Edge Computing Driven Scene-aware Intelligent Augmented Reality for Manual Assembly

Mingyu Fu
Beijing University of Posts and Telecommunications
Wei Fang (fangwei@bupt.edu.cn)
Beijing University of Posts and Telecommunications
Shan Gao
Beijing University of Posts and Telecommunications
Jianhao Hong
Beijing University of Posts and Telecommunications
Yizhou Chen
Beijing University of Posts and Telecommunications

Research Article

Keywords: Intelligent augmented reality, scene-aware, edge computing, manual assembly, smart factory

Posted Date: July 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-736006/v1

Edge computing driven scene-aware intelligent augmented reality for manual assembly

Mingyu Fu, Wei Fang*, Shan Gao, Jianhao Hong, Yizhou Chen

School of Modern (School of Automation), Beijing University of Posts and Telecommunications, Beijing, China

Wei Fang(✉): fangwei@bupt.edu.cn

Tel: +86 010-62281368

Abstract: Wearable augmented reality (AR) can superimpose virtual models or annotation on real scenes, and which can be utilized in assembly tasks and resulted in high-efficiency and error-avoided manual operations. Nevertheless, most of existing AR-aided assembly operations are based on the predefined visual instruction step-by-step, lacking scene-aware generation for the assembly assistance. To facilitate a friendly AR-aided assembly process, this paper proposed an Edge Computing driven Scene-aware Intelligent AR Assembly (EC-SIARA) system, and smart and worker-centered assistance is available to provide intuitive visual guidance with less cognitive load. In beginning, the connection between the wearable AR glasses and edge computing system is established, which can alleviate the computation burden for the resource-constraint wearable AR glasses, resulting in a high-efficiency deep learning module for scene awareness during the manual assembly process. And then, based on context understanding of the current assembly status, the corresponding augmented instructions can be triggered accordingly, avoiding the operator's cognitive load to strictly follow the predefined procedure. Finally, quantitative and qualitative experiments are carried out to evaluate the EC-SIARA system, and experimental results show that the proposed method can realize a worker-center AR assembly process, which can improve the assembly efficiency and reduce the occurrence of assembly errors effectively.

Keywords: Intelligent augmented reality, scene-aware, edge computing, manual assembly, smart factory

Acknowledgments: The authors gratefully acknowledge the supports of the Beijing Natural Science Foundation (3204050), the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems (VRLAB2020B05), the National Key Research and Development Program of China (No. 2019YFC0119200).

Ethical approval: We confirm that this manuscript is original and has not been published, nor is it currently under consideration for publication elsewhere.

Conflict of interest: The authors declare no competing interests.

1 Introduction

In the Industry 4.0 era, products with high quality and mass customization have been growing at a surprising pace, and researches have shown that the assembly process costs about 50 percent of the production developing time while nearly takes up about 20 percent of the total manufacturing time currently[1]. Although assembly automation technology has made great developments, most discrete assembly operations for complex products are still performed by manual operations, which imposes a great challenge to the operator, especially for the newer due to lack of experience[2]. Therefore, the ability to sense, monitor, characterize, and support workers for highly complex assembly with intuitive guidance have become imperative.

Along with emerging information and communication technologies, augmented reality (AR) can superimpose virtual models or text on real scenes, enabling a novel information delivery and leading to cognitive load reduction [3][4][5]. Recently, AR has been widely used in different industrial scenes, such as design[6], assembly[7], and warehouse[8], and intuitive and visual work guidance can result in more efficient manual actions.

Although visual AR instructions for assembly has been demonstrated significant for their promising potential, it has been found that the increase in performance through AR depends on the complexity and nature of the task[9], and most of the current AR assembly instructions produce are designed and scheduled in advance, where the step-by-step augmented guidance is triggered by human intervention or default procedure[10]. When faced with the varying assembly items and personalized assembly tasks, this rigid instruction procedure is difficult to adapt to changing assembly scenarios. Thus, a friendly and scene-aware AR assembly system is still expected for increasing assembly productivity while minimizing assembly time and errors.

With the advancement in artificial intelligence (AI), especially in deep learning, machines are now able to perform as well as, or even better than humans in various tasks, such as image recognition, and it is be concluded that the use of AI technology in combination with AR could lead to more efficient, interactive and intelligent manufacturing applications[11]. To improve manual assembly performance, instead of the commercialized AR device, a video see-through AR system is established with a fixed screen and powerful PC [12], and then the AR assembly system with the support of a deep learning network for tool detection is presented, which can provide on-site instructions for operators. While Park[13] combines deep learning-based object detection and instance segmentation to provide more effective visual guidance on Hololens, enabling the user to more easily and quickly perform complex assembly tasks. Nevertheless, these deep learning based AR assembly methods mainly focus on the tool detection or segmentation around the objects to be assembled, lacking context understanding about the assembly process, and the launch of step-by-step AR guidance still depends on the human intervention or the predefined procedure, lacking natural interactive mechanisms between the operators and the assembly components.

In addition, these AI-based methods put forward a higher computing resource requirement, although the state-of-the-art AR glasses, such as Hololens, can provide recognition performance with the AR assembly task, it is still too heavy for long-term wear and too expensive to wide deploy on the shop floor. While the lightweight AR glasses[14], such as BT-300, are friendlier for long-term assembly instruction in the actual industrial scene, but it is difficult to provide high computational complexity scene recognition and segmentation when performing scene-aware AR assembly tasks.

Thus, this paper proposed an Edge Computing driven Scene-aware Intelligent AR Assembly (EC-SIARA) system

to realize the smart and worker-centered assembly assistance method. With the help of edge computing (EC), the semantic understanding of the current assembly status is perceived by the high-efficiency scene recognition algorithm. Then, the corresponding augmented guidance for manual operation can be activated accordingly, alleviating the cognitive burden for the operator while generating AR instruction procedures, and thus a more flexible and intuitive AR assembly experience is available. The main contributions of this paper are summarized as follows:

1 An intelligent AR assembly system including an edge computing module is proposed, enhancing the computing capacity of the current lightweight wearable AR device for assembly scene awareness.

2 The assembly instruction can be superimposed accordingly based on the scene understanding about the manual assembly status, alleviating the mental burden to trigger the next AR instruction while performing AR assembly.

3 The EC-SIARA system can reduce the assembly time and errors significantly in a manual assembly task compared with the conventional paper-based manual instruction.

The remainder of the paper is organized as follows. Related works are detailed in Section 2. Our proposed is presented in Section 3, in terms of system description about the EC-SIARA system, wearable AR-based edge computing, and deep learning based assembly status awareness. The experimental setup and evaluation metrics are described in Section 4. Finally, Section 5 provides the conclusions of this study.

2 Related works

Currently, lots of studies have been conducted to provide user-centered task assistance using wearable AR devices, while deep learning-based methods are widely used in image-based recognition applications. Nevertheless, due to the complexity and uncertainty of the industrial assembly environments, there is still a lack of research on the utilization of real-time semantic understanding by lightweight wearable AR to effectively support manual assembly. In this section, we discuss previous research on AR-aided manual assembly and deep learning-based object detection for wearable AR.

