Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-024-02533-w

Camera-based safety system for collaborative assembly

Elvira Chebotareva[1] · Maksim Mustafin[1] · Ramil Safin[1] · Tatyana Tsoy[1] · Edgar A. Martinez-García[2] · Hongbing Li[3] · Evgeni Magid[1,4]

Received: 6 June 2024 / Accepted: 13 November 2024

Abstract
Collaborative assembly represents one of the most prevalent practical applications of collaborative robots in intelligent manufacturing. Developing intelligent systems to ensure safety of collaborative assembly processes requires a special attention. In this work, we introduce a visual safety system designed to monitor hazardous situations that may occur during collaborative assembly, potentially resulting in operator injuries. Unlike many other vision-based systems, we solely rely on data from two RGB cameras, without acquiring additional depth information from other sensors. These cameras provide top and side projections of a collaborative workspace. The safety system assesses a current level of a risk by employing two neural network YOLOv8-cls models. These models are pretrained on the ImageNet dataset and subsequently fine-tuned on our dataset. Upon identifying a potential hazard, the system employs our proposed algorithm to determine whether to slow down or halt a robot's motion. Additionally, the system integrates with a visual control system that utilizes an operator gesture control throughout an assembly process. We further conduct experiments to compare our system's assessment with an assessment of human experts. An analysis of the experiments demonstrated a high level of correlation between the evaluations of the autonomous system and the human experts. Benefits of the proposed system encompass its relative cost-effectiveness and ease of setup.

Keywords Intelligent manufacturing · Human–robot collaboration · Human–robot interaction · Collaborative assembly · Vision-based safety system

Introduction

Typically, collaborative assembly entails a higher degree of operational complexity compared to other tasks executed by industrial manipulators (Petzoldt et al., 2023). At the same time, safety concerns regarding collaborative assembly constitute a key area of research that garners significant attention within the intelligent manufacturing scientific community (Keshvarparast et al., 2024; Faccio et al., 2023). An essen-

Elvira Chebotareva, Maksim Mustafin, Ramil Safin, Tatyana Tsoy, Edgar A. Martinez-García, Hongbing Li and Evgeni Magid contributed equally to this work

✉ Evgeni Magid
magid@it.kfu.ru

Elvira Chebotareva
elvira.chebotareva@kpfu.ru

Maksim Mustafin
mustafin@it.kfu.ru

Ramil Safin
safin.ramil@it.kfu.ru

Tatyana Tsoy
tt@it.kfu.ru

Edgar A. Martinez-García
edmartin@uacj.mx

Hongbing Li
lihongbing@sjtu.edu.cn

[1] Laboratory of Intelligent Robotic Systems, Intelligent Robotics Department, Institute of Information Technology and Intelligent Systems (ITIS), Kazan Federal University, 35 Kremlin Street, Kazan 420008, Russian Federation

[2] Department of Industrial Engineering and Manufacturing, Institute of Engineering and Technology, Manuel Díaz H. No. 518-B Zona Pronaf Condominio, 32315 Ciudad Juarez, Mexico

[3] Department of Instrument Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[4] School of Electronic Engineering, Tikhonov Moscow Institute of Electronics and Mathematics, HSE University, Moscow 123592, Russian Federation

Published online: 22 November 2024

tial direction in this domain involves an intelligent safety systems' development (Hanna et al., 2019; Gualtieri et al., 2020).

A development of intelligent manufacturing processes, which involve concurrent collaboration between robots and humans in a shared workspace (including joint assembly) should comply with international safety standards that regulate requirements for human–robot interaction (Galin & Meshcheryakov, 2019; Li et al., 2023). Primarily among these standards is ISO/TS 15066:2016 standard (International Organization for Standardization, 2016), as well as ISO 10218-1 (International Organization for Standardization, 2011a) and ISO 10218-2 (International Organization for Standardization, 2011b) standards, which are awaiting for updates. However, it is noteworthy that these standards offer rather general recommendations. In practice, integrators of collaborative robotic systems often encounter a necessity for customizing safety modes to suit their specific tasks (Bdiwi et al., 2022). Consequently, intelligent safety systems must possess a flexibility in configuration to accommodate particular aspects of implemented cases.

