APEM journal

Advances in Production Engineering & Management
Volume 16 | Number 4 | December 2021 | pp 393–404
https://doi.org/10.14743/apem2021.4.408

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

A deep learning-based worker assistance system for error prevention: Case study in a real-world manual assembly

Riedel, A.[a,*], Gerlach, J.[a], Dietsch, M.[a], Herbst, S.[a], Engelmann, F.[a], Brehm, N.[a], Pfeifroth, T.[a]

[a]Department of Industrial Engineering, Ernst-Abbe University of Applied Sciences, Jena, Germany

A B S T R A C T

Modern assembly systems adapt to the requirements of customised and short-lived products. As assembly tasks become increasingly complex and change rapidly, the cognitive load on employees increases. This leads to the use of assistance systems for manual assembly to detect and avoid human errors and thus ensure consistent product quality. Most of these systems promise to improve the production environment but have hardly been studied quantitatively so far. Recent advances in deep learning-based computer vision have also not yet been fully exploited. This study aims to provide architectural, and implementational details of a state-of-the-art assembly assistance system based on an object detection model. The proposed architecture is intended to be representative of modern assistance systems. The error prevention potential is determined in a case study in which test subjects manually assemble a complex explosion-proof tubular lamp. The results show 51 % fewer assembly errors compared to a control group without assistance. Three of the four considered types of error classes have been reduced by at least 42 %. In particular, errors by omission are most likely to be prevented by the system. The reduction in the error rate is observed over the entire period of 30 consecutive product assemblies, comparing assisted and unassisted assembly. Furthermore, the recorded assembly data are found to be valuable regarding traceability and production improvement processes.

A R T I C L E   I N F O

Keywords:
Deep learning;
Machine learning;
Industry 4.0;
Smart manufacturing;
Manual assembly;
Assistance system;
Error prevention;
Object detection

**Corresponding author:*
alexander.riedel@eah-jena.de
(Riedel, A.)

Article history:
Received 14 October 2021
Revised 3 December 2021
Accepted 4 December 2021

1. Introduction

The continuing global trend towards highly customized products, small batch sizes and short product life cycles is leading to a variety of challenges that manufacturing companies have to face [1, 2]. To meet these challenges, manufacturing systems must be designed more flexible both in terms of the production process and the product itself, as well as the deployment of employees. In practice, a flexible product and production means rapidly changing processes to which workers must adapt. However, flexible deployment of employees and tasks reduces the available training time per employee.

For economic and technological reasons, complex products with small batch sizes are preferably produced in a manual assembly environment [3]. However, in this type of production, product quality and assembly time depend heavily on the characteristics of the workers, such as their qualifications, working conditions and experience. The constraints of flexibility can counteract these desired characteristics, as, for example, a short-term temporary worker is not able to acquire many years of experience. At the level of the production process, high product complexity and variance lead to higher cognitive demands on each worker [4]. The resulting combination of limited worker skills and complex tasks causes the need for error preventing assembly assistance systems. These systems, also referred to as Cyber-Physical-Systems (CPS) [5], use sensors to collect data of the assembly, process them and provide visual instructions to the worker [6]. Generally, assembly assistance systems have shown to both reduce the error rate [7] and increase the workers productivity [8]. Methods for analysing worker motions in motion time systems have been recently advanced [9] and might be suitable to give assistance during assembly processes.

Like the workplaces they are used at, these systems need to be flexible and adaptable to a variety of products, processes, and workers. A modular architecture with a central unit and one or more plug-and-play cyber-physical units would meet the requirement for flexibility and adaptability of such systems. The use of deep learning object detection is crucial to the proposed method, as it allows to detect visible assembly errors and the visually observable assembly status at a high accuracy in contrast to traditional computer vision methods [10]. Thus, the accurate detection of every assembly part's position and status at all times can be considered as a crucial task and information source for other assembly assistance modules to work.

In this paper we propose a modular assembly assistance system combining various modules, applied to a complex real-world product on a manual assembly workstation. The system is driven by a deep learning object detection model to detect the assembly status and potential assembly errors from RGB video. The information derived from the object detection unit are used to display the current work instruction, errors, warnings and guarantee full traceability of each work step. The information about the current work step provided by the object detection unit is also used by a pick-by-light module and an electric screwdriver. Additionally, any type of worker assistance module that works on information which can be visually retrieved from the assembly process is attachable to the system. The system architecture is displayed in Fig. 1.

To validate the assumed error prevention potential, we conducted experiments in which test subjects carried out the manual assembly of an explosion-proof tube lamp with and without assembly assistance. The assembly product was provided by the internationally operating company R. Stahl AG for this investigation. Especially in safety-critical areas such as explosion-proof product manufacturing, the risk of assembly errors must be reduced to near zero.