2.1 AR-aided manual assembly

Over the past decade, due to the seamless integration capacity between the virtual objects and actual scene, more and more research regarding AR has become popular in manufacturing scenarios [15], which can be applied for practical task assistance with intuitive visual instructions. Wang et al [7] conducted a comprehensive review for AR assembly, it reveals the features and advantages of AR and quantified the industrial assembly in terms of key performance indicators. They divided literature related to AR assembly into three categories: AR assembly guidance; AR assembly training; and AR assembly design, and the AR-aided visualization enables the worker to request process-relevant information of an assembly line during the production process. The significant reductions of errors and time in AR assembly tasks have been proven by Boeing, which stated “This has tremendous potential to minimize errors, cut down on costs and improve product quality” [16].

Uva et al. [17] presented a spatial AR (SAR) system for working stations, and the visual instructions were projected directly on the objects to be maintained, and their research shows that SAR instructions have significant benefits compared to paper-based instructions in tasks. Accordingly, Radkowski et al. [10] sought to investigate the relationship between visual features and the complexity of an AR assembly task. The results indicate that the difficulty of a task does not significantly affect one's assembly performance. Nevertheless, assembly is highly correlated with

converting a given 3D model into visual instructions, which give information about operations, tools, assembly procedure that is also defined in simple terms as process planning, thus a friendly worker-center AR assembly related to the actual assembly process is more appropriate to reduce operator's cognitive load. Zhu et al. [18] presented a context-aware AR system to assist the operators in routine and ad hoc manual tasks, it can analyze the contexts of the tasks and provide relevant information to the operators by rendering the information on the real equipment and environment.

In general, lots of experiments indicate that AR-assisted assembly is more efficient in terms of completion time and error rates, and it has also been found that the increase in performance through AR depends on the complexity of the assembly task[9]. To make the AR-aided assembly more friendly, Mourtzis et al [19] proposed an automated approach for remotely supporting assembly workstations, and this makes the application adaptive to production re-scheduling, as the operator is given the assembly instructions based on the task currently assigned to the production schedule, without previous preparation. The high level of flexibility in the task may support the production of highly customized products. Nevertheless, the semantic-based approach is still unavailable when building AR instructions related to the actual assembly process.

Deshpande et al. [20] believe that a user-oriented visual cognitive mechanism has an important impact on the artificial assembly process. In other words, the cognitive mechanism of visual cues can guide and improve the collaborative performance of AR assembly. A novel human cognition-based interactive AR assembly guidance system is developed [21], which can provide various modalities of guidance to assembly operators for different phases of the user cognition process during assembly tasks, facilitating the interaction between the user and the rendered content. While Wang et al [22] describe an AR collaborative assembly platform: SHARIDEAS, and it applied a generalized grey correlation method to integrate user cues and scene cues, leading to appropriate and intuitive assembly guidance for local workers.

In the end, although great improvements about AR assembly are achieved during the past decade, existed methods mainly focus on the AR instruction and procedure related to geometric consistency, locking of high-efficiency scene understanding while doing the manual assembly process, and thus a friendlier worker-centered visualization and interaction for practical and intelligent assembly assistance are worth of further study.

2.2 Deep learning based AR assembly

Given the achievement of AI, especially in deep learning, machines are now able to perform as good as, or even better than humans in various tasks, such as image recognition [23], image segmentation [24], et al. And AI has been utilized in all aspects of the manufacturing process and supply chain from design, operations management, maintenance and assembly[25], which also promote the deep learning-based object detection to be applied in AR applications.

Region-based CNN (R-CNN) is one of the most representative topics in the field of object detection, and it can identify a rectangular region as an object in an image based on the features of each rectangular region, and then fully connected layers are applied to the identified features and classify the objects [26]. However, because the CNN is applied to a considerable number of rectangular areas, the computation costs are high. Therefore, the Fast R-CNN has been proposed [27], which reduces the time complexity by applying the CNN to the entire image, identifying a rectangular area, and applying fully connected layers.

Abdi et al. presented a framework for traffic sign recognition based on a deep learning method and AR, and in which augmented virtual objects are superimposed onto a real scene [28]. Based on the utilization of built-in sensor data such as GPS, IMU, and magnetometer data, Rao et al. presented a mobile outdoor AR method by adopting a vision-based deep learning object detection approach [29]. Given the error analysis of YOLO compared to Fast R-CNN, experimental results show that YOLO makes a significant number of localization errors, and which has relatively low recall compared to region proposal-based methods[30].

Although some studies utilized deep learning-based object detection methods for AR applications, they provided limited visual augmentation such as 2D-based static images. Park et al. present a new wearable AR method for task assistance, which is based on the combination of deep learning-based object detection and instance segmentation with AR-based spatial relation[13]. Ferraguti et al. [31] proposed a wearable AR approach for the online quality assessment of polished surfaces. As originally conceived, AR used a sensor-rich head-mounted display as a compute engine. This provided a deeply immersive user experience, but with the inconvenience of tethering. In these applications, the operators focus on performing a real-life task, while some mobile devices act like a virtual instructor or personal assistant to give her prompt guidance.

Nevertheless, to accurately understand an operator's state, the applications have to leverage the enormous data captured by relevant sensors on mobile devices and apply a huge amount of computation for interpretation. This computation is usually too heavy to be run entirely on a lightweight AR device, which is constrained by its weight, size, and thermal dissipation[32]. What is more, although current AR hardware is ready for deployment in some areas, it might lack the maturity to be deployed for more demanding tasks until now[33].

In summary, most past work has only used AR as a tool to supplement assembly activities, the determination of what data to display and the visual guidance procedure is mainly dependent by the operator on the shop floor, resulting in varying performance among workers with different experience level. These limitations would be eliminated by using AI as a tool in the computational framework. Therefore, given the lightweight wearable AR device, an edge computing driven SIARA system is needed for worker-center AR assembly tasks.

3 Methodology

3.1 System description of the EC-SIARA

As shown in Figure 1, the flow chart of the proposed EC-SIARA system is depicted, and the system mainly consists of two parts: the edge computing part and the mobile AR assembly part. First of all, the connection between the Edge Cloud (EC) and AR devices is established in advance. Thus, the real-time image captured by the wearable AR device can be delivered to the Cognitive Virtual Machine (Cognitive VM) deployed on the EC, alleviating the computational burden for the lightweight AR devices. That is to say, through integrating the Cognitive VM into the EC, the high-efficiency scene awareness about the manual assembly actions is achieved. Accordingly, the next visual guidance can be triggered and rendered automatically based on the current assembly action. As a result, the human intervention to activate the step-by-step AR assembly process is avoided, and the scene-aware visual guidance can be superimposed on the real industrial scene automatically, leading to a more intelligent AR assembly operation.