According to ISO 10218-1, mechanical hazards are identified as primary significant hazards that robots may pose. Therefore, in scenarios where a production process does not entail a direct mechanical contact between a robot and a human, one of safety system's objectives should be to avert any collisions between a human and a moving robot (Zhang et al., 2022; Kanazawa et al., 2021). In collaborative assembly settings, where processed parts are exchanged between a robot and a human, this task becomes notably intricate due to a presence of the moving robot and the human within a shared workspace. Presently, this problem remains open and poses a challenge for researchers in the field of intelligent manufacturing (Proia et al., 2022) and seamless human–robot collaboration in industrial applications (Makris et al., 2024).

The ISO/TS 15066:2016 standard provides the following methods of collaborative operations: safety-rated monitored stop, hand guiding, speed and separation monitoring, power and force limiting (International Organization for Standardization, 2016). To prevent a direct mechanical contact, employed practical methods include halting a robot upon detecting a potential collision, reducing a velocity when approaching a person, and altering a robot's trajectory (Scimmi et al., 2019). Our research is motivated by a necessity to create an affordable, swiftly deployable vision-based safety system tailored to monitor and prevent potentially dangerous situations that could occur during collaborative assembly processes. Similar systems are particularly sought after in small and medium-sized enterprises seeking to implement intelligent manufacturing processes that involve collaborative assembly (Cencen et al., 2018). Integrating such systems can enhance overall workplace safety, mitigate risks of staff injuries, and reduce other hazards' probabilities.

To prevent potential emergency situations that could result in operator injuries, collaborative robotic cells are equipped with monitoring systems, which leverage various sensor types, including distance and motion detection sensors, force and inertia sensors, and diverse camera configurations (Cherubini & Navarro-Alarcon, 2021). Collisions between a robot and a human operator can be detected using sensors that are placed directly on the robot (Katsampiris-Salgado et al., 2024b) or employing external monitoring (Katsampiris-Salgado et al., 2024a).

Kamezaki et al. (2024) proposed an approach for constructing a dynamic collaborative workspace based on data from three laser rangefinders. Selvaraj et al. (2023) presented a system with a ZED2i stereo vision sensor for human detection within a collaborative environment. Wong et al. (2024) proposed a method for distinguishing between intentional and unintentional physical interactions with a collaborative robot, utilizing a touch input, a user body posture, and a user gaze.

An example of a practical implementation for collecting data about operator actions during collaborative assembly is an artificial intelligence-based monitoring system presented in Gkournelos et al. (2023), which utilizes data from a static high-definition 3D camera, a wearable 3D camera mounted on an operator's headset, and two wearable IMU sensors.

A significant advantage of robot–human collision avoidance methods based on computer vision is an absence of a necessity to attach additional sensors to a robot or an operator. Commonly employed camera types include RGB-D cameras and stereo cameras (Kozamernik et al., 2023; Maric et al., 2021; Amaya-Mejía et al., 2022). RGB camera-based computer vision systems are less commonly used in collaborative robotics for collision prediction and prevention. However, in some cases, a choice of RGB cameras may be justified due to lower cost, simplified image processing, and reduced computational requirements.

An objective of this research is to design a monocular vision based control system that ensures safe collaborative assembly processes within the intelligent manufacturing paradigm of Industry 4.0 (Barari & de Sales Guerra Tsuzuki, 2021). We address a common scenario where a human and a robot collaborate in assembling a product, with the assembly object alternating between the robot and the human. We consider a collaborative assembly case that involves one operator and a six degrees of freedom industrial manipulator. A distinctive aspect of our scenario is a feedback between the operator and the robot through contactless control, which is based on a recognition of operator's gestures (Mustafin et al., 2023a). A primary objective of our safety system is to avert hazardous situations where the operator may sustain mechanical injuries due to a direct contact with the robot. In this paper, we present the safety system, which relies on image classification from two projections of the collabora-