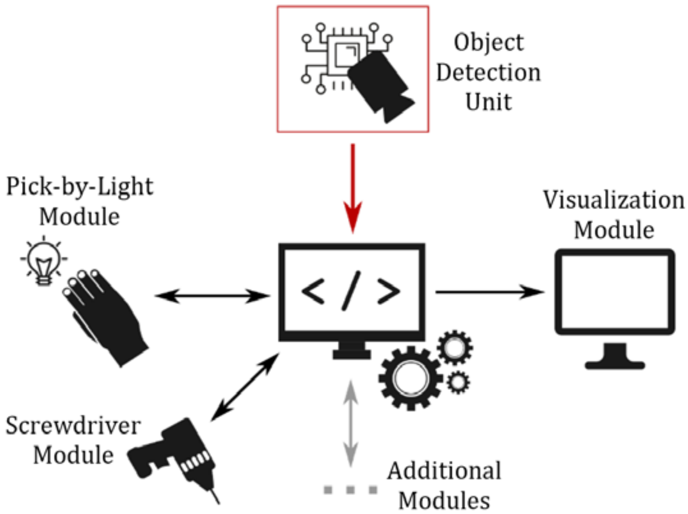

Fig. 1 Schematic architecture of the proposed error preventing assembly assistance system.

The use of object detection in manual assembly and modular assembly assistance systems has been described in literature [11-14]. However, the combination of both, their implementation effort and, most importantly, their quantitative effect on a real-world assembly task, remains mostly unclear. To address this lack of knowledge about modern assembly assistance systems, this work aims to provide:

- Architectural details of a modular assembly assistance system powered by state-of-the-art deep learning-based object detection technologies
- Implementational details of the proposed assistance system
- A quantitative analysis of the error prevention potential based on experiments conducted with test subjects

Related work

The proposed assembly assistance system characterized in Fig. 1 mainly relies on (1) deep learning object detection and (2) a modular approach. Both aspects will be considered and compared to existing methods in literature.

The application of deep learning-based object detection in manual assembly has been widely investigated in terms of feasibility and regarding how the acquired information can be used in the assembly process. The use cases include e.g., counting assembly parts to support the handling process [12], detecting small electrical parts during assembly [15] or recognizing work steps in a virtual reality environment [16]. These approaches underline the purpose and the importance of using advanced object detection in manual assembly, however they are not implementing the method in a productive assembly assistance system or evaluate its benefit. In a real-world scenario, the final goal of applying object detecting in manual assembly tasks would be to increase productivity by reducing the error rate or assembly time. This case study acts as a logical follow-up for many object detection assistance system concept studies by determining the productivity impact of this method.

Fraunhofer IPA propose a multi-modal worker assistance system "MonSiKo" combining 3D object detection, acoustic detection, and motion tracking, collecting data to operate a pick-by-light system and work step detection [17]. The sensor-fusion approach of vision, acoustics and inertial sensing is a reliable source of data but brings an expensive installation and high workplace invasiveness. A qualitative or quantitative analysis of the manual assembly process improvements was not carried out.

Oestreich *et al.* show the quantitative effect of a basic assembly assistance system on the assembly duration during the training period [18]. The proposed system includes digital instruction visualization and a proximity sensor for detecting workpieces. Each completed work step is manually confirmed by the user with the proximity sensor being the only source of indirect sanity checking the assembly process. Although a quantitative user study is conducted, the effect on assembly errors is not evaluated as they can hardly be directly assessed without the use of an object detection method.

Kaczmarek *et al.* present a vision-only approach to monitor the progress in a manual assembly task [19]. By using an RGB-depth camera, different height layers and areas can be defined as regions, where a located object is either assembled or not assembled. The method prevents assembly errors caused by unassembled parts, however the use of hard-coded areas and layers instead of machine learned features raises doubts about the universal and flexible use in a real-world use case.

A study conducted by Faccio *et al.* proposes a more advanced vision-only approach to error and progress monitoring [20]. Here, the workers hand positions are compared to virtual pre-defined three-dimensional control volumes to determine correct or wrong hand positions. The assistance system provides visual feedback based on the hand reaching into the control volumes. In a follow-up work, the authors conduct user studies to compare the assembly durations with and without the assistance system [21]. The proposed system shortens the average assembly duration by 22 %. However, the error reduction rate using the system is not provided. The use of pre-defined control volumes is inflexible in the case of various workers having different types of hand movements. Deep learning object detection methods could provide a sufficient way to directly observe the object status and potential errors.

The case study provided by Rocha *et al.* presents a vision-based assembly assistance system, based on detecting hand movements over control points [22]. The system follows a modular approach, using the collected movement data to determine the components status. Rocha *et al.*

focus on implementational details and data collection rather than on determining error or assembly duration reduction.

The literature review underlines the current research situation, where deep learning-based object detection is underrepresented in assembly assistance systems and their impact on key performance indicators like error rate has hardly been studied. Due to the large variety of assistance systems described in literature, our proposed method will act as a meta-approach of the most advanced and lightweight systems.

