

A Conditional Variational Auto-encoder Model for Reducing Musculoskeletal Disorder Risk during a Human-Robot Collaboration Task

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2023, Vol. 67(1) 425–431
Copyright © 2023 Human Factors and Ergonomics Society
DOI: 10.1177/21695067231192538
journals.sagepub.com/home/pro



Liwei Qing¹, Bingyi Su¹, Ziyang Xie¹, Sehee Jung¹, Lu Lu¹, Hanwen Wang¹, Xu Xu¹, and Edward P. Fitts¹

Abstract

In recent years, there has been a trend to adopt human-robot collaboration (HRC) in the industry. In previous studies, computer vision-aided human pose reconstruction is applied to find the optimal position of point of operation in HRC that can reduce workers' musculoskeletal disorder (MSD) risks due to awkward working postures. However, the reconstruction of human pose through computer-vision may fail due to the complexity of the workplace environment. In this study, we propose a data-driven method for optimizing the position of point of operation during HRC. A conditional variational auto-encoder (cVAE) model-based approach is adopted, which includes three steps. First, a cVAE model was trained using an open-access multimodal human posture dataset. After training, this model can output a simulated worker posture of which the hand position can reach a given position of point of operation. Next, an awkward posture score is calculated to evaluate MSD risks associated with the generated postures with a variety of positions of point of operation. The position of point of operation that is associated with a minimum awkward posture score is then selected for an HRC task. An experiment was conducted to validate the effectiveness of this method. According to the findings, the proposed method produced a point of operation position that was similar to the one chosen by participants through subjective selection, with an average difference of 4.5 cm.

Keywords

Human-robot interaction, Occupational safety, Robotics

Introduction

Musculoskeletal disorders (MSDs) are among the most common occupational injuries, which affect muscles, nerves, tendons, joints, and cartilage (Gallagher & Schall Jr, 2017). Epidemiology studies have shown that risk factors for MSDs include repetitive motions, awkward postures, and excessive force exertion over a prolonged work period (Kamat et al., 2017). Therefore, improving workers' posture is crucial in reducing the risk of MSDs.

To improve workers' postures in different workplace settings, a number of observational methods have been made based on ergonomic assessments, such as Rapid Upper Limb Assessment (RULA) (McAtamney & Corlett, 1993) and Rapid Entire Body Assessment (REBA) (Hignett & McAtamney, 2000). Yet, manually performing these observation methods can be extremely labor-intensive. Thus, there are also methods having been proposed to automatically obtain workers' posture. For example, joint angles can be measured by wearable sensors, such as inertial measurement units (IMUs). Yet, these wearable sensors can interfere with the natural body motions of workers during operation

(Ahmad et al., 2013). Alternatively, computer-vision algorithms can be utilized to infer human postures from images or videos. In the study conducted by Martinez (2017), human motion was well reconstructed using a simple and lightweight deep neural network. The fast-developing machine learning methods appear to be capable to form a more efficient way of understanding and improving workers' postures (Li, et al., 2020; Wang et al., 2021). However, computer vision-aided posture assessment heavily relies on the workplace configuration and hardware. Obstacle in camera view as well as lens distortion can affect the accuracy of reconstructed poses (Chakraborty et al., 2018).

Along with direct measurements, workers' postures also can be simulated through biomechanical modeling and

¹Department of Industrial & Systems Engineering, North Carolina State University, Raleigh, NC, USA

Corresponding Author:

Liwei Qing, Department of Industrial & Systems Engineering, North Carolina State University, 915 Partners Way, Raleigh, NC 27695-7001, USA.

Email: lqing@ncsu.edu

optimization. For instance, some researchers studied joint angle acquisition from an inverse kinematics perspective. In El Makrini et al. (2022), the simulated pose was obtained by solving the ordinary differential equations (ODE) of a linkage model representing human body. An earlier example of analytical human modeling can be found at Tolani et al. (2000), where the authors solved a set of inverse kinematics equations suitable for different body dimensions. Yet, there could be an infinite number of feasible poses to fit a kinematics chain and thus an objective function needs to be adopted. Although biomechanical modeling and optimization can provide plausible predictions of joint angles as well as muscle force for certain occupational tasks, their applicability in a real-world scenario remains a matter of concern. This is because most of these optimization-based methods are only validated in a strictly controlled simulation environment (Borbély & Szolgay, 2017).

