Design of an ergonomic App for entire rapid body assessment based on Mask RCNN

Shuai Wu¹, Ziwen Chen¹, Xinqiao Zhao¹, Maoyun Yao¹, Zihao Wang¹ and Shaolong Kuang¹,²,³

¹School of Mechanical and Electrical Engineering, Suzhou University, Suzhou, China
²Jiangsu Provincial Key Laboratory of Advanced Robot Technology, Suzhou University, Jiangsu
³E-mail: slkuang@suda.edu.cn

Abstract. The rapid entire body assessment (REBA) is a rapid and semi-quantitative ergonomic assessment method for those who engaged in manual handling and/or standing work for a long time. In order to provide a self-assessment tool for operators, an App software based on Mask RCNN is proposed in this paper. The software is developed by adopting the architecture of a mobile terminal App combined with an imaging processing server. The main functions of the App are video capturing of work process, the operation of the process of REBA, human-machine interaction, etc., while the server is work for processing the video imaging transmitted from the App, key-points extraction of worker’s body from the images for work posture identification. One thousand working scene photos marked by VIA are used for training and testing based on the Microsoft COCO dataset to obtain a reliable target detection model. Experiment of container handling scene shows that the App evaluation software has achieved higher evaluation efficiency and accuracy. The validation of this method has been proved compared with manual evaluation based on REBA. And other work posture evaluation methods will be developed in the future to form an ergonomic evaluation software system.

1. Introduction

At present, physical-dominated work and overtime work are prevalent in developing countries. Under this condition, work-related musculoskeletal diseases (WMSDs) have becoming one of the important reasons for workers' disability and productivity decline in these countries [1]. Research shows that factors associated with the workplace, personal psychology and social risk factors are the mainly contributions of the WMSDs. To prevent the occurrence of the WMSDs, many effective ergonomic assessment methods were developed for evaluating the work process. And people have great interest in evaluating the risk factors related to WSMDs and subsequent ergonomic interventions in the workplace [2]. At present, the methods for evaluating risk factors of WMSDs include self-evaluation, work posture evaluation and direct measurement [3], etc. Comparing with other methods, the self-assessment method is usually less accurate, and has not been widely used. The direct measurement method, which is using the wearable devices, obtains workers' working posture information directly. Although direct measurement could provide higher precision data, due to the influence of working environment, some wearable devices limit the freedom of movement. And due to technical and financial limitations, the direct measurement method is a niche product too [4]. As the work posture evaluation method is semi-quantitative and can be easily used with accuracy in most manual workplaces, it plays an important role for ergonomic evaluation.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
Several methods, like rapid upper limb assessment (RULA) [5], rapid entire body assessment (REBA) [6], ovako working posture analysis system (OWAS) [7] and concise exposure index (OCRA index) [8], etc., are the most often used techniques in the industry. Despite their simplicity, the effectiveness of these methods has largely depended on the accuracy of input variables. During evaluation process, when the researchers obtain data, they usually interrupt the operator's operation process. Moreover, the accuracy of the collected data is affected by the subjective factors of researchers with ambiguity, which limits the application of these methods in the environment of workshop site.

With the development of computer and software technology, several commercial ergonomic evaluation tools based on these methods have been developed, such as JACK, Double-E. The JACK is currently an excellent ergonomics software, which powerful functions of static and dynamic human-machine simulation. It evaluates the right posture of workers through human modeling and modeling the accurate virtual situations [9]. However, human modeling requires a lot of human and financial supported. With the different situations between different operators, it is difficult to implement human modeling for every worker, thus it is not suitable for real-time analysis of the workers on the job site. Double-E captures the images of each action, and analyzes the action through live video, which is more accurate than the data calculated by traditional IE (Industrial Engineering) [10]. But, Double-E could only perform manual segmentation and analysis on video. The operation of Double-E requires professional knowledge, and it is not efficient and intelligent enough. Moreover, all the evaluation tools mentioned above are expensive with skills and have great limits when they are sold to developing countries.

With the widely used smart phones, Apps have been the popular tools as a mode of application in many scenes [11]. If an ergonomic evaluation method can be realized to be an App, it will be changed to an easy used tool, that widely used for product and production site improving, and for enhance product competitiveness and corporate management capability.