2. Materials and methods

To comprehensively perceive the assembly process and simultaneously provide aid to the worker, we chose a vision-only approach. Based on the assembly part's state, a pick-by-light module, an electric screwdriver and an according work instruction is triggered during the assembly process. The product to be assembled is a complex explosion-proof tube lamp which require 34 work steps including plugging, clicking, screwing, and testing operations. To carry out the quantitative error analysis, two groups of 10 participants each assembled the product with and without the assistance system.

2.1 System architecture

The systems individual modules are interconnected via the MQTT network protocol. The key information shared in the Broker/Client architecture is the currently observed workpiece status. This status might be the presence of the workpiece, the absence of a workpiece or a potential faulty assembly. Based on this information, the current work step is derived to present the associated instruction via the visualization module. When detecting a faulty assembly, the visualization module will also present warning messages to the worker. Both the pick-by-light module and the electric screwdriver are activated based on the current work step. For traceability purposes, the grab status and tightening status of the pick-by-light and screwdriver module are published as MQTT messages and stored in a database. Assembly part status messages are stored in a database as well. The complete MQTT messaging scheme is displayed in **Napaka! Vira sklicevanja ni bilo mogoče najti.**.

To derive the work step from assembly parts, each work step is mapped to the presence or absence of certain key objects for this respective step. This approach guarantees flexibility in case of changing workflows.

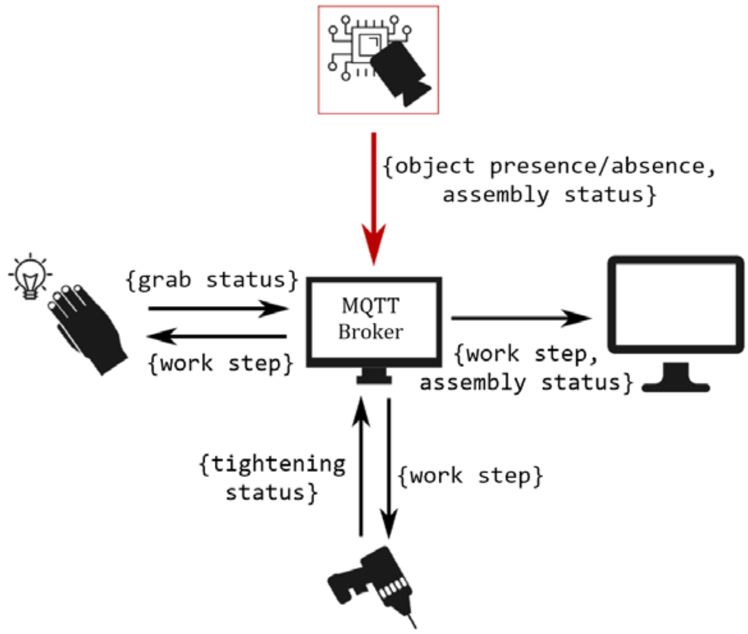

Fig. 2 MQTT messaging scheme of the proposed assembly assistance system. The figure pictures the information flow and content between the modules

The single modules run on edge devices such as Raspberry Pi (pick-by-light module, screw-driver module, MQTT broker), Nvidia Jetson AGX Xavier (object detection unit) or a capacitive industrial panel PC (visualization module). The use of individual devices for each module allows the easy extension or reduction of the proposed assistance system.

2.2 Deep learning-based object detection

The workpiece status (presence, absence, faulty) is acquired via the state-of-the-art single shot object detection model YOLOv4 [23]. The model was chosen because it is capable of delivering real-time results at a high accuracy on GPU edge devices such as the Nvidia Jetson AGX Xavier. To maximize the inference speed, the trained model was optimized using Nvidia's TensorRT runtime. The SDK applies a variety of optimization techniques such as precision calibration, layer fusion, or multi-stream execution [24].

In total, 1512 training images containing 5485 object instances and 790 validation images containing 1468 object instances were used for training the model. The training data consists of 41 individual object classes, so each class is represented by 134 training instances on average. The small training set size is reasonable due to the low intra-class and high inter-class variance of the assembly parts. The training was conducted on a Nvidia GeForce 1650 Super GPU using the Darknet Framework [25]. The mean average precision at 50 % intersection over union (mAP@.50) reached 99.81 % after 6 training epochs for the validation data set. This evaluation metrics gives proof that the model is capable of predicting all classes at a satisfactory level.

Object and error detection can be difficult with small parts or parts that change their appearance only slightly after assembly. The chosen camera perspective (aerial view, see Fig. 2) makes it impossible, for example, to detect the correct vertical engagement of a part in a holder. Identifying the correct torque value by vision can be considered very difficult, so bolts are only checked for their presence. It has been shown that parts that can appear in many different shapes, e.g. thin wires, require a larger number of training data to achieve an accurate detection result.