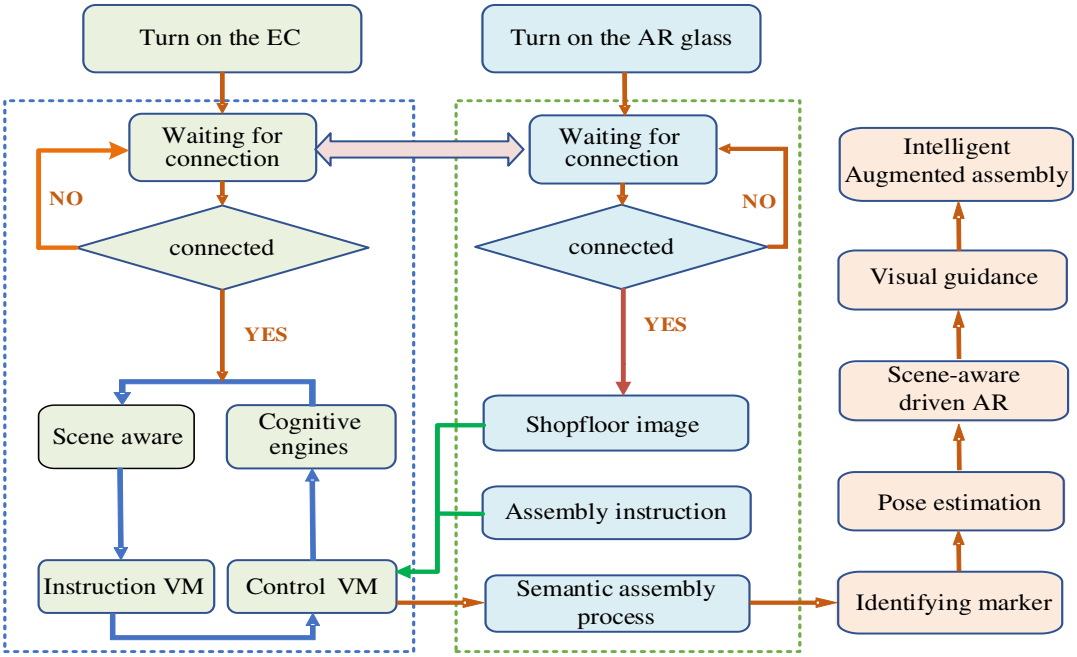


Figure 1 The flowchart of the proposed EC-SIARA system

3.2 Edge computing for mobile AR device

With the development of mobile computing power, visualized route directions can be rendered on the glasses and the users can directly follow the routes without paper reading. However, when encountering semantic understanding of the real assembly scene, it's difficult for current lightweight AR glasses to meet these computing performance requirements. In this paper, an edge computing module is applied to offload the data to the EC for AR assembly scene awareness. It can not only solve the problem of insufficient computing power, but also accelerates the speed of the whole EC-SIARA system. As shown in Figure 2, the edge computing module within the EC-SIARA system consists of four parts:

1. *Data collection and transmission*. Real-time images are collected from AR glasses and encapsulated according to the transmission protocol. The input data can be delivered to the EC through only one or two transmission hops. The Central Cloud (CC) can also be included in the edge computing system to deal with certain tasks, for example, the EC can be deployed to the industrial fields while the CC can be used to manage the ECs and sensors.

2. *Multiple VMs structure*. The Control Virtual Machine (Control VM) is responsible for the interaction between server and clients, at the same time, it controls the VMs' interaction within the edge computing system. The servers, usually the devices with high computing power, are set up as the host machine, working as the EC of the system. The Direction Virtual Machine (Direction VM) and Cognitive VMs process the data and give instruction information. The multiple VMs are connected by the Control VM, collecting the output data streams from each VM and delivering the data streams to the individual VM according to the datatypes and the VMs' requests.

3. *Content-based task distribution*. To deal with different contents, multiple Cognitive VMs can be set up accordingly. When input data is sent to the system, the visual information and other contents are separated by the Control VM, then they are packaged as certain datatype and distributed to specific Cognitive VMs to process. Besides, both the package of data and the data distribution are controlled by the Control VM. All in all, the Control VM is acting as the brain of the whole system.

4. *The Instruction VM and the output*. An Instruction VM is designed to store and regulate the whole manual assembly process and the visual guidance. After the awareness of the current assembly process by Cognitive VMs, the Control VM requests the Instruction VM to give the next augmented guidance and help the Instruction VM to pass the instruction data in JavaScript form to the AR glasses through the WebSocket.

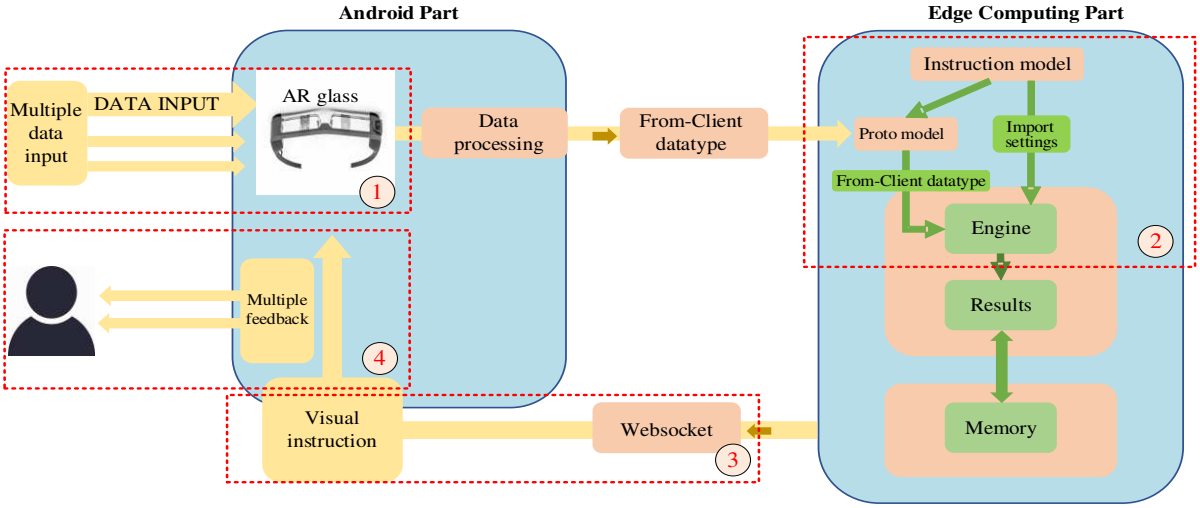


Figure 2 Schematic diagram of the edge computing system for AR assembly

To deploy different VMs in the edge computing module, there is a mechanism to find different virtual machines. Control VMs are started first, followed by the various Cognitive VMs, and they are connected through a private virtual network. The UPnP server provides a standardized, broadcast mechanism connection service to enable the Control VM to control different Cognitive VMs, and then AR glasses can receive the information flow of the augmented guidance through the TCP/IP protocol.

3.3 Deep learning for AR assembly

Currently, most AR assemblies are based on predefined procedures, and human intervention is essential to trigger the step-by-step assembly operations visual instruction, which is cumbersome for the worker when performing manual assembly on the shop floor. In this paper, based on the scene awareness of the actual assembly status, the EC-SIARA system can activate the AR guidance automatically related to the current assembly status, improving the efficiency of AR assembly while alleviating the workers' cognitive burden.