Human-robot collaboration (HRC) is a burgeoning workplace configuration in recent years and brings great opportunities for a wide range of industrial sectors, including manufacturing, healthcare (Olaronke et al., 2017), and agriculture (Vasconez et al., 2019). This work configuration takes advantage of the endurance of robots and the dexterity and ability of humans to react to unpredicted environments. Yet, the HRC can also present occupational safety and health challenges. For example, when a robot is designed to pass an object to a worker, the predetermined one-size-fit-all point of operation position (where workers pick up the object) may lead to awkward postures for workers with different anthropometry dimensions, thereby increasing the risk of developing MSDs (Anita et al., 2014).

To improve workers' postures in HRC environments, one can adopt a reinforcement learning approach (Xie et al., 2022) to automatically change the point of operation position for each individual workers. Specifically, a computer vision technique was adopted to directly obtain workers' postures and the associated MSD risk score. The location of robot end effector then moved gradually in a way to alter workers' posture with a goal to minimize the associated MSD risk score. While this work can improve individual workers' postures by providing personalized work configurations, the proposed method may fail in real-world scenarios because it requires a well-arranged working environment where cameras are placed without any occlusions.

In this study, we propose a novel data-driven generative model to simulate workers' postures in an HRC task and use these simulated postures to find the optimal point of operation position. Generative modeling aims to train algorithms to synthesize new data by learning latent representations of the input data and then sampling from the latent space (Goodfellow et al., 2020; Kingma & Welling, 2013; Mirza & Osindero, 2014). In particular, a conditional variational auto-encoder (cVAE) model is trained in this study using an open-access human posture dataset (Li, et al., 2020). The trained model can output simulated yet plausible worker postures

with diverse combinations of body sizes and task constraints (e.g., simulate a posture of a 1.75m tall worker with hand position at a specific height). Next, an awkward posture score is calculated to evaluate the MSD risks associated with all generated postures. The posture associated with a minimum score is then chosen for setting the point of operation position for an HRC task. The effectiveness of this data-driven method was verified an experiment by comparing the yielded point of operation positions with the participants' self-selected ones.

Method

The pipeline of minimizing workers' MSD risk scores includes the following components: cVAE algorithm, posture scoring, and algorithm validation, as presented in Figure 1. We implemented the cVAE algorithm using the method outlined in Li et al. (2021). For readers' convenience, we present a brief overview of this method here, with detailed information available in the reference paper.

cVAE Algorithm

Dataset Preparation for Training. The training dataset used in this study contains full-body anatomical landmark positions of 11 participants during a variety of lifting motions, as illustrated in Figure 2a. In these lifting tasks, participants were asked to lift a box from the ground to different levels of a shelf, with the lowest level around the floor height and the highest level of 1.63 m. The 3-D coordinate of 35 markers placed on their body (Figure 2a) was measured by an optical motion tracking system (MotionAnalysis).

Model. Our posture optimization pipeline leveraged a cVAE model, an extension of the standard auto-encoder and variational auto-encoder that provides generative capabilities to the latent space. During training, the loss function for cVAE comprised a "reconstruction term," which minimized the input-output differences, and a "regularization term," which approximated the latent distribution to a standard normal distribution. The total loss function is expressed as:

$$loss = \|x - \hat{x}\|^2 + KL[N(\mu_x, \sigma_x)N(0,1)] \quad (1)$$

Where x represents the input data, $\hat{x}$ represents the generated data, μ_x and σ_x are the mean and standard deviation of the encoded distribution, and $KL[N(\mu_x, \sigma_x), N(0, 1)]$ is the Kulback-Leibler divergence between the encoded distribution and a standard Gaussian. Through minimizing the loss (1) during the training process, the model can effectively generate plausible human postures.

To impose constraints over the simulated postures (e.g., the simulated posture needs to have a specific hand height), the reconstruction term of a cVAE was modified to incorporate the pre-defined conditions c , expressed as $\hat{x} = P(x|z, c)$.