Collecting the training and validation data and annotating all object instances took a workload of roughly 10 hours for a single person. Given the high complexity of the product with 34 work steps and 41 object classes to annotate, this temporal expenditure seems reasonable, as it can also be performed by untrained personnel. The inference is performed on a Nvidia Jetson AGX Xavier, publishing the detected objects in real-time via MQTT messages. For collecting the training data as well as inferencing, an IMX 178 image sensor was used, mounted above the working area. An example of the inference for a single frame is displayed in Fig. 2.

Fig. 2 Inference of a single frame by the trained YOLOv4 model. The inference results are published via MQTT and the corresponding work step or assembly faults are derived based on the detected objects.

2.3 Pick-by-light module

As pick-by-light is one of the most widespread assembly assistance systems used in industry [26], the proposed meta-assistance-system was equipped with such system. The module operates based on the published current work step derived by the object detection unit. All 21 boxes of material needed in the assembly process are provided with LED lights to guide the worker in finding and picking. Photoelectric barriers are used to detect the successful grabbing process, that is also stored in a database for traceability reasons.

2.4 Electric screwdriver module

The electric screwdriver is a required tool in the assembly process of the studied explosion-proof tube lamp. The required parameters such as torque and speed are set according to the current work step. Also, the screwdriver is only active during the relevant work steps to preserve the correct working sequence. For traceability purposes, the tightening parameters angle, torque, and speed are published and stored in a database. As the electrical screwdriver is necessary in the assembly of the tube lamp, it is used for both the assisted and unassisted assembly workplace. In case of the unassisted assembly, the activation and the correct torque value are set according to the number of screwing processes. In case of a faulty screwing, it has to be set manually for the following screws.

2.5 Visualization module

The work instruction is visualized on a 15" touch screen panel PC. According to the current work step, the respective instruction is presented. After completion and detection by the object detection unit, the next work step instruction is moved on to automatically. For each assembly task, a meaningful image as well as a clear text instruction is presented. Additionally, workers can watch a short video sequence of the assembly at the touch of a button.

3. Experimental setting

The experimental setup refers to the assembly process of a complex tube lamp with support of the proposed assembly assistance system (experimental group) and a control group without said system. Two physically identical workstations are set up for the assembly experiments as shown in Fig. 3. The work instructions for both groups are identical content-wise but paper-based in the control group and screen-based in the experimental group.

The assembly process takes 34 individual work steps including, among other tasks, 21 steps of plugging parts, 5 cable connections, 3 screwing processes, and 2 functionality tests. The assembly parts are made of materials such as plastic, aluminium, steel, glass, and electrical components. Their sizes range from 1 mm cable wires to a 50 cm long tube housing. The large variety of assembly processes, materials, and part sizes ensures the relevance of the assembly error data composition for different kinds of products. In addition, the assembly process diversity makes the results transferable to other products and modern assistance systems.

In both the experimental group and the control group 10 test subjects assembled tube lamps for a duration of 8 hours. During this period, at least 30 tube lamps could be assembled by each test person. The subjects have never assembled the exact lamp before but might have different general knowledge in manual assembly. They were not given any further information on the assembly process except for their work instructions. The whole process was led and supervised by experienced engineers in manufacturing. During the assembly process, the supervisors manually recorded every occurring assembly error.

Fig. 3 The workstation on the left is set up without the proposed assistance system and relies on the paper based work instruction on the far left of the picture. The workstation on the right side is equipped with a pick-by-light system and a work instruction visualization. Those are operated with the data supplied by the object detection unit.

Error assessment

To determine error modes for each work task, an error assessment was performed. As a basis for subsequent analyses, the tasks are described according to the hierarchical task analysis (HTA) method [27]. HTA shows all possible ways of interaction with a system as each task is divided into sub-tasks until no further sub-divisions can be done. Based on the possible assembly states, a human reliability analysis such as THERP (technique for human error-rate prediction) [28] is used to map the states to relevant error categories. As the aim of this work is a comparative error reduction study rather than an error analysis, so THERP was used to identify errors. Due to the high complexity of the assembly product, only safety- and functionality-relevant error modes are taken into further consideration. In total, the study supervisor team identified 70 potential error states. Errors made by the test subjects are counted as such if no error recovery takes place within the next work step. This approach respects the real-world scenario the study is aiming for. Workers typically fix errors but should not reverse too many steps due to time constraints.

According to guideline VDI 4006-2 [29], the identified errors were classified into the four categories described in Table 1. Using error categories provides the results at a higher interpretability level. The difference in error composition between the experimental and control group might be due to specific assistance modules, so each module's influence can be discussed.

Comparing the 70 relevant error states, 18 (25 %) are considered errors by omission, 27 (39 %) execution errors, 9 (13 %) errors by confusion, and 16 (23 %) quantitative errors. The following analysis is intended to compare the error occurrence for the assisted and unassisted assembly lines.