In this paper, Yolov3 is chosen as the basic framework for scene recognition of the manual assembly process, which is based on the R-CNN algorithm and has some improvements over it. It discarded the sliding window approach adopted by the R-CNN, instead, lots of potential bounding boxes by using the neural network are generated. As a result, fewer potential bounding boxes are generated and thus improve the recognition speed comparing to the R-CNN. During the training and testing process, the high efficiency of the Yolov3 can guarantee the performance of semantic recognition of the assembly status during the assembly process. In the proposed network structure (as shown in Figure 3), each residual module consists of two convolution layers and a shortcut connection. The residual module is added to provide conditions for deeper network layers, where Darknet-53 is the backbone network of the cognitive module, which consists of 53 convolution layers. For each convolution layer, the original feature map size will be reduced to 1/2 of the original input size, and the convolution layer is followed by a BN layer and a leaky Relu layer.

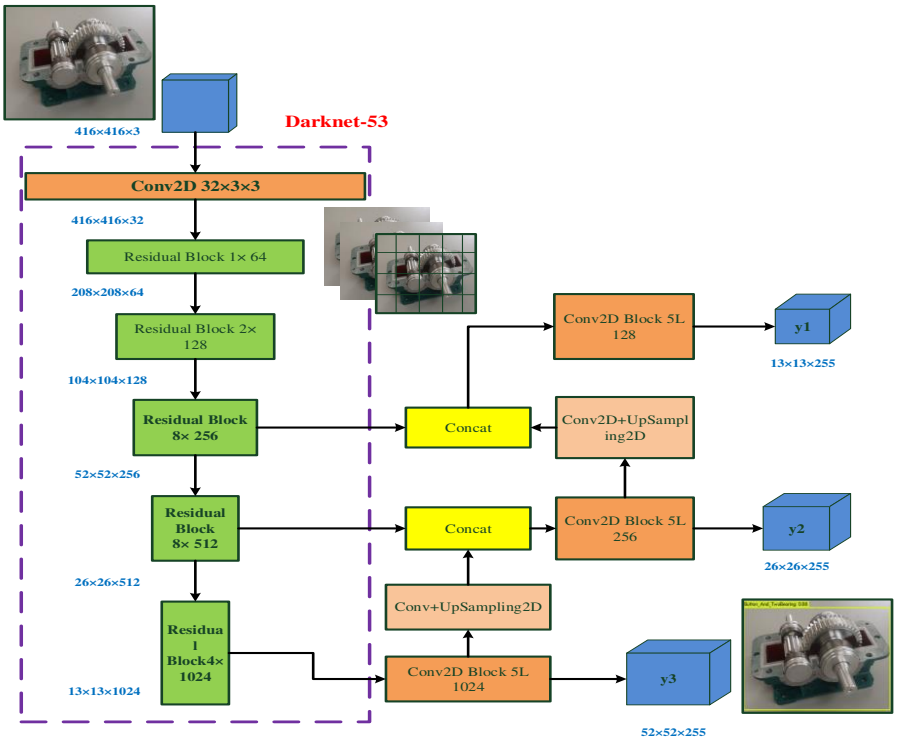


Figure 3 The network structure of the proposed actual scene recognition

The loss function of the proposed cognitive network is presented as follows:

$$
\begin{aligned}
&Loss = \lambda_{coord} A + \lambda_{coord} B - C - \lambda_{coord} D - E \\
&A = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [(x_i - \hat{x}_i^j)^2 + (y_i - \hat{y}_i^j)^2] \\
&B = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i^j})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i^j})^2] \\
&C = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \\
&D = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \\
&E = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} [\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - P_i^j)]
\end{aligned}
\tag{1}
$$

where I_{ij}^{obj} represents whether the j anchor box of the i^{th} grid is responsible for identifying the target, which means that the j anchor box of the i^{th} grid. The anchor box is responsible for identifying the target. $\hat{C}_i^j$ is the parameter confidence level. When there is one anchor box in all anchor boxes and the IOU of the ground truth box of the identified object is the largest, only its $\hat{C}_i^j$ =1, and the other anchor boxes are set to 0.

The scene-aware method can predict the boundary frame of the target object, and divide the image into small blocks that are not overlapped. Each block is responsible for the prediction of the boundary box (as shown in Figure 4). The small cell prediction boundary box needs 4 values to represent the prediction position, namely, the center coordinate point (x, y) and the length and width of the prediction box.

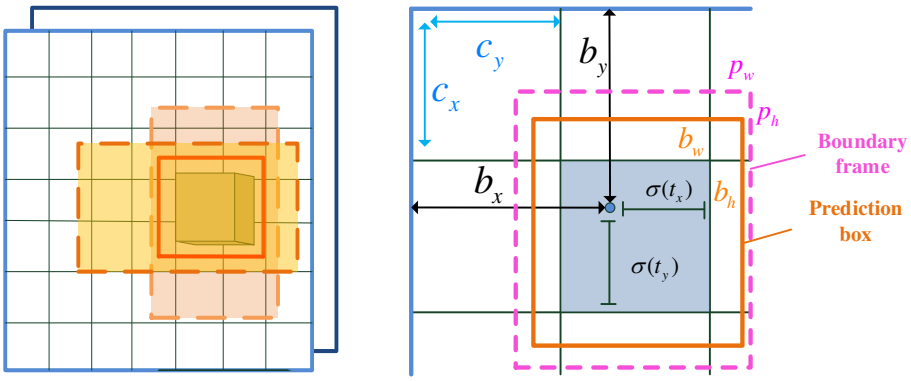


Figure 4 The prediction diagram for the assembly items recognition

The centroid is the border selected as the center when clustering, where IOU (International over Union) is usually applied to demonstrating the performance of the object recognition, and the larger the IOU value, the smaller the distance between the two boxes and the closer the cluster analysis box size is to the label box. Thus, the quantitative criteria $d(\bullet)$ is derived by:

$$d(B,C)=1-IOU(B,C) \tag{2}$$

where B represents the frame obtained by cluster analysis (box), C represents the border selected as the center during clustering (Centroid).

As shown in Figure 4, (b_x,b_y) is the center coordinate positions, b_w and b_h are the widths and heights of the prediction box, respectively, c_y is the distance from the upper left corner of the current grid to the upper left corner of the image after normalization, p_w and p_h are the width and height of a priori box, respectively. σ is the sigma function, and t_x,t_y,t_w,t_h are the parameters to be trained to calibrate the size of the prediction box to the actual box.