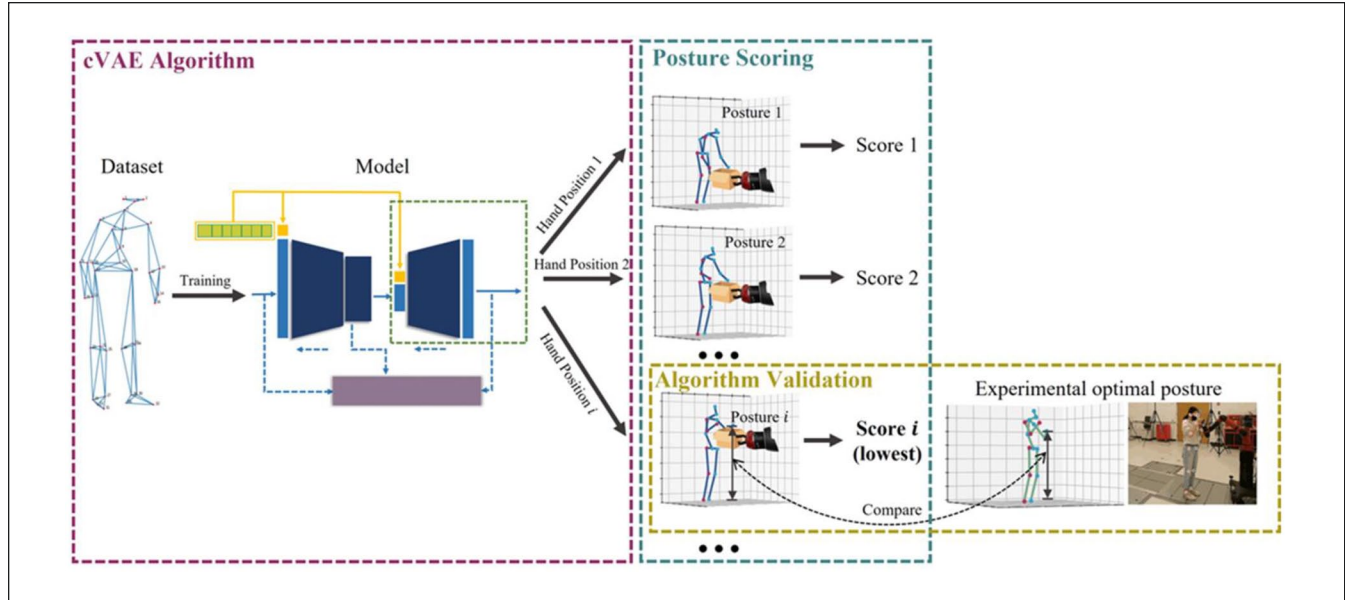


Figure 1. Pipeline of minimizing workers' MSD risk scores via a conditional variational auto-encoder (cVAE). The pipeline consists of three components: (1) cVAE algorithm for training a posture generation network, (2) Posture Scoring module for online posture evaluation and selection, and (3) Algorithm Validation for verifying the selected optimal position against the self-selected point of operation position.

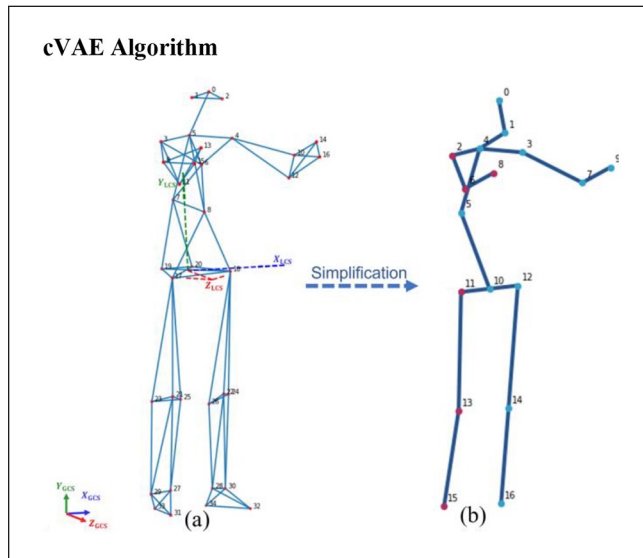


Figure 2. Dataset illustration. (a) Marker placement details. (b) A simplified skeletal model with 17 key points.