Table 1 Selection of error categories according to VDI 4006-2, that occured during the assembly process of the explosion-proof tube lamp

Type (Abbreviation)	Description	Example
Error by omission (OM)	Some action was not carried out	- Worker forgot to plug in two cable wires - Worker forgot to perform functionality test
Execution error (EX)	Something is wrongly set or selected	- Worker chose a wrong plastic cap - Worker sets wrong torque value for screwing
Error by confusion (CE)	Something is done instead of something else	- Worker reverses positive and negative pole connections - Worker confuses two similar looking plastic parts
Quantitative error (QE)	Something is too much or too little	- Worker applies two layers of insulating film instead of one - Worker only screws 7 of the 8 required screws

4. Results and discussion

To analyse the occurrence of assembly errors, different metrics were determined. The following evaluation considers the first 30 assemblies of the test subjects, as each person was able to complete at least this number of full assemblies during the execution of the experiment.

4.1 Error prevention analysis

To track the test subject's training process over the course of the assemblies, the errors were accumulated for every five assemblies. For the case of this analysis, the errors are counted total without being classified into error types. For both the assisted assembly group (AA) as well as the unassisted assembly (UA) group, the number of errors decreases over the number of performed assemblies (Fig. 5). The AA group records a decrease from an average of 5.0 errors in assemblies 1-5 to an average of 0.20 in assemblies 26-30, which is a decline of 95 %. The UA control group shows 8.5 errors in assemblies 1-5 and 1.3 errors in assemblies 26-30, resulting in a decline of 85 %. While both groups show a training progress in terms of error reduction, the assisted assembly group produces fewer error states at each point in the assembly process. During the first 15 assembly operations the number of errors steadily decreases in both groups. In the following 15 assemblies, the number of assembly errors remains relatively constant at a lower level. This leads to the assumption that there is an initial training phase followed by a skilled work phase.

The total number of errors in 30 individual assemblies decreases from an average of 17.4 in the unassisted assembly group to 9.2 when using the proposed assistance system, which corresponds to an error reduction of 51 %. The results apply to both the initial training phase and the following skilled phase.

The generally smaller interquartile range in the assisted assembly group indicates a smaller statistical dispersion of the number of errors in this group. The lower spread over the number of test subjects signals a more stable assembly environment. The general error reduction effect provided by the assembly assistance system brings a variety of positive effects in the studied use case. First, the required amount of supervision during the training period can be reduced. Second, the number of rejects due to errors is reduced. Third, the assembly time is reduced due to less rework caused by faulty assemblies. Since these positive effects are difficult to quantify, we assume a linear correlation of each factor with the number of total errors.

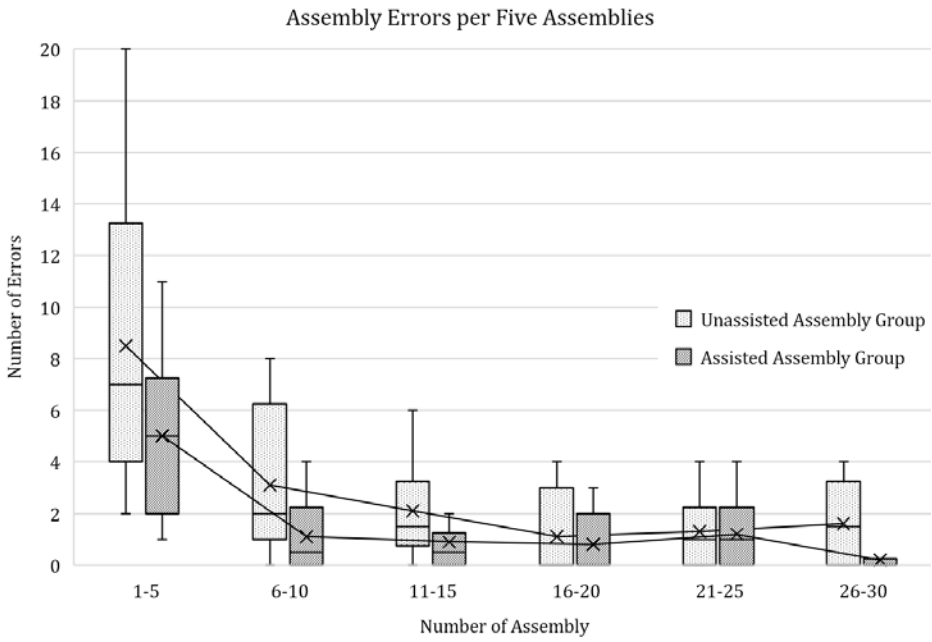

Fig. 4 Number of assembly errors encountered in assembly using the assistance system and without using the assistance system. The errors are accumulated for five assemblies to show the potential training effect during assembly.