Researches show that many people engaged in manual handling or standing for a long time have great contribution to overall fatigue and work efficiency of the operators, and the REBA method is a suitable ergonomics tool used in this field [6]. Therefore, an ergonomic App evaluation tool based on the REBA method is proposed and developed in this paper by using the most popular machine learning algorithm of Mask RCNN. It is feasible to perform ergonomic evaluation of the worker’s posture without disturbing the operation. This App can be applied to the labor-intensive operation sites of automobile manufacturing, transportation and other industries in order to meet the urgent needs of ordinary design engineers and enterprises.

Figure 1. REBA evaluation flowchart.
2. Rationale of REBA ergonomic evaluation method

The REBA method integrates the body posture, load, grasping tools, activity frequency and other factors of the work into an evaluation process, which set the corresponding scores of each factor by referring to the score query table, as shown in Figure 1. The larger the score, the higher harm of the operation posture of the body.

The detailed evaluation process of REBA method is as follows. Firstly, the scores of a body posture are calculated by dividing the body into six parts: neck, trunk, leg, lower arm, upper arm, and wrist, of which trunk, neck and leg are part A, and upper arm, lower arm and wrist are part B. Each part of the body has its score according to the posture during work. The score A and score B are obtained by referring to score A1 in Figure 2 and B1 in Figure 3 with correspondence pose respectively. And the score C and score D are obtained by adding load and grip score respectively. Then the score E is further obtained according to C1. Secondly, the REBA total score is obtained by combining activity frequency score. Finally, the risk level is determined by relationship between score and risk standard. The scores of a body posture of part A and part B, the load and grip scores are shown in Figure 4.

![Figure 2. Part A scoring process.](image1)

![Figure 3. Part B scoring process.](image2)

![Figure 4. Part A and Part B body part diagrams.](image3)
According to C1 in Figure 5, the overall scores of A and B are obtained from the total scores of part A and B, and the final score of REBA is calculated by adding activity frequency score C2. According to the score of REBA, the risk level can be determined by consulting the risk grade C3.

![Figure 5. REBA score and risk level.](image)

The result of REBA score ranges from 1 to 15 and sets the risk level from 1 to 4 as shown in Figure 5. In Figure 5 the measure for each risk level shall be taken is also listed.

3. Ergonomics evaluation App software design method

3.1. Software architecture design and information flow analysis

During evaluation processing, video of human motions is split into static image frames for key-points extraction. As the high computational resource is required for video image processing, the ergonomic evaluation program is divided into two parts: the android program (App Application) and server program (video file processing). The Android program takes the mobile phone as the carrier and is mainly responsible for interaction with users and video collection. The server program is mainly responsible for lots of images computing. The overall structure is shown in Figure 6.

![Figure 6. Overall structure diagram of ergonomic evaluation software.](image)

During the evaluation, the evaluator turns on the App software to record the complete working condition of the evaluated person in one working cycle and records the entire body movements in the working process. Then the video file is converted into a byte stream by the App and sent to the server. Server is used to read the byte stream and complete frame extraction. The deep learning model based on Mask RCNN of human joint recognition is used to detect the key points of the human body in the
working image and to obtain the angle information between the joints during the operation. Then, the image is recombined into video and the angle information of the images are collected and sent back to the App. Subsequently, the App evaluator manually checked, and supplemented the parameter information that cannot be obtained through server video analysis. Finally, the final score is calculated by using REBA method step by step in the android program and gives feedback opinion to the user.

3.2. App interface design
The design of the App interface mainly considers the efficiency and ease of operation, because the main groups targeted by this design are not only professional evaluators, but also workers from the work site. Therefore, the interface should be simple enough for convenient operating [12]. The App interfaces are designed including three parts: instruction, evaluation test, and result information. The main evaluation work is completed under the evaluation test interface, and the operator can click according to the actual situation. The overall functional composition is shown in Figure 7.