The cVAE algorithm consisted of one encoder network and one decoder network, with the encoding process expected to be the reverse process of the decoding process. During inference, sample z was randomly taken from the trained encoded distribution $N(\mu_x, \sigma_x)$, and both c and z were fed into the decoder. After training, the decoder converged to a conditioned latent space, allowing for control over the output

simulated postures. Figure 3 illustrates the structure of our cVAE algorithm. In this study, we formulated the posture generation problem as predicting the human posture given the location of both hands, both feet, and the body height (7 constraints in total). The left-hand location was described by two parameters, H_{lh} and W_{lh} . H_{lh} is the vertical distance between the hand (mid of radial styloid and ulnar styloid) and body center (mid of anterior superior iliac spine and posterior superior iliac spine), while W_{lh} is the horizontal distance between the hand and body center. The left-foot location W_{lf} was interpreted by the horizontal distance between the feet (mid of medial malleolus and lateral malleolus) and body center. The right hand and right foot were similarly defined. Therefore, the condition c belonged to a 7-dimensional space $R^7 : (Height, H_{lh}, W_{lh}, H_{rh}, W_{rh}, W_{rf}, W_{lf})$. The input x and output postures $\hat{x}$ belonged to 105-dimensional space R^{105} (35 landmarks $\times$ 3 dimensions).

Posture Scoring

This study employed a method similar to RULA (McAtamney & Corlett, 1993) for assessing posture safety and health. RULA is an ergonomic assessment tool used to evaluate the risk of MSDs based on a systematic observation of posture and associated body discomfort, with a scoring system that guides user to identify high-risk postures. As depicted in Figure 4, we utilized continuous linear functions to interpolate the original RULA score for posture scoring. The original discrete RULA has limited sensitivity when different

