file with a.exe extension. The designed interface is designed in such a way that the user can customize it in every aspect, as seen in Fig. 4. The interface's main features are the selection of the posture tests, customization of wrong posture reference values, and customization of notification frequency. The settings also have five different profiles to save the made customizations.

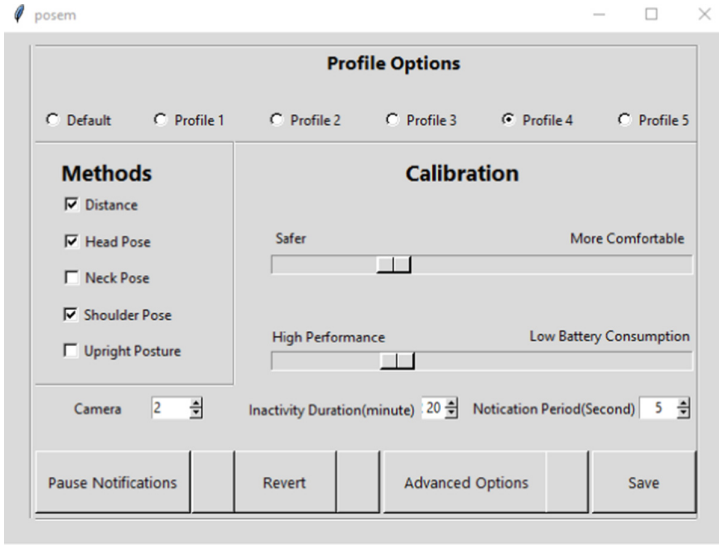


Fig. 4. Improved user interface

4 Results and Discussions

Using the images reserved from the dataset for testing, the high performed models, i.e., PersonLab [23], METU [24], Associative Emb. [25], and OpenPose [5] models were compared with the proposed method in Table 1 with respect to accuracy and precision. Since the most critical parameter for the applicability of the study is speed, the proposed method performed better than other methods.

Table 1. Performance comparison of the proposed method and the methods in the literature.

Models	Performance metrics		
	Accuracy	Precision	Speed (fps)
PersonLab	0.687	0.877	42.8
METU	0.705	0.890	53.6
Associative Emb.	0.655	0.868	31.4
OpenPose	0.642	0.862	68.2
Developed method	0.679	0.871	75.1

5 Conclusion

Health is one of the significant elements that reduce efficiency in organizations. In addition, one of the most critical expense items in most countries is healthcare. This study aimed to contribute to the country's economy by identifying situations that reduce working efficiency and quality of life. In order to achieve this aim, an optimized pose estimation model is proposed using the transfer learning technique. With the developed application, posture disorder analysis will be performed, and health problems that may occur in the waist, neck, and joint regions will be prevented. Visual disturbances will be prevented by analyzing the distance to the monitor, working environment lighting, and usage time. With the help of the proposed model, the work environment will be analyzed dynamically to support healthy working conditions.

References

1. Sarp, S., Demirhan, H., Akca, A., Balki, F., Ceylan, S.: Work in progress: activating computational thinking by engineering and coding activities through distance education. In: 2021 ASEE Virtual Annual Conference Content Access, July 2021
2. Szeto, G.P., Straker, L., Raine, S.: A field comparison of neck and shoulder postures in symptomatic and asymptomatic office workers. *Appl. Ergon.* **33**(1), 75–84 (2002)
3. Sarp, S., Kuzlu, M., Cali, U., Elma, O., Guler, O.: An interpretable solar photovoltaic power generation forecasting approach using an explainable artificial intelligence tool. In: 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1–5. IEEE, February 2021
4. Sarp, S., Kuzlu, M., Wilson, E., Cali, U., Guler, O.: The enlightening role of explainable artificial intelligence in chronic wound classification. *Electronics* **10**(12), 1406 (2021)
5. Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y.: Real-time multi-person 2D pose estimation using part affinity fields. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7291–7299 (2017)
6. Wahlström, J.: Ergonomics, musculoskeletal disorders and computer work. *Occup. Med.* **55**(3), 168–176 (2005)
7. Hidalgo, G., Cao, Z., Simon, T., Wei, S.-E., Joo, H., Sheikh, Y.: OpenPose library. <https://github.com/CMU-Perceptual-Computing-Lab/openpose>
8. Sarp, S., Kuzlu, M., Zhao, Y., Cetin, M., Guler, O.: A comparison of deep learning algorithms on image data for detecting floodwater on roadways. *Comput. Sci. Inf. Syst.* **19**(1), 397–414 (2022)
9. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: The IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969 (2017)
10. Fang, H.-S., Xie, S., Tai, Y.-W., Lu, C.: RMPE: regional multi-person pose estimation. In: The IEEE International Conference on Computer Vision (ICCV), pp. 2334–2343 (2017)
11. Pishchulin, L., Jain, A., Andriluka, M., Thormählen, T., Schiele, B.: Articulated people detection and pose estimation: reshaping the future. In: CVPR (2012)
12. Gkioxari, G., Hariharan, B., Girshick, R., Malik, J.: Using k-poselets for detecting people and localizing their keypoints. In: CVPR, pp. 3582–3589 (2014)
13. Papandreou, G., et al.: Towards accurate multi-person pose estimation in the wild. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4903–4911 (2017)

14. Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., Sun, J.: Cascaded pyramid network for multi-person pose estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7103–7112 (2018)
15. Xiao, B., Wu, H., Wei, Y.: Simple baselines for human pose estimation and tracking. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11210, pp. 472–487. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01231-1_29
16. Sarp, S., Kuzlu, M., Cetin, M., Sazara, C., Guler, O.: Detecting floodwater on roadways from image data using Mask-R-CNN. In: 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 1–6. IEEE, August 2020
17. Pishchulin, L., et al.: DeepCut: joint subset partition and labeling for multi person pose estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4929–4937 (2016)
18. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2D human pose estimation: new benchmark and state of the art analysis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3686–3693 (2014)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
20. Szegedy, C., et al.: Going deeper with convolutions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
22. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: MobileNetV2: inverted residuals and linear bottlenecks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520 (2018). (mobilnet)
23. Papandreou, G., Zhu, T., Chen, L.-C., Gidaris, S., Tompson, J., Murphy, K.: PersonLab: person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. LNCS, vol. 11218, pp. 282–299. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01264-9_17
24. Kocabas, M., Karagoz, S., Akbas, E.: MultiPoseNet: fast multi-person pose estimation using pose residual network. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11215, pp. 437–453. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01252-6_26
25. Newell, A., Huang, Z., Deng, J.: Associative embedding: end-to-end learning for joint detection and grouping. In: NIPS (2017)
26. Raj, B., Osin, Y.: An overview of human pose estimation with deep learning (2019). <https://www.kdnuggets.com/2019/06/human-pose-estimation-deep-learning.html>. Accessed 1 June 2021
27. Bogdanov, Y.: Understanding couple walking on snow near trees during daytime, 1 January 2018. https://unsplash.com/photos/XuN44TajBGo?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText. Accessed 1 June 2021

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