![Figure 7. Overall function frame diagram of App.](image)

3.3. Server-side human joint angle detection design based on Mask RCNN
The core process of App-based REBA evaluation is human key-points identification, which detects the human body and corresponding key-points position from the input images [13]. In the COCO data set, the key-points of the human body are represented as 17 joints, namely nose, left and right eyes, left and right ears, left and right shoulders, left and right elbows, left and right wrists, left and right hips, left and right knees, left and right ankles. To facilitate the processing, the server is used to transcode (MP4 to AVI) and split the original video frame uploaded from the App by using FFmpeg video processing tool and OpenCV image processing function library.

At the same time, in order to reduce the workload of data set production and improve the accuracy of human posture angle detection in the image, Mask RCNN method is adopted to realize the detection of human key points and obtain the angle information between human joint points [14].

3.3.1. Mask RCNN network structure design. Mask RCNN is a target segmentation detection framework proposed by He et al. [15] based on Faster R-CNN. Here, a two-stage Mask RCNN image processing framework is proposed, as shown in Figure 8.

The work flow of the proposed Mask RCNN method is as follows. The marked images of work site are sent to CONV Layers for processing to obtain feature maps, and then the existing Region of Interest (RoI) is extracted from the feature maps through a Region Proposal Network (RPN). Then the RoI is aligned so that the candidate frame coordinates predicted by RPN are mapped to the feature map. Finally, the RoI is output in three branches, including Mask branch obtained throughFully Convolutional Networks (FCN) sampling, Classification branch and Box Regressor branch obtained
through full connection layer. Transfer learning method is adopted to continue training while using FCN and full connection layer to mark human image targets.

![Figure 8. Mask RCNN network structure.](image)

CONV Layers uses the deep residual network ResNet101 to achieve the purpose of training deeper network layers. In order to get better integration of bottom-to-top feature maps, Feature Pyramid Network (FPN [16]) and the 5 outputs of ResNet101 are added together. The left bottom feature layer (C1-C5) is subjected to 1*1 convolution to obtain the same channel as the previous feature layer. The upper feature layer (P2-P5) is up-sampled to obtain the same length and width as the next feature layer and added to obtain a blended new feature layer. P2-P5 will be used to predict the bounding of the object in the future, P2-P6 will be used to train RPN, of which P6 is only used for RPN network.

The RPN is used to obtain RoI in feature images. Based on the structure of full convolution neural network, two convolution layers, cls-layer and reg-layer, are added [17]. In the design, a 256-dimensional fully connected feature is generated on feature maps by using a 3*3 sliding window, and then two cls-layer and reg-layer branches are generated by using the generated 256-dimensional fully connected feature to classify feature images and perform border regression.

Considering the large difference that may occur when feature maps are restored, RoI Align operation is adopted to fix the pixel values of four points coordinates through bilinear interpolation, thus making discontinuous operation continuous and achieving the purpose of reducing the error when the original image is returned. Firstly, RoI is divided into 7*7 bins, then bilinear interpolation is carried out on each bin, and max pooling is carried out after interpolation to obtain the final 7*7 RoI, thus completing the operation of RoI Align. Finally, these images through the Fully connected layers implement the Classification of objects and obtain the bounding box position of each class. The other path passes through FCN to realize Mask prediction of objects and complete detection of objects in images.

3.3.2. Data set making. Before training data using Mask-RCNN network, labeling of human joint points in working images needs to be completed manually. 1,000 working photos of operators were taken and selected from the work site, and the photos were labeled and stored in JSON format using
image label making software VIA. During the labeling process, information such as human joint position and Mask polygon node position were labeled with reference to the data format of Microsoft COCO data set [18]. Some labeling results are shown in Figure 9. Among them, 800 images will be used as training data for network training, and the remaining 200 images will be used as test sets to adjust the model parameters to improve the identification accuracy of human joint points.

![Figure 9. VIA labelled dataset.](image)

3.3.3. Human joint angle information detection training process. In order to simplify the training model, the transfer learning method is introduced into the fully connected layers and FCN of Mask-RCNN to continue training. The pre-training weight obtained by training the Mask-RCNN model on COCO data set is used as the initial weight first, then 800 labeled training set images are sent into the model, iterative training is carried out, and training parameters are continuously optimized, so as to improve the accuracy of model identification on the job site, and finally the target detection model is obtained.