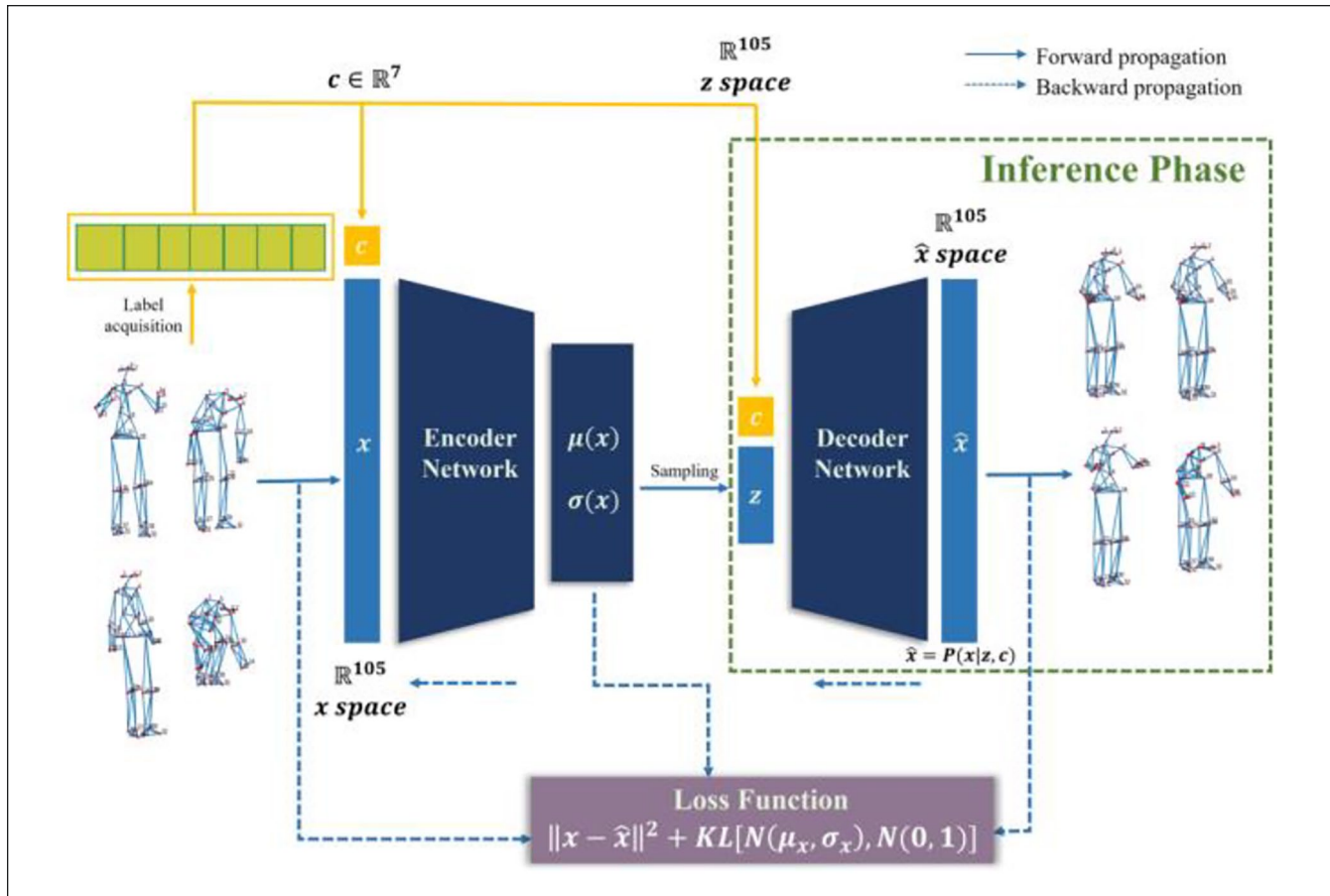


Figure 3. Structure of the proposed cVAE model.

joint angles fall within the same range, which poses a challenge when determining an optimal robot position, as different human postures may yield the same RULA score. Additionally, the original RULA score employs a step function that can lead to sudden score jumps over similar postures, which may not accurately reflect the actual MSD risk changes among similar postures.

To map a posture to a score, the simplified skeleton model was used. The simulated 35-key-point skeleton was simplified to a skeleton with only 17 key points (Figure 2b). Such a simplified skeleton are widely adopted for pose representation in previous studies (El Makrini et al., 2022; Martinez et al., 2017; Pavllo et al., 2019). Additionally, penalties are applied for squatting and unbalanced posture. Specifically, when the angle of the worker's upper leg or knee exceeds 15 degrees (indicating squatting) or when the discrepancy between the left and right knee angles is greater than 15 degrees (indicating unbalance), a penalty score of 4 is assigned. The score for each joint is then aggregated to derive an overall posture score, with lower scores indicating lower risk of MSDs. Consequently, the posture among all simulated postures with the lowest score is identified as the posture for setting the optimal point of operation position of an HRC task.

Algorithm Validation

An HRC task were performed to verify the proposed algorithm for setting the point of operation position. In this task, participants were asked to stand in front of a robot and perform a fine assembly task together with a robot partner. Specifically, they needed to insert a fine wire into a board held by the robot. The initial position of the end-effector was randomly assigned at a low position and subsequently adjusted by the participants themselves (limited to the vertical axis) based on their perceived comfort. Once the operation position was determined, the participants performed the wire insertion task three times under each location as well as the initial end effector location. The adjusted end-effector position was recorded as it was considered as the reference optimal position from a psychophysical perspective (Snook & Ciriello, 1991). To find the optimal point of operation position that yields the posture with lowest score through the proposed cVAE method, we simulated human workers' posture at different hand height from low to high and obtained corresponding postural scores for each hand height. We then found the hand height with the minimal postural score and used this height as the optimal point of operation position.