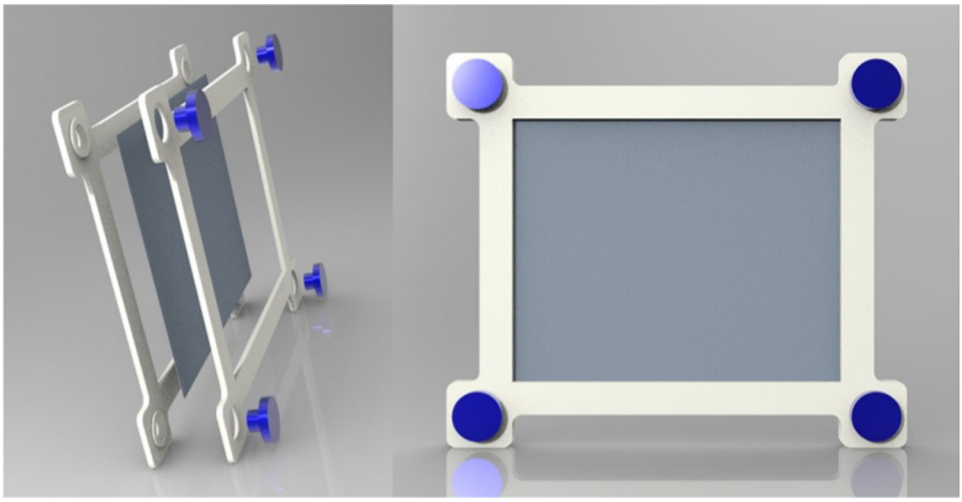

Fig. 1 An assembled object comprising of a small two-layered frame and secured by four blue rivets: a side view before the assembly (left) and a front view of the ready product (right) (Color figure online)

tive workspace. In laboratory settings, we replicated a typical collaborative assembly scenario, assessed a current risk level using our developed system, and gauged a correlation of a system's autonomous assessment with a human experts' assessment.

Materials and methods

Collaborative assembly use case description

We examined a typical collaborative assembly scenario adaptable to situations where some assembly tasks are performed by a human while others are delegated to a robot. In this setup, the human is required to transfer an assembly object to the robot and retrieve it, remaining within a shared workspace and continuing with the tasks. A safety system should continuously monitor motion of the human and the robot, preventing any mechanical contact between the moving robot and the human, as such contact could potentially endanger the human.

Previously, our team conducted a series of pilot experiments (Mustafin et al., 2023a, b), which provided insights into user experience when interacting with a robot through gestures during collaborative assembly. This work was based on the previous findings and utilized the virtual control system UR-VC (Mustafin et al., 2023a) for contactless control of the collaborative robot UR3e during the assembly process. UR-VC system enables control of UR collaborative robots through augmented reality elements, represented by on-screen buttons and simple operator gestures based on fingers' closing and opening.

In this paper, we consider a scenario of assembling a small frame with a picture, which is inserted into the frame and secured with rivets (Fig. 1). An operator selects a paper card with a picture and places it between foreground and background sections of the frame. Subsequently, the robot fastens the two frame parts together using the rivets.

A sequence of the collaborative assembly process is as follows. An operator initiates the object assembly within the shared workspace and places assembly components in a pre-defined area within the shared workspace. Using an appropriate gesture, the operator requests the UR-VC system to initiate the assembly on the robot's side. Next, the robot carries out its part of assembly operations. Once the robot completes its actions, the operator retrieves the assembled object. If the operator needs to pause the assembly, he/she can pause the robot with a corresponding gesture and then resume the robot's actions with another gesture (if necessary). Emergency stop of the robot could be activated with a corresponding gesture or with an emergency stop button.

Collaborative work cell configuration

The collaborative assembly cell (Fig. 2) contains the following equipment:

- Universal Robots (UR) 3e manipulator;
- A work table;
- A web camera for the UR-VC interface;
- A display for presenting operator data and UR-VC system messages;
- Top-view Basler acA1300-200uc camera mounted above the work table at a height of 1.24 m;
- Side-view Basler acA1300-200uc camera positioned at a distance of 3.15 m from the work table.