Subsequently, 200 test sets are sent to the obtained target detection model for testing, and when the test does not meet the requirements, the model and training parameters are returned to be readjusted and retrained, until the model result with the final weight parameter optimization is obtained. Finally, the key points detected by the human body are linked together to obtain the required angle information of the joint points.

4. Implementation of REBA evaluation software

4.1. Software operation interface

Under Android Studio programming environment, Java 1.8.0_162 is used to for developing the mobile phone App. The software uses simple and vivid hand-drawn cartoon image of the worker, as the icons, and the final interface effect diagram that is shown in Figure 10.

The basic interface and operation process of the software are as follows: (1) the user shall click the software icon shown in Figure 10(A) of the desktop for login in, and then click to enter the Figure 10(B) login interface. If the user does not have an account and needs to register first. If there is an account the user can login. (2) Enter Figure 10(C) interface after login, it’s including three major functions: usage method, start test and history record. Users can click from left to right, such as Figure 10(D), to introduce the principle of REBA method and software usage, so as to facilitate users to understand the software and get started quickly. (3) Figure 10(E) is an introduction to software function; the processed operation video can be viewed in the video result, and the user can click PAUSE button to view and evaluate the posture at any time, as shown in Figure 10(F). The unidentified information in the video can be checked manually according to Figure 10(G). (4) Finally, clicking the confirm button, the software will automatically complete the evaluation. The software can complete continuous evaluation for many times. Users can view the evaluation values at any time in the historical records as required, as shown in Figure 10(H).
4.2. Pre-processing of video shooting in workplace

The investigator takes pictures of the working site by using the camera of the mobile phone. Before starting, the camera angle needs to be calibrated to ensure the accuracy of the test results. The person first adjusts the angle between the arm and the trunk to approximately 90 degrees, and then the investigator observes whether the angle between the arm and the trunk on the screen is 90 degrees with the help of grid lines on the screen, as shown in Figure 11.

4.3. Unidentified information processing

As the current algorithm cannot extract all human postures, an alternative solution similar to the traditional questionnaire is provided, which is completed through interactive modes such as user selection and manual check at App, as shown in Figure 12.
5. Experiments and results

5.1. Model training experimental and its results
The weights of the initial coco dataset before transfer learning were used to identify key-points, and the results of key-points identification that compared with the final model after training were shown in Figure 13. It can be seen that, both the lower arm of the left with gloves and the knee joint shielded by the milling machine can be successfully extracted, which proves the effectiveness of the method.

![Comparison of identification effect before and after transfer learning.](image)

5.2. App software test experiments and results
Here, a person of handling process is investigated. The person, who was moving a 25 kg box from the ground to a chair with a height of 50 cm for about 15 seconds, was captured by the designed App. Then the streaming video was transferred to the server for body posture key-points computing. After computing by the server, the angle of the body and the video synthesized from the processed pictures were sent to the mobile App. Feature images at certain moments acquired through video pause are shown in Figure 14 (A, B, C).

Since all critical postures must be evaluated in the workplace, during the experiment, the pause key was used to randomly select the posture at a certain time in the working video for evaluation, as shown in Figure 14(D). Result showed that the score of body posture under work condition is 10 after calculated by the ergonomic App, as shown in Figure 14(E) and Figure 14(F) in details. We know that this kind of work has a high-risk level, and the posture is harmful to the body. It is suggested to improve the posture as soon as possible.

On the other hand, 5 male and 5 female ergonomics undergraduates in our school were asked to conduct this evaluation as the investigators by manual calculating, which compared with the App to evaluate the work posture shown in Figure 14(D). Results of manual scoring, average value and ergonomic App scoring are shown in Figure 15.

It shows that 7 of 10 people's evaluation scores are equal to that of the software, and the other 3 people have a deviation of 1 point in some parts of the evaluation, with a relative deviation of 10%. The average value of 10 times results is 10.1, and the relative error compared with ergonomic App is 1%. In addition, 10 investigators take about 10 minutes to identify the angle from the images obtained by the video pause and then to complete the calculation, while it only takes about 30 seconds for ergonomic App to process them.