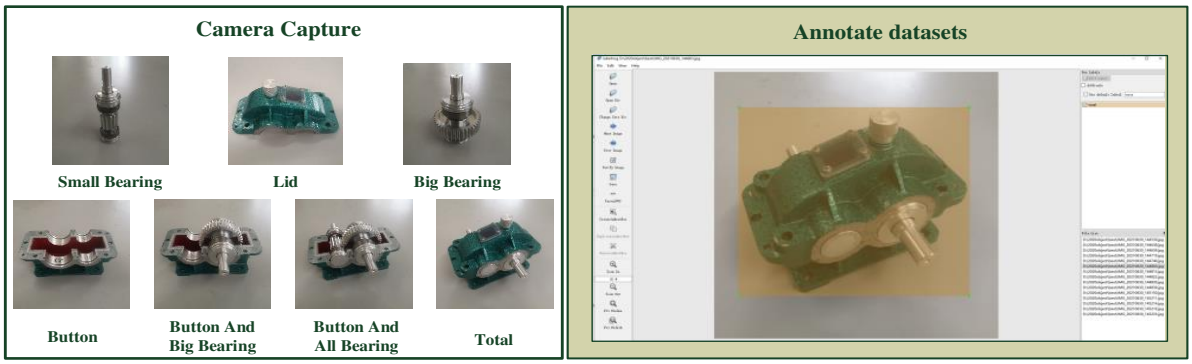
$$\begin{aligned} b_x=\sigma(t_x)+c_x,\ \ b_y=\sigma(t_y)+c_y \\ b_w=p_w e^{t_w},\ \ b_h=p_h e^{t_h} \end{aligned} \tag{3}$$

Based on the lightweight network structure, high-efficiency recognition on the actual assembly status is carried out with the help of edge computing, resulting in a scene-aware intelligent AR assembly.

3.4 Scene awareness for the assembly process

To integrate the semantic recognition model into the EC-SIARA system, certain datasets involved in the assembly process are applied to train and verify the network model. In this paper, based on the manual assembly operations, the training dataset (839 pictures) is divided into seven categories: lid, small bearing, big bearing, button, button, and big bearing, button and tow bearings, and total (shown in Figure 5 (a)). During the training process, the learning parameters and coefficients are constantly modified to get higher accuracy and recall rates.

The dataset is divided into the training set and the testing set, conventionally, the ratio of the volumes of the training set and the testing set is 8:2. The training set is used to train the recognition model, while the test set is used to evaluate the accuracy and recall rates of the model.



(a) (b)

Figure 5 Preparation of the training dataset for the assembly process

As a supervised learning method, the pictures of the assembly process need to be annotated manually in the following labeling process. For a specific picture, different people mark the picture separately and give the picture a subjective grade according to the picture's quality (shown in Figure 5 (b)). Then the picture with the highest grade is selected as the best one, which is included in the dataset. In this way, the model trained can be improved and eventually get relatively high confidence. To improve the accuracy of the model and make the loss function (loss value) drop rapidly during the training process, the Yolov3 automatically clusters the test sample set in advance, then a set of anchor clustering numbers are obtained as the training reference standard.

The training sample set is pretreated to make the learning rate equals 0.1, so that the model iterates until the loss value have no significant change, and then the training is suspended. The learning rate values are selected at equal intervals, and the iterations are suspended after the same number of iterations. The minimum loss value of each group of learning rates is recorded visually through the training log generated by training. The least-square method is used to fit the recorded data. Through error estimation calculation, the data points are selected to fit the quadratic curve. According to the fitting curve, when the learning rate = 0.074, the minimum loss value can be reached, it can be found that the confidence about the assembly status recognition is reliable (shown in Figure 6).

Average Precision on detecting differ states

State	Average Precision
Total	93.3%
Lid	89.1%
Small bearing	92.4%
Big bearing	85.6%
Base	84.1%
Base and big bearing	89.3%
Base and tow bearings	90.7%

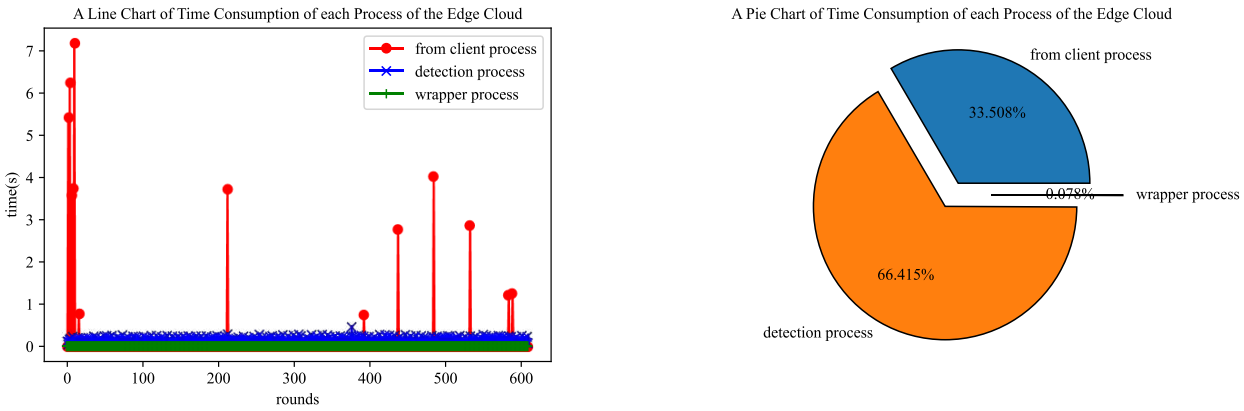
Figure 6 The confidences of the assembly process recognition

4 Experiments

4.1 Evaluation on assembly process cognition

To evaluate the Cognitive VMs' performance under the edge computing system, experiments on connection performance between the PC station and the wearable AR device are carried out. There are generally two kinds of data that need to be collected and evaluated: 1. The time consumption of the launch of the edge computing system: how long would it take for EC to load the Yolov3 model and to be prepared as the server; (2) The time consumption of each process of the EC during the detection and instruction: how long would it take for EC to load the data from the wearable AR device, to recognize the certain image, and to wrapper the result as a certain datatype.

In this paper, we use the PC (Core i7-10870, 2.20GHz, 8Cores) as the EC, and a wearable AR device (Android 10.0) as the client to collect information and display augmented instructions. Through evaluating each process of the EC during detection and instruction, we find that the detection process is most time-consuming, up to about 66.42% of each round's time consumption, and due to the surge of certain round's from-client process's time-consumption, the communication between client and server is most unsteady (shown in Figure 7(a)). By comparison, the wrapper process is swift and steady, and each round's time consumption is almost the same and the process only takes up 33.508% of total time consumption (shown in Figure 7(b)). That's to say, the Edge computing system generally provides the user a rather swift and steady connection service, and the user may apply it to the industry as long as it can meet the basic need. During the experiment, some factors such as the network status and the distance between the client and the server may affect the communication of EC which leads to unstable communication time. To evaluate the ability proposed Cognitive VM for AR assembly process, the consumption of time while detecting different items is recorded and evaluated.