It is important to emphasize that the cell lacks any physical (security) barriers, which allows an operator to move freely within the shared workspace. Examples of views captured by the two cameras are depicted in Fig. 3.

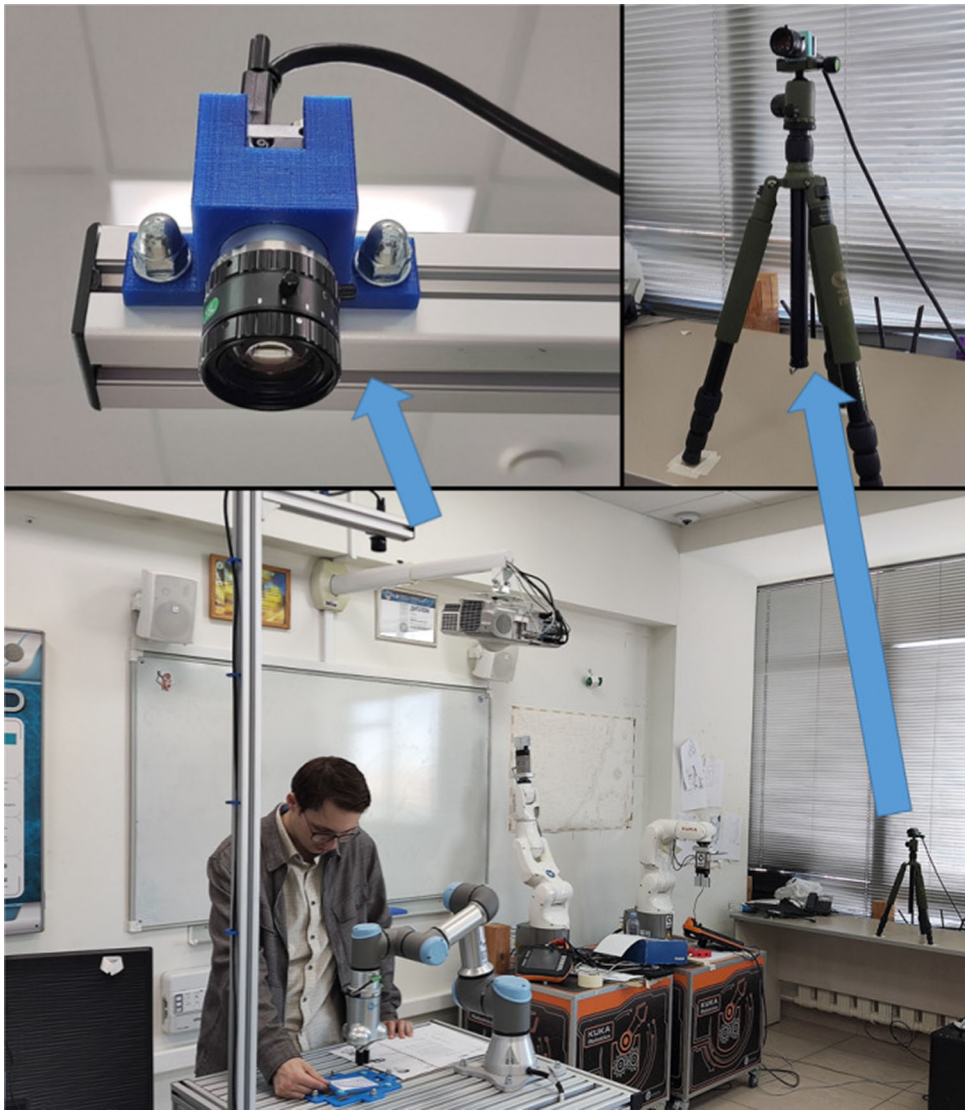

Fig. 2 A collaborative work cell setup: UR3e, an operator, the gesture-based control system