It also shows that the evaluation used by the ergonomic App has high efficiency and accuracy in automatic identification, and its advantages will be more obvious when the evaluation workload is in large scale.
6. Conclusions and prospects
An ergonomics evaluation method of REBA is introduced, designed, and implemented based on the Android App software. The method uses Mask RCNN to establish the identification and evaluation model of human posture angle, with mode of automatic identification combined with manual selection to obtain the operation posture to gain the most evaluation efficiency and accuracy. This method will not interfere with the operator's work during the data acquisition process. Experimental results show that the software evaluation results are reliable.

Moreover, this evaluation App is significant for labor-intensive enterprises in developing countries that do not pay much attention on ergonomics [19]. It provides a simple and self-evaluation software for employees and the investigators. For industrial engineers, it provides convenient tools for on-site improvement. It is of great significance of this kind of evaluation tool that used to build a harmonious
working atmosphere between man and machine, to avoid accidents such as cumulative injuries, and to save national medical resources. More ergonomics valuation methods will be developed to form an ergonomic evaluation software system in the future.

Acknowledgements
Work was financially supported by Guangdong Province Key Field R & D Plan Project: Research on Robot Perception and Human-Machine Integration Technology (2019B090915002).

References
[1] Abareshi F, Yarahmadi R and Solhi M 2015 Educational intervention for reducing work-related musculoskeletal disorders and promoting productivity Int. J. Occup. Saf. Ergon 21 480-85
[2] Occhipinti E and Colombini D 2007 Updating reference values and predictive models of the OCRA method in the risk assessment of work-related musculoskeletal disorders of the upper limbs Ergonomics 50 1727-39
[3] Chander D S and Cavatorta M P 2017 An observational method for postural ergonomic risk assessment (PERA) Int. J. Ind. Ergon 57 32-41
[4] Bao S, Howard N and Spielholz P 2007 Two posture analysis approaches and their application in a modified rapid upper limb assessment evaluation Ergonomics 50 2118-36
[5] Mcatamney L and Corlett E N 1993 RULA: a survey method for the investigation of work-related upper limb disorders Appl. Ergon 24 91-99
[6] Hignett S and Mcatamney L 2000 Rapid entire body assessment (REBA) Appl. Ergon 31 201-05
[7] Karhu O, Kansi P and Kuorinka I 1977 Correcting working postures in industry: A practical method for analysis Appl. Ergon 8 199-201
[8] Occhipinti E 1998 OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs Ergonomics 41 1290-1311
[9] Magistris G D, Micaelli A and Evrard P 2013 Dynamic control of DHM for ergonomic assessments Int. J. Ind. Ergon 43 170-80
[10] Ma Y Z, Gao C C and Xu Yan 2015 The informational work study based on software Double-E Val. Eng. 8 7-10
[11] Voas J, Michael J B and Van G M 2012 The mobile software app takeover IEEE Softw. 29 25-27
[12] Antrobus V, Burnett G and Krehl C 2017 Driver-passenger collaboration as a basis for human-machine interface design for vehicle navigation systems Ergonomics 60 321-32
[13] Barbosa J, Lee K and Lee S 2016 Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier BMC Med. Imaging 16 23
[14] He K, Gkioxari G, Dollár P and Girshick R 2020 Mask R-CNN IEEE Trans. Pattern Anal. Mach. Intell 42 386-97
[15] Ren S, He K, Girshick R and Sun J 2017 Faster R-CNN: towards real-time object detection with region proposal networks IEEE Trans. Pattern Anal. Mach. Intell 39 1137-49
[16] Woo S, Hwang S and Jang H D 2019 Gated bidirectional feature pyramid network for accurate one-shot detection Mach. Vision. Appl 6 543-55
[17] Sarikaya D, Corso J J and Guru K A 2017 Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection IEEE Trans. Med. Imag 367 1542-49
[18] Vinyals O, Toshev A and Bengio S 2016 Show and tell: lessons learned from the 2015 ms coco image captioning challenge IEEE Trans. Pattern Anal. Mach. Intell 394 652-63
[19] Golabchi A, Han S and Fayek A R 2016 A fuzzy logic approach to posture-based ergonomic analysis for field observation and assessment of construction manual operations Can. J. Civ. Eng 43 294-303