(a) (b)

Figure 7 The time consumption analysis about edge computing is driven assembly recognition

The results (depicted in Figure 8) also demonstrated that the time consumption for recognition of different items is stable and swift, normally between 0.1 seconds to 0.16 seconds while encountering different assembly components. It shows that the Yolov3 within the EC module can deal with the different items' recognition, and which are the basic features of the EC-SIARA system. The results illustrate the availability and collaborative processing capability of the edge computing system, leading to a friendlier EC-SIARA system.

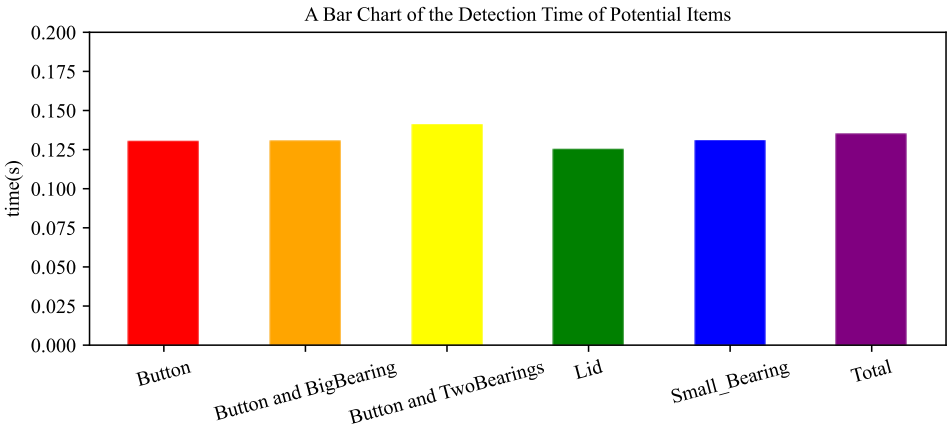


Figure 8 The time escaping against varying assembly process

Thus, the step-by-step AR assembly process can be activated according to the scene awareness by the edge computing module. As shown in Figure 9, the process of reducer assembly is shown in the form of the screenshots captured by the lightweight AR device BT-300. Based on the accurate localization derived from the fiducial marker, the visual guidance is superimposed on the real assembly scene for the operator. At the same time, the deep learning model is applied to recognize the current assembly process, and through EC's Instruction VM to indicate the next assembly operation. Then, the operator can follow the instructions, such as the text cues in the red box to accomplish the manual assembly work efficiently. The result of the recognition is processed by the EC's Instruction VM and the corresponding 3D virtual model is conveyed by the edge system to the AR glasses to display the next assembly process. For example, the EC-SIARA system analyzes the captured scene image and recognizes that the operator has installed the follower wheel, then it automatically prompts that the next assembly step is to install the driving wheel.

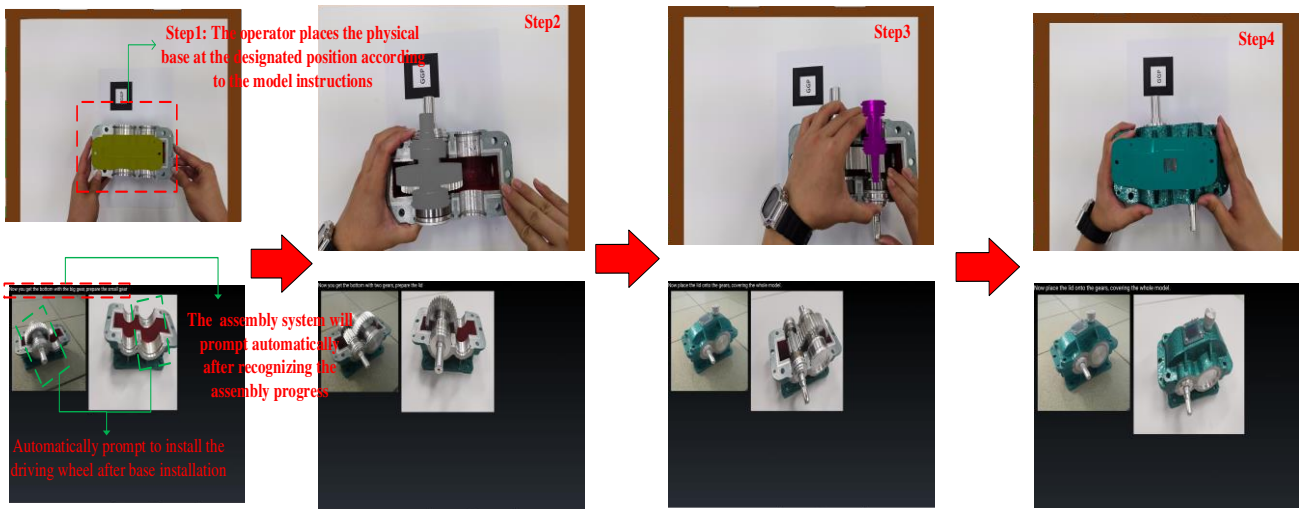


Figure 9 Edge computing driven assembly process cognition

4.2 Evaluation on the EC-SIARA System

In this section, comparative experiments are carried out to evaluate the EC-SIARA system. For fair comparisons, the participators with/without the proposed system execute the same manual assembly operation, and two novice group are assigned the same assembly task, avoiding the influential fact that they are getting familiar with the

assembly process during the experiment. In addition to evaluate the influence on wearing AR glasses for experienced workers, workers with/without the EC- SIARA system are also evaluated, and the flow chart of the experiment is shown in Figure 10.

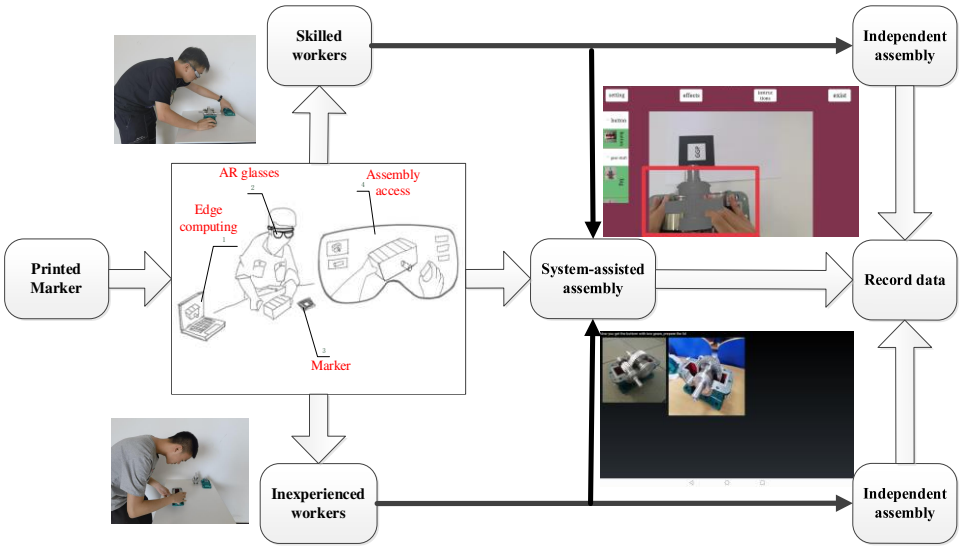


Figure 10 Experiment process for intelligent AR assembly

To evaluate the assistant effect of the EC-SIARA on the assembly process, four different groups are defined to accomplish the manual assembly task. In detailed: Group 1: Inexperienced worker A; Group 2: Inexperienced worker B with EC-SIARA system; Group 3: skilled worker; Group 4: skilled worker with glasses but don't apply the EC-SIARA system, and the pictures about different assembly scenes are depicted in Figure 11.