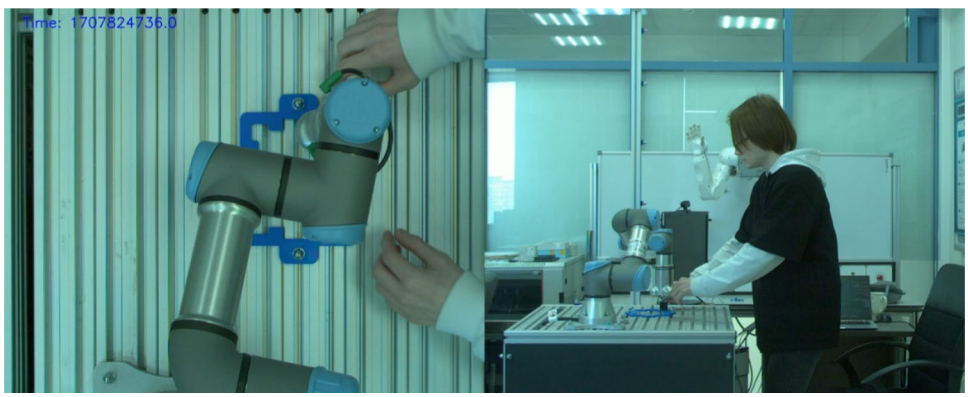

Fig. 3 Views from the top (left) and side (right) cameras

Table 1 Safety zones of the work cell

Zone type	Expected time T to a mechanical contact event
Safe zone	$T > \Delta t_1$
Moderate hazard zone (Increased attention zone)	$\Delta t_2 \leq T \leq \Delta t_1$
Critically hazardous zone (Danger zone)	$T < \Delta t_2$

Assessment of safety levels

The ISO/TS 15066:2016 standard (International Organization for Standardization, 2016) declares hazard identification and risk reduction as prioritized aspects in a development of collaborative robotic systems. In accordance with this, our developed safety system aims to reduce hazards arising from brief or quasistatic contact between an operator and the robot during the collaborative assembly task outlined in Sect. 2.1. According to ISO 10218-1, these hazards primarily comprise mechanical hazards and may result in potential consequences such as impact, crushing, or trapping. Our proposed monitoring system assesses a current level of danger based on visual sensory data from two RGB cameras and makes decisions on continuing a normal operation, limiting joints' velocities, or a full stop of the robot.

We propose to determine a current danger level based on a current location zone of an operator. The cell's workspace was divided into three zones: a safe zone, a moderately hazardous zone, and a critically hazardous zone. A similar simplified approach for a danger level assess is convenient in scenarios that require a real-time decision-making and is frequently employed in intelligent manufacturing (Wang et al., 2023; Malm et al., 2019). A comprehensive example illustrating a design of dynamic safety zones within a hybrid robotic cell can be found in Karagiannis et al. (2022).

Dimensions of the zones were established based on ISO/TS 15066:2016 standard specifications and consider specific characteristics of the manufacturing process in accordance with Table 1. When an operator stays in the safe zone, a mechanical contact between the operator and the robot would occur in a time that exceeds Δt_1 if the operator moves along a shortest path towards the robot. If the operator in the moderately hazardous zone (referred to as an increased attention zone) moves along a shortest path towards the robot, the mechanical contact would occur between Δt_1 and Δt_2 time. Finally, in the critically hazardous zone (termed a danger zone) a mechanical contact would occur in a time shorter than Δt_2, assuming the operator moves along a shortest path towards the robot.

Time parameters Δt_1 and Δt_2 are determined based on available data on the operator velocity, a current velocity of the robot, and a system response time. If the operator velocity is not limited by a technological process, ISO/TS 15066:2016 suggests assuming it as 1.6 m/s in a direction that reduces a separation distance the most. Alternatively, the operator speed can be determined according to other relevant specifications, including an empirical approach. The system response time is determined experimentally during its setup; it involves practical testing and calibration to assess a time required for the safety system to detect potential hazards and react accordingly.