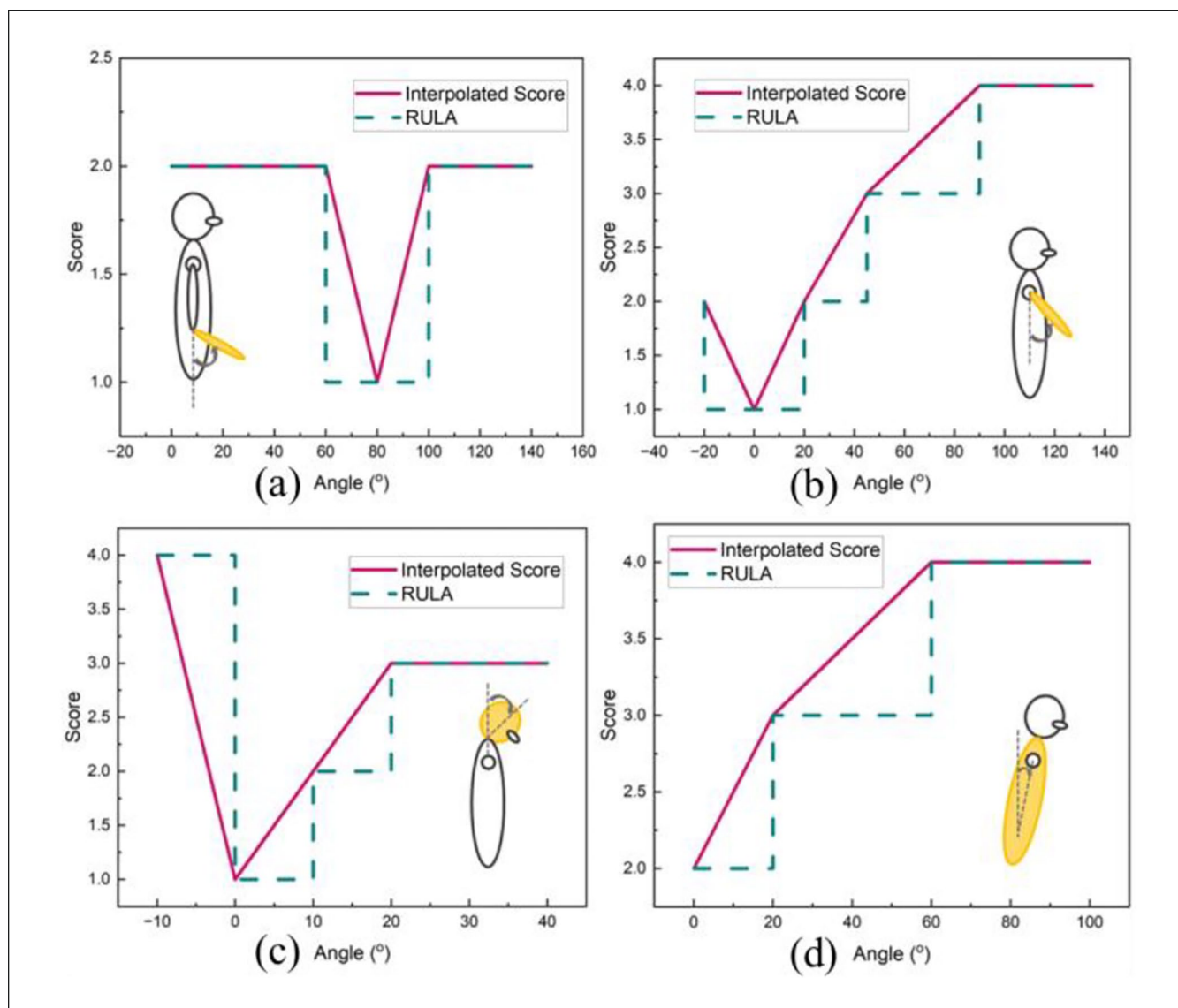


Figure 4. Interpolated posture scoring functions. (a) Lower arm. (b) Upper arm. (c) Neck. (d) Trunk.

To validate this optimal point of operation position generated from the cVAE method, eight participants (4 males and 4 females; age 26.0 ± 3.5 years; height 1.72 ± 0.12 m; weight 75.13 ± 16.72 kg) were recruited. They were asked to adjust the end-effector to their most comfortable position to perform the task as if they needed to perform this task for 8 hours a day. The initial robot end effector (i.e., robot gripper) position in the vertical direction was set at a very low position (1.055 m above the ground). This was to avoid participants simply accepting the initial position as an acceptable position. A webcam (Model: MF920P, Spedal) was placed three meters away from the robot to capture the images of the participants during experiments. An open-sourced VideoPose3D model (Pavlo et al., 2019) was applied to reconstruct participants' body posture.

Results

Figure 5a shows a series of simulated human postures at different hand height where the fine assembly HRC task is performed. All these simulated postures were generated by the proposed cVAE model. Each posture was then mapped to a posture score (y-axis of Figure 5a). In this example, the posture with the minimal score occurs when the hand is 1.38 m above the ground.

We also validated our approach by comparing the point of operation position yielded from the cVAE model and that yielded from participant's self-selection for each participant. The middle and right subfigures in Figure 5b show a reconstructed human pose through VideoPose3d when a participant is doing the assembly task at the self-selected position.