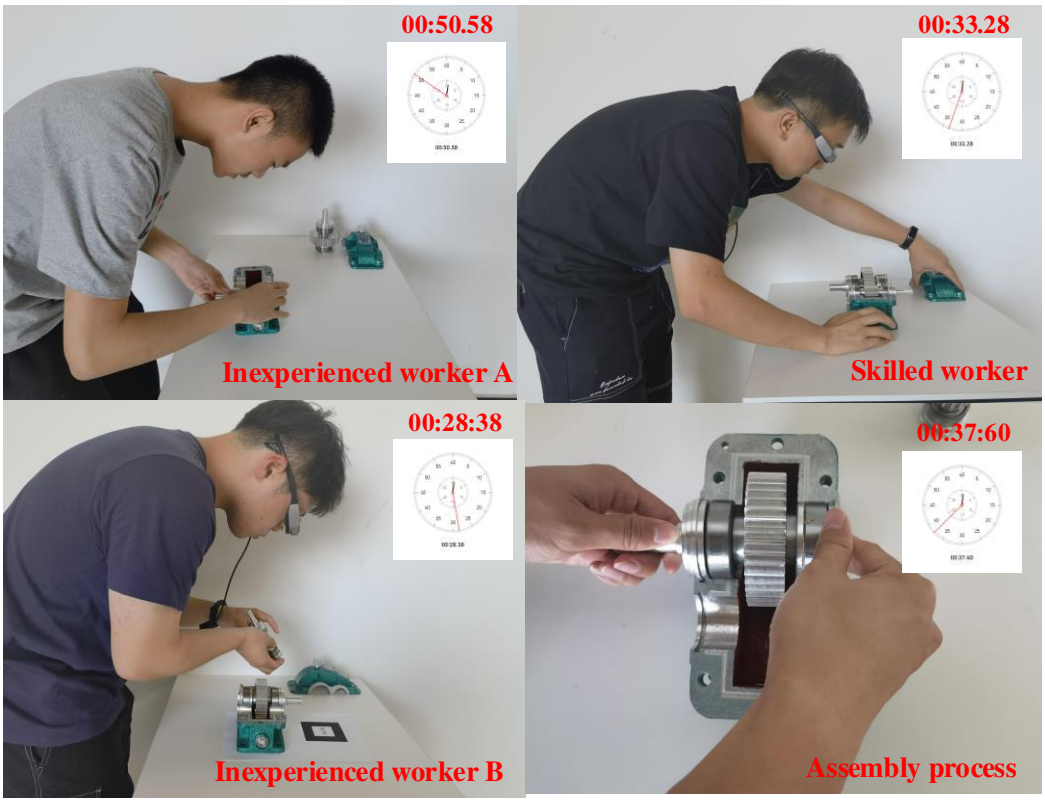


Figure 11 Comparative assembly experimental process for reducer

To obtain reliable experimental results, four repeated experiments are carried out among different groups, and the detailed results are collected and demonstrated in Table 1 and intuitive comparisons are depicted in Figure 12. We can find that the novice operator with the help of the proposed EC-SIARA system can finish the assembly task within about 33s, which is about 50% improvement in the efficiency without the EC-SIARA system. What is more, the assembly errors are avoided with the help of the scene-aware AR assistant, the result also illustrated that the proposed method can result in a more intelligent worker-center situation for manual assembly.

Besides, it is also interesting to find that the skilled workers spend the least time when installing independently, and it only needs about 30s to accomplish the assembly task on average. Experimental results also show that operator wearing AR glasses slightly lowers the assembly speed, which mainly due to the hindrance of the AR glasses when performing manual assembly operations. Compared with skilled assemblers, inexperienced workers who use the EC-SIARA system still take a longer time to assemble the reducer, the reason may be that inexperienced workers are not familiar with assembly parts, even if they have the EC-SIARA's assistance, they have to look for some parts which take a long time. The results also demonstrate that the performance of the EC-SIARA system on the shop floor is also related to the assembly task, which is more meaningful when encountering complex assembly tasks and has a more significant effect for beginners.

Table 1 Quantitative statistic on comparative assembly operations

	Experiment 1		Experiment 2		Experiment 3		Experiment 4		Average	
	Time/s	Accuracy	Time/s	Accuracy	Time/s	Accuracy	Time/s	Accuracy	Time/s	Accuracy
Group 1	50.58	75%	52.30	75%	50.11	100%	49.88	100%	50.72	87.5%
Group 2	33.28	100%	35.45	100%	32.18	100%	33.09	100%	33.50	100.0%
Group 3	28.38	100%	28.91	100%	29.66	100%	30.20	100%	29.29	100.0%
Group 4	37.60	100%	43.66	100%	39.67	100%	38.85	100%	39.95	100.0%

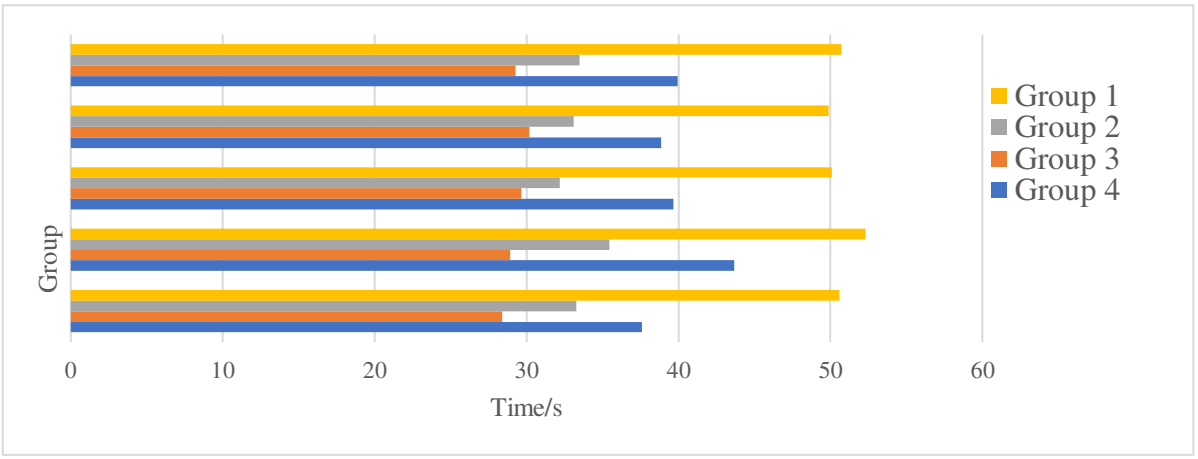


Figure 12 Comparative performance of AR assembly system among workers

5 Conclusions

To improve the intelligence level for manual assembly, this paper introduces an EC-SIARA method, which can understand the current assembly status and provide intuitive visual guidance for manual operators. Based on the EC system, a high-efficient semantic recognition for the assembly scene is available for lightweight wearable AR device,

and the scene-aware the step-by-step assembly sequence is realized automatically, alleviating the mental burden of operators to trigger the next AR instruction consciously. This EC driven intelligent AR system will be able to work autonomously in the assembly environment with minimal human intervention, resulting in a friendly worker-center AR operation. Experimental results also show that the proposed EC-SIARA method can not only improve the assembly efficiency but also improve the assembly accuracy for practical manual assembly tasks.