We conducted empirical observations to compute an average velocity of an operator during the collaborative assembly. The assessment considered velocities of the operator's hands, head and upper body. Monitoring male operator assembly operations for 5 min we obtained his average velocity of 0.31 m/s. Using this value along with the system response time and the robot emergency stopping time, we derived the time threshold values as $\Delta t_1 = 0.9$ s and $\Delta t_2 = 0.38$ s. These values indicate that the robot should decelerate its velocity when an operator approaches the moving robot to a distance less than 28 cm and should stop when the operator gets closer than 12 cm. It should be noted that the specified zones' width values were calculated based on characteristics of the particular assembly process and require an adjustment for each specific case, including a new dataset collection and labelling. Figure 4 illustrates a safety zones arrangement relatively to the robot; Fig. 5 demonstrates which particular distances the safety system calculates based on an outcome of the image processing.

Based on the analysis of visual data from the top-view and side-view cameras, the proposed system classifies a robotic cell state into one of three danger levels (Table 2):

- A low hazard level does not require any actions from the safety system and it continues monitoring in the normal mode;
- A medium hazard level indicates that an operator is likely to enter the critically hazardous zone imminently. In response, the system slows down the robot;
- A high hazard level indicates that the robot should be stopped immediately.

A robot velocity is continuously adjusted by the system as a response to a current level of danger according to Table 1.

Fig. 4 The hazard zones layout diagram: Danger zone (red), Increased attention zone (yellow), Safe zone (green)

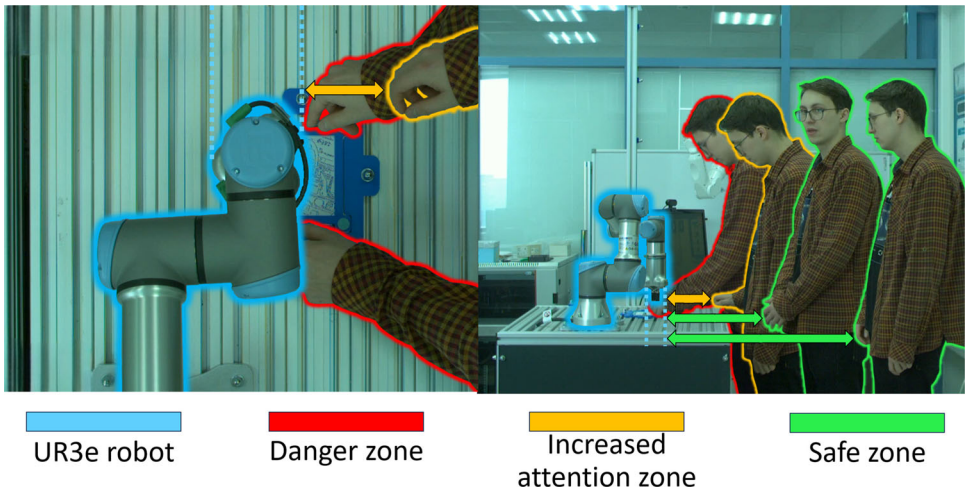

Fig. 5 Examples of the distance calculation between the robot and the operator. In the danger zone the human boundary overlaps with the robot boundary for the top and side cameras' images (red); in the increased attention zone the distance values (the yellow arrows) are within [12,28] cm (yellow); in the safe zone the values of the two examples (the green arrows) are greater than 28 cm (green) (Color figure online)

Table 2 Danger levels identified by the safety system and corresponding responses

Danger level	Actions of the safety system	The robot state
Low	Workspace monitoring	The robot operates in the normal mode
Medium	The robot velocity limitation. An operator is notified of the increased hazard level	The robot starts moving at a limited speed
High	The robot stops. An operator is informed of the high hazard level and is requested to exit the hazardous zone	The robot is stopped and will not resume the operation until further instructions from the system

In the safe zone a maximal angular velocity of the joints is limited to 60 deg/s with a tool speed of 250 mm/s when moving and 100 mm/s during positioning. In the increased attention zone the maximally allowed angular velocity of the joints decreases to 12 deg/s, with the tool speed reduced to 50 mm/s when moving and 20 mm/s during positioning. In the danger zone, the robot halts all joints. In addition to limiting the joint velocities, the safety system displays on the work cell monitor a corresponding message for each danger level.