To track the assistance system's ability to prevent certain types or error modes, the number of errors per category (Table 1) are analysed. The results show that the AA group produces fewer errors for each of the four error types (Fig. 6). Especially errors by omission (59 % error reduction), quantitative errors (50 % error reduction) and execution errors (42 % error reduction) are effectively prevented by the assembly assistance system. As the system allows the execution of the following work step only if the current work step is visually detectable, errors by omission are actively prevented. Additionally, the pick-by-light system is likely to prevent several types of errors by reducing the cognitive load and the risk of assembling incorrect parts. The control of the electric screwdriver by the assistance system prevents incorrectly set torques and the forgetting of screwdriving operations. The qualitative analysis of the error prevention potential for each assistance module coincides with the quantitative results previously explained.

In the underlying use case, it is important not to trade error prevention for additional assembly time. For this reason, the workplace invasiveness of the system was kept as low as possible. Comparing the assembly times of both groups, AA group is able to assemble even faster on the first runs. After the initial training period, both groups show similar assembly times. The detailed analysis of the assembly times should not be part of this work, but the results indicate that the proposed assembly assistance system does not affect the worker in any undesirable way.

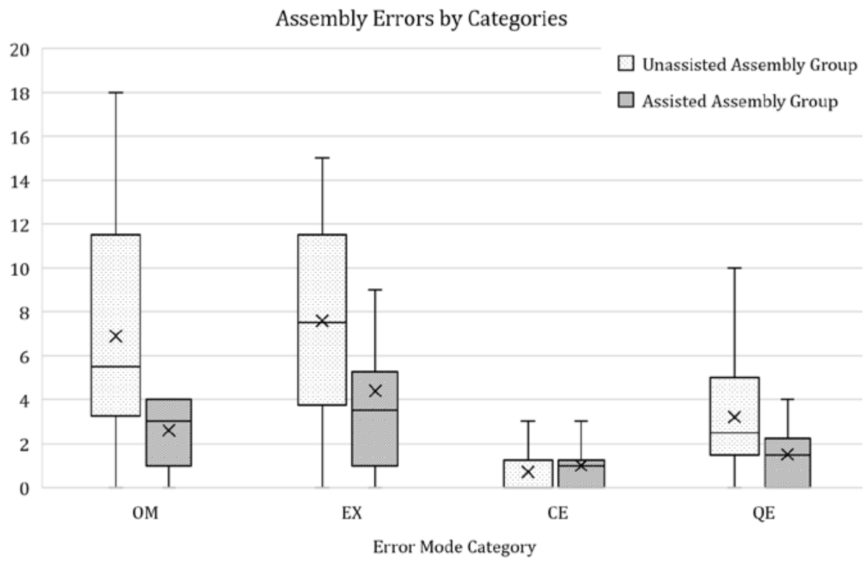

Fig. 6 Number of assembly errors per category summed for all 30 assemblies per test person. The four categories are errors by omission (OM), execution errors (EX), errors by confusion (CE) and quantitative errors (QE).

4.2 Additional insights

In addition to the error prevention aspect, the system's ability to provide traceability should also be discussed. During the execution of the conducted experiments, assembly data of 10 workdays were recorded. In this period, about 600 tube lamps were assembled in a total of 11,000 work steps. The traceability data include the relative position of every visible component at every time in the assembly process. This information provided by the object detection unit can be very valuable in tracing individual assembly errors or providing evidence of critical assembly steps. Especially in safety-critical assembly environments like the presented explosion-proof use case product, this high grade of traceability might a competitive advantage. In addition to storing the acquired assembly data, it can also be processed to gain meaningful insights into the production process and further increase the production efficiency. Unsupervised-learning methods might be used to cluster the data and find irregularities that could lead to the identification of additional errors.

The economic value of the proposed system is highly dependent on the product, labour costs and whether other measures have already been taken to avoid production errors. The cost per error also depends on the industry and potential threat. This external failure cost like warranty expenses, legal costs associated with claims or recall costs are also highly sector and product dependent. Generally, the proposed assistance system will be amortised faster in high-priced product manufacturing and safety regulated sectors like explosion-proof manufacturing. A detailed estimate of the economic value is likely to be the subject of follow-up research. To give a rough estimate for the costs of the proposed assistance system we can say that the hardware equipment (worker display, cameras, GPU, edge computers, etc.) used in our experiments is currently available for about 5,000€. In practice the implementation of such a system will be comparable to a software project including all phases (requirements analysis, implementation, test), which is by far the largest cost factor. Currently, such a project must be considered as an individual software development project because there is no standard software product existing. However, as described above, there are open-source software components available for the training and application of neural networks that can accelerate the development enormously.

5. Conclusion

This research paper studies a real-world manual assembly use case to determine the error prevention potential of a state-of-the-art assembly assistance system. The chosen approach is a combination of the most recent advances in assembly assistance systems and deep learning. Our proposed system relies on a deep learning object detection model to accurately identify the state of each assembly part in real-time. The object detection module supplies data to other assistance modules such as pick-by-light or the electrical screwdriver which are used within the system. In order to analyse the system's error prevention ability, a complex manual assembly product was provided by the company R. Stahl AG. The explosion-proof tube lamp is assembled in 34 work steps using different assembly techniques. Since the product is used in safety-critical environments, finding methods to avoid assembly errors is highly relevant. In experiments with test subjects, it was found that the proposed assistance system can prevent 51 % of the assembly errors compared to a control group without the use of assistance. Especially errors by omission are effectively prevented, as the system supervises the assembly process and notices the non-assembly of components. Due to the number of test persons, the number of individual assemblies and the variety of assembly techniques, this case study can also be representative for other comparable assistance systems and complex products. Additionally, it is discussed how the collected data can further be used as valuable sources of information.