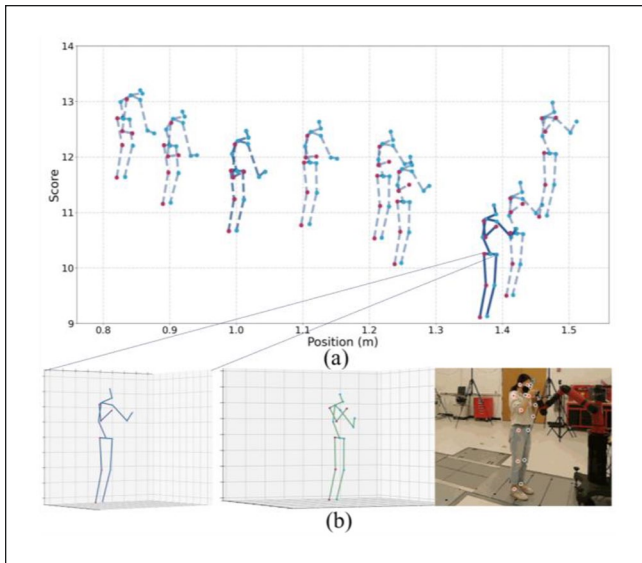


Figure 5. (a) Simulated postures with different hand height. The solid skeleton represents the posture with the minimal posture score. (b) Comparison between the simulated-based posture and the posture observed during the experiment. Left: the simulated posture with the minimal posture score; Middle and Right: subjectively selected posture during the experiment.

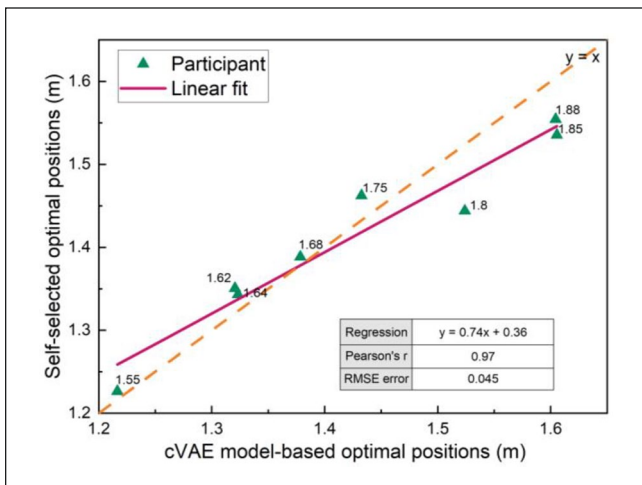


Figure 6. Comparison between the cVAE model-based optimal point of operation positions and self-selected optimal positions among all participants. Numbers by each green triangle markers show participants' height. The red solid line depicts the linear regression, while the orange line is the diagonal.

The simulated body postures with minimal score are similar to the self-selected postures observed during the experiment. As shown in Figure 6, the correlation coefficient (Pearson's r) is 0.97 between the point of operation position yielded from the two methods, and the RMSE is 0.045 m.

Discussion

In this study, we present a cVAE-based method for predicting a point of operation position during an HRC task for workers with different body height. This predicted position is associated with a less risk of MSDs. This method can improve workers' musculoskeletal health by providing personalized work configurations while avoiding direct measurement of human pose at work.

While our study presents a promising approach to optimizing job configurations for HRC tasks, it is important to address some limitations. First, the human posture data used to train the cVAE model were collected from lifting tasks, where most participants chose symmetric lifting motions to lift a regular rectangle box. We chose this dataset for its ready accessibility. With the addition of constraints, it serves as a viable resource for simulating postures in the current application. Yet, the generalizability of the current cVAE model remains unclear. It is questionable whether the current cVAE model can simulate any asymmetric postures or motions. Future study should consider training a more robust cVAE model with a wide range of body motion data. Second, we admit that the reconstructed body pose through VideoPose3D, which is used to calculate the reference point of operation position, may carry errors. A previous study showed that the joint location error due to computer-vision algorithm is around 7.7% (Wang et al., 2021). For a more precise validation, a laboratory-grade motion tracking system would be needed. Third, it is possible that the subjectively selected position of point of operation during a fine assembly task does not necessarily lead to a posture associated with the minimal muscle activities or joint forces. Other subjective concerns, such as the distance from eyes to the point of operation, may also play an important role in decision-making. Further research is needed to develop an objective method for determining the reference body posture that can minimize the risk of MSDs.

Conclusion

By using a data-driven approach and highlighting the importance of preventing MSDs, this study paves the way for future research in HRC task design. The proposed cVAE model was trained using an open-access posture dataset with 7 labels constraining the output and can simulate human postures. The MSD risk of the simulated postures can be evaluated by a posture scoring protocol that is similar to RULA. For an HRC assembly task, the optimal point of operation position is selected based on the MSD risk assessment among all simulated postures. This optimal position was verified against the self-selected point of operation position. The comparison shows the effectiveness of the proposed method in reducing MSD risks during an HRC task.

Acknowledgments

This manuscript is based upon work supported by the National Science Foundation under Grant #2024688.