The two following sections provide a detailed description of how our safety system assesses a current level of danger based on the analysis of visual data.

Visual data processing

RGB cameras are less frequently employed in collaborative robotics for identifying and preventing hazardous situations, unlike depth cameras and stereo cameras. However, their availability, compactness, and ease of use render them appealing for a widespread adoption. Consequently, in this project, we intentionally opted for this type of a visual sensor for the safety system. Nonetheless, the proposed approach can be implemented with other types of cameras as well.

We assess a current danger level using visual data from two RGB cameras. As depicted in Fig. 3, the top-view camera's field of view captures the operator's hands and a portion of the robot during collaborative work, whereas the side-view camera captures the operator and the robot from the side. With this information, the safety system determines the appropriate danger level (Table 1) corresponding to a current situation.

Various approaches can be applied for analyzing observations of a human–robot collaborative work process. The first approach entails determining a current position of an operator relative to the robot, addressing the question "Where is the operator?" [e.g., (Rodrigues et al., 2022a, 2023; Forlini et al., 2024)]. To achieve this, the safety system implements real-time human recognition or body part recognition in captured by the two cameras images. The system measures a distance from the human to a source of danger (the moving robot). This approach enables the system to analyze a spatial relationship between the operator and the robot, facilitating an assessment of potential hazards and appropriate safety responses.

The second approach involves addressing a question "Is the controlled area free from foreign objects?" [e.g., (Saleem et al., 2024)]. In this scenario, the safety system assesses whether there are foreign objects within a designated zone and whether a zone's appearance deviates from its normal state. This approach focuses on detecting any anomalies or unexpected objects within the workspace, allowing the system to identify potential hazards and take appropriate safety measures.

To address these questions, each approach can utilize various computer vision and deep machine learning methods, including image classification, object detection, image segmentation, and feature extraction. For example, successful solutions for determining an operator's position in a video frame reported application of the MediaPipe framework (Lugaresi et al., 2019) and YOLOv8 Pose family models (Ultralytics, 2024a). These human pose detection methods allow achieving high accuracy in answering a question "Where is the human?" under conditions favorable for human detection in an image. However, these methods carry risks of encountering false negatives, wherein despite an actual presence of an operator in a hazardous area, a computer vision system fails to detect it. Such situations may arise, for instance, when an operator is obscured by a robot or when an operator's position deviates from typical instances in a dataset on which a model was trained.

The risks can be reduced by answering a question "Is the controlled area free from foreign objects?" and by focusing not on tracking an operator's position, but on determining a state of the hazard zones. This approach seems more reliable since it is simpler to identify a limited range of safe states of the system in which foreign objects are absent in hazardous areas, rather than analyzing a diverse range of situations related to a human behavior in a work area. However, this method may lead to false positives, where dangerous situations could be mistakenly identified as safe.

A combination of the two approaches could minimize risks associated with false positive and false negative detection. At the same time an additional analysis of visual data in real-time increases computational costs and may significantly increase the system's response time. Therefore, we focused on determining a current state of the observed system by tracking if an operator is located in the safe zone, in the increased attention zone or in the danger zone.

In this study, we used the YOLOv8-cls model for image classification, which is provided by the Ultralytics library (Ultralytics, 2024b). However, the proposed method permits using any other model that is suitable for image classification. The YOLOv8-cls model was selected due to its high speed, sufficient accuracy, low computational resource requirements, and ease of use.