Despite the high error prevention capability, the economic value of implementing the proposed assistance systems remains unclear. In order to determine this value, case-dependent analysis have to be carried out that include the cost of individual errors, labour costs and the value of additional traceability. Although the test subjects did not express any concerns, the acceptance and applicability of the system in a production environment needs to be intensively studied. In addition to the implementation details for the use in a company's production ecosystem, legal aspects also need to be clarified depending on local legislation.

In follow-up studies, the time-saving potential of the assembly assistance system could be investigated, and ways found to quantify its economic benefits. The potential of the recorded assembly data should also be further investigated using machine learning methods.

Acknowledgement

The authors would like to thank the Carl-Zeiss-Foundation for their support within the research project "Smart Assembly". Acknowledgements also go to our industrial partners for providing assembly products and exchange of knowledge.

References

[1] Koren, Y. (2010). *The global manufacturing revolution: Product-process-business integration and reconfigurable systems,* John Wiley & Sons, New Jersey, USA, doi: 10.1002/9780470618813.

[2] Zhou, F., Ji, Y., Jiao, R.J. (2013). Affective and cognitive design for mass personalization: Status and prospect, *Journal of Intelligent Manufacturing,* Vol. 24, No. 5, 1047-1069, doi: 10.1007/s10845-012-0673-2.

[3] Metzmacher, A.I., Hellebrandt, T., Ruessmann, M., Heine, I, Schmitt, R.H. (2019). Aligning the social perspective with the technical vision of the smart factory, In: Schmitt, R., Schuh, G. (eds.), *Advances in Production Research*, WGP 2018, Springer, Cham, Switzerland, 715-729, doi: 10.1007/978-3-030-03451-1_69.

[4] Merkel, L., Atug, J., Berger, C., Braunreuther, S., Reinhart, G. (2018). Mass customization and paperless assembly in the learning factory for cyber-physical-production systems: Learning module 'from paperbased to paperless assembly', In: *Proceedings of 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT),* Mumbai, India, 270-271, doi: 10.1109/ICALT.2018.00130.

[5] Villani, V., Sabattini, L., Czerniaki, J.N., Mertens, A., Vogel-Heuser, B., Fantuzzi, C. (2017). Towards modern inclusive factories: A methodology for the development of smart adaptive human-machine interfaces, In: *Proceedings of 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1-7, doi: 10.1109/ETFA.2017.8247634.

[6] Quint, F., Loch, F., Orfgen, M., Zuehlke, D. (2016). A system architecture for assistance in manual tasks, In: Novais, P., Konomi, S. (eds.), *Ambient intelligence and smart environments, Volume 21: Intelligent Environments,* IOS Press BV, Amsterdam, The Netherlands, 43-52, doi: 10.3233/978-1-61499-690-3-43.

[7] Fast-Berglund, Å., Fässberg, T., Hellman, F., Davidsson, A., Stahre, J. (2013). Relations between complexity, quality and cognitive automation in mixed-model assembly, *Journal of Manufacturing Systems,* Vol. 32, No. 3, 449-455, doi: 10.1016/j.jmsy.2013.04.011.

[8] Hinrichsen, S., Bendzioch, S. (2019). How digital assistance systems improve work productivity in assembly, In: Nunes, I.L. (ed.), *Advances in Human Factors and Systems Interaction*, Springer International Publishing, 332-342, doi: 10.1007/978-3-319-94334-3_33.

[9] Turk, M., Pipan, M., Šimic, M., Herakovič, N. (2020). Simulation-based time evaluation of basic manual assembly tasks, *Advances in Production Engineering & Management,* Vol. 15, No. 3, 331-344, doi: 10.14743/apem2020.3.369.

[10] Zamora-Hernández, M.-A., Castro-Vargas, J.A., Azorin-Lopez, J., Garcia-Rodriguez, J. (2021). Deep learning-based visual control assistant for assembly in Industry 4.0, *Computers in Industry,* Vol. 131, Article No. 103485, doi: 10.1016/j.compind.2021.103485.

[11] Min, Y., Zhang, Y., Chai, X., Chen, X. (2020). An efficient pointLSTM for point clouds based gesture recognition, In: *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Seattle, USA, 5760-5769, doi: 10.1109/CVPR42600.2020.00580.

[12] Börold, A., Teucke, M., Rust, A., Freitag, M. (2020). Deep learning-based object recognition for counting car components to support handling and packing processes in automotive supply chains, *IFAC-PapersOnLine*, Vol. 53, No. 2, 10645-10650, doi: 10.1016/j.ifacol.2020.12.2828.