ORCID iDs

Bingyi Su  <https://orcid.org/0009-0008-0806-8890>

Sehee Jung  <https://orcid.org/0000-0001-7980-3911>

References

- Ahmad, N., Ghazilla, R. A. R., Khairi, N. M., & Kasi, V. (2013). Reviews on various inertial measurement unit (IMU) sensor applications. *International Journal of Signal Processing Systems*, 1(2), 256–262.
- Anita, A. R., Yazdani, A., Hayati, K. S., & Adon, M. Y. (2014). Association between awkward posture and musculoskeletal disorders (MSD) among assembly line workers in an automotive industry. *Malaysian Journal of Medicine and Health Sciences*, 23–28.
- Borbély, B. J., & Szolgay, P. (2017). Real-time inverse kinematics for the upper limb: a model-based algorithm using segment orientations. *Biomedical Engineering Online*, 16(1), 1–29.
- Chakraborty, B. K., Sarma, D., Bhuyan, M. K., & MacDorman, K. F. (2018). Review of constraints on vision-based gesture recognition for human–computer interaction. *IET Computer Vision*, 12(1), 3–15.
- El Makrini, I., Mathijssen, G., Verhaegen, S., Verstraten, T., & Vanderborght, B. (2022). A Virtual Element-Based Postural Optimization Method for Improved Ergonomics During Human-Robot Collaboration. *IEEE Transactions on Automation Science and Engineering*, 19(3), 1772–1783.
- Gallagher, S., & Schall, M. C., Jr. (2017). Musculoskeletal disorders as a fatigue failure process: evidence, implications and research needs. *Ergonomics*, 60(2), 255–269.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144.
- Hignett, S., & McAtamney, L. (2000). Rapid entire body assessment (REBA). *Applied Ergonomics*, 31(2), 201–205.
- Kamat, S. R., Zula, N. E. N. M., Rayme, N. S., Shamsuddin, S., & Husain, K. (2017). The ergonomics body posture on repetitive and heavy lifting activities of workers in aerospace manufacturing warehouse. *IOP Conference Series: Materials Science and Engineering*, 210(1), 12079.
- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *ArXiv Preprint ArXiv:1312.6114*.
- Li, L., Chen, K., Chen, K., & Xu, X. (2020). A depth camera-based full-body reconstruction method for body pose training in occupational safety. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1), 1541–1545.
- Li, L., Prabhu, S., Xie, Z., Wang, H., Lu, L., & Xu, X. (2021). Lifting posture prediction with generative models for improving occupational safety. *IEEE Transactions on Human-Machine Systems*, 51(5), 494–503.
- Li, L., Xie, Z., & Xu, X. (2020). MOPED25: A multimodal dataset of full-body pose and motion in occupational tasks. *Journal of Biomechanics*, 113, 110086.
- Martinez, J., Hossain, R., Romero, J., & Little, J. J. (2017). A simple yet effective baseline for 3d human pose estimation. *Proceedings of the IEEE International Conference on Computer Vision*, 2640–2649.
- McAtamney, L., & Corlett, E. N. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2), 91–99.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *ArXiv Preprint ArXiv:1411.1784*.
- Olaronke, I., Ojerinde, O., & Ikono, R. (2017). State Of The Art: A Study of Human-Robot Interaction in Healthcare. *International Journal of Information Engineering and Electronic Business*, 3, 43–55.
- Pavlo, D., Feichtenhofer, C., Grangier, D., & Auli, M. (2019). 3d human pose estimation in video with temporal convolutions and semi-supervised training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7753–7762.
- Snook, S. H., & Ciriello, V. M. (1991). The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34(9), 1197–1213.
- Tolani, D., Goswami, A., & Badler, N. I. (2000). Real-Time Inverse Kinematics Techniques for Anthropomorphic Limbs. *Graphical Models*, 62(5), 353–388.
- Vasconez, J. P., Kantor, G. A., & Cheein, F. A. A. (2019). Human–robot interaction in agriculture: A survey and current challenges. *Biosystems Engineering*, 179, 35–48.
- Wang, H., Xie, Z., Lu, L., Li, L., Xu, X., & Fitts, E. P. (2021). A Single-Camera Computer Vision-Based Method for 3D L5/S1 Moment Estimation During Lifting Tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1), 472–476.
- Xie, Z., Lu, L., Wang, H., Su, B., Liu, Y., & Xu, X. (2022). Mitigating the risk of musculoskeletal disorders during human robot collaboration: a reinforcement learning approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1), 1543–1547.