In the future, based on the remote edge computing capacity, we will give further research on the pose estimation from the assembly scene with regression method, instead of arranging markers within the actual scene in advance, and a friendlier worker-centered intelligent AR assembly is expected.

Reference

[1] Michalos G, Makris S, Papakostas N, Mourtzis D, Chryssolouris G (2010) Automotive assembly technologies review: Challenges and outlook for a flexible and adaptive approach, CIRP J Manuf Sci Tec 2(2):81-91.

[2] Servan J, Mas F, Menendez J, Rios, J (2012) Using augmented reality in AIRBUS A400M shop floor assembly work instructions. The 4th Manufacturing Engineering Society International Conference 1431(1):633-640.

[3] Simoes B, Amicis RD, Barandiaran I, Posada J (2019) Cross reality to enhance worker cognition in industrial assembly operations. Int J Adv Manuf Technol 105:3965-3978.

[4] Alves JB, Marques B, Dias P, Santos BS (2021) Using augmented reality for industrial quality assurance: a shop floor user study. Int J Adv Manuf Technol 115:105-116.

[5] Masood T, Egger J (2019) Augmented reality in support of Industry 4.0-Implementation challenges and success factors. Robot Comp Integ Manuf 58:181-195.

[6] Wang ZB, Ong SK, Nee AYC (2013) Augmented reality aided interactive manual assembly design. Int J Adv Manuf Technol 69(5-8):1311-1321.

[7] Wang X, Ong SK, Nee AYC (2016) A comprehensive survey of augmented reality assembly research. Adv Manuf 1:1-22.

[8] Fang W, An Z (2020) A scalable wearable AR system for manual order picking based on warehouse floor-related navigation. Int J Adv Manuf Technol, 109(7):2023-2037.

[9] Uva AE, Gattullo M, Manghisi VM, Spagnulo D, Cascella GL, Fiorentino M (2018) Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. Int J Adv Manuf Technol 94:509-521.

[10] Radkowski R, Herrema J, Oliver J (2015) Augmented reality-based manual assembly support with visual features for different degrees of difficulty. Int J Hum Comput Int 31(5):337-349.

[11] Sahu CK, Young C, Rai R (2020) Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review. Int J Prod Res, DOI: 10.1080/00207543.2020.1859636.

[12] Lai ZH, Tao WJ, Leu MC, Yin Z (2020) Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. J Manuf Syst 55:69-81.

[13] Park K, Kim M, Choi S H, Lee J Y (2020) Deep learning-based smart task assistance in wearable augmented reality. Robot Comp Integ Manuf 63:101887.

[14] Miller J, Hoover M, Winer E (2020) Mitigation of the Microsoft HoloLens' hardware limitations for a controlled product assembly process. Int J Adv Manuf Technol 109:1741-1754.

[15] Danielsson O, Holm M, Syberfeldt A (2020) Augmented reality smart glasses in industrial assembly: Current status and future challenges. J Ind Inf Integr 20(1):100175.

[16] Makris S, Karagiannis P, Koukas S, Matthaiakis A (2016) Augmented reality system for operator support in human–robot collaborative assembly. CIRP Ann-Manuf Techn 65(1):61-64.

[17] Uva AE, Gattullo M, Manghisi VM, Spagnulo D, Cascella GL, Fiorentino M (2018) Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. Int J Adv Manuf Technol 94:509-521.

[18] Zhu J, Ong SK, Nee AYC (2014) A context-aware augmented reality system to assist the maintenance operators. International Journal on Interactive Design and Manufacturing, 8:293-304.

[19] Evans G, Miller J, Pena MI, MacAllister A, Winer E (2017) Evaluating the Microsoft HoloLens through an augmented reality assembly application. Proceedings on Degraded Environments: Sensing, Processing, and Display, 101970V, DOI: 10.1117/12.2262626.

[20] Mourtzis D, Zogopoulos V, Xanthi F (2019) Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling. Int J Adv Manuf Technol 105:3899-3910.

[21] Deshpande A, Kim I (2018) The effects of augmented reality on improving spatial problem solving for object assembly. Adv Eng Inform 38:760-775

[22] Wang X, Ong S.K, Nee A.Y.C (2016) Multi-modal augmented-reality assembly guidance based on bare-hand interface, Adv Eng Inform 30:406-421.

[23] Wang, Z, Wang Y, Bai X, Huo X, Zhou J (2021). SHARIDEAS: A smart collaborative assembly platform based on augmented reality supporting assembly intention recognition. Int J Adv Manuf Technol 115:475-486.

[24] Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. pp.234-41.

[25] Zhang X, Ming X, Liu Z, Yin D, Chen Z, Chang Y (2019) A reference framework and overall planning of industrial artificial intelligence (i-ai) for newapplication scenarios. Int J Adv Manuf Technol, 101(9-12):2367-2389.

[26] Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation, IEEE Conference on Computer Vision and Pattern Recognition, pp.580-587.

[27] Girshick R (2015) Fast R-CNN. IEEE International Conference on Computer Vision, pp.1440-1448.

[28] Abdi L, Meddeb A (2017) Deep learning traffic sign detection, recognition and augmentation. Proceedings of the Symposium on Applied Computing, pp.131-136.

[29] Rao J, Qiao Y, Ren F, Wang J, Du Q (2017) A mobile outdoor augmented reality method combining deep learning object detection and spatial relationships for geovisualization, Sensors, 17:1951.

[30] Redmon J, Farhadi A (2017) YOLO9000: Better, Faster, Stronger. IEEE Conference on Computer Vision and Pattern Recognition. pp.6517-6525.

[31] Ferraguti F, Pini F, Gale T, Messmer F, Storchi C, Leali F, Fantuzzi C (2019) Augmented reality based approach for on-line quality assessment of polished surfaces. Robot Comp Integ Manuf 59 158-167.

[32] Ha K, Chen Z, Hu W, Richter W, Pillaiy P, Satyanarayanan M (2014) Towards wearable cognitive assistance, Proceedings of the 12th annual international conference on Mobile systems, applications, and services, pp. 68-81.

[33] Preum S M, Shu S, Ting J, Lin V, Williams R, Stankovic J, Alemzadeh H (2018) Towards a cognitive assistant system for emergency response. ACM/IEEE 9th International Conference on Cyber-Physical Systems. pp. 347-348.