[13] Ozdemir, R., Koc, M. (2019). A quality control application on a smart factory prototype using deep learning methods, In: *Proceedings of 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT),* Lviv, Ukraine, 46-49, doi: 10.1109/STC-CSIT.2019.8929734.

[14] Pierleoni, P., Belli, A., Palma, L., Palmucci, M., Sabbatini, L. (2020). A machine vision system for manual assembly line monitoring, In: *Proceedings of 2020 International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, 33-38, doi: 10.1109/ICIEM48762.2020.9160011.

[15] Tavakoli, H., Walunj, S., Pahlevannejad, P., Plociennik, C., Ruskowski, M. (2021). Small object detection for near real-time egocentric perception in a manual assembly scenario, *Computer Science, Computer Vision and Pattern Recognition,* Cornell University, from *https://arxiv.org/abs/2106.06403*, accessed October 17, 2021.

[16] Eversberg, L., Grosenick, P., Meusel, M., Lambrecht, J. (2021). An industrial assistance system with manual assembly step recognition in virtual reality, In: *Proceedings of 2021 International Conference on Applied Artificial Intelligence (ICAPAI),* Halden, Norway, 1-6, doi: 10.1109/ICAPAI49758.2021.9462061.

[17] Egeler, R., Wimpff, D.-P., Jauch, C., Wiedenroth, S.J., Wolf, A., Ruck, M., Wohlfeld, D. (2017). MonSiKo – Adaptives Montageassistenz- und Interaktionssystem mittels 3D-Szenenanalyse und intuitiver Mensch-Technik Kommunikation, Schlussbericht, Stuttgart, Germany, from *http://publica.fraunhofer.de/dokumente/N-590407.html*, accessed October 14, 2021.

[18] Oestreich, H., Töniges, T., Wojtynek, M., Wrede, S. (2019). Interactive learning of assembly processes using digital assistance, *Procedia Manufacturing,* Vol. 31, 14-19, doi: 10.1016/j.promfg.2019.03.003.

[19] Kaczmarek, S., Hogreve, S., Tracht, K. (2015). Progress monitoring and gesture control in manual assembly systems using 3D-image sensors, *Procedia CIRP*, Vol. 37, 1-6, doi: 10.1016/j.procir.2015.08.006.

[20] Faccio, M., Ferrari, E., Galizia, F.G., Gamberi, M., Pilati, F. (2019). Real-time assistance to manual assembly through depth camera and visual feedback, *Procedia CIRP*, Vol. 81, 1254-1259, doi: 10.1016/j.procir.2019.03.303.

[21] Pilati, F., Faccio, M., Gamberi, M., Regattieri, A. (2020). Learning manual assembly through real-time motion capture for operator training with augmented reality, *Procedia Manufacturing,* Vol. 45, 189-195, doi: 10.1016/j.promfg.2020.04.093.

[22] Rocha, C.A.P., Rauch, E., Vaimel, T., Garcia, M.A.R., Vidoni, R. (2021). Implementation of a vision-based worker assistance system in assembly: A case study, *Procedia CIRP*, Vol. 96, 295-300, doi: 10.1016/j.procir.2021.01.090.

[23] Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M. (2020). YOLOv4: Optimal speed and accuracy of object detection, *Computer Science, Computer Vision and Pattern Recognition,* Cornell University, from *http://arxiv.org/abs/2004.10934*, accessed October 14, 2021.

[24] NVIDIA deep learning TensorRT documentation, from *https://docs.nvidia.com/deeplearning/tensorrt/developer-guide/index.html*, accessed October 14, 2021.

[25] Redmon, J., Darknet: Open Source Neural Networks in C, from *http://pjreddie.com/darknet/*, accessed October 14, 2021.

[26] Stockinger, C., Steinebach, T., Petrat, D., Bruns, R., Zöller, I. (2020). The effect of pick-by-light-systems on situation awareness in order picking activities, *Procedia Manufacturing,* Vol. 45, 96-101, doi: 10.1016/j.promfg.2020.04.078.

[27] Annett, J. (2003). Hierarchical task analysis, In: Hollnagel, E. (ed.), *Handbook of cognitive task design*, Vol. 2, Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, 17-35.

[28] Kirwan, B. (1994). *A guide to practical human reliability assessment*, First edition, CRC Press, London, United Kingdom.

[29] VDI. VDI 4006 Blatt 2:2017-11, Human reliability – Methods for quantitative assessment of human reliability. (2017), from *https://www.vdi.de/*, accessed October 14, 2021.

[30] Jiang, C., Xi, J.T. (2019). Dynamic scheduling in the engineer-to-order (ETO) assembly process by the combined immune algorithm and simulated annealing method, *Advances in Production Engineering & Management,* Vol. 14, No. 3, 271-283, doi: 10.14743/apem2019.3.327.