To prepare datasets for classifying images from the top and side cameras, we used a video of the standard assembly process. A pre-trained experienced operator performed a complete assembly cycle several times in a standard mode, adhering to all safety regulations. The operator's safety was also controlled by an instructor using an emergency stop button, which the instructor could press at any time. It should be noted that reproducing hazardous situations with a moving robot for collecting training data is not possible under real

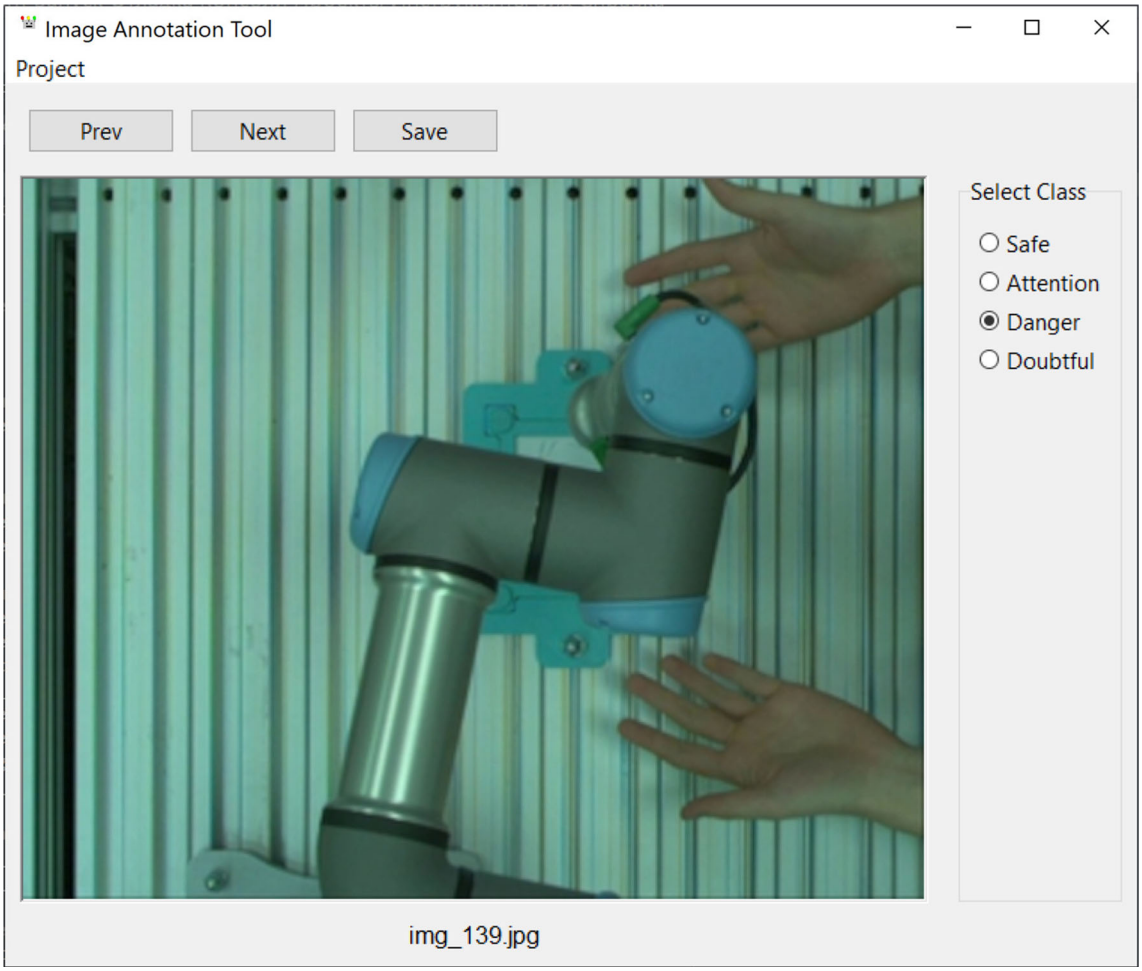

Fig. 6 The annotation tool graphical interface: with buttons above an image a user can move forwards and backwards between images and save results; to the right of the image one of four options could be selected for the annotation, which cover the three hazard classes and inability to make a deterministic decision (be selecting *Doubtful* option)

conditions, as it contradicts the safety standards for operating industrial robots and puts an operator at risk. Therefore, for gathering images that correspond to non-standard situations, we employed a stationary unpowered robot.

We invited two experts with experience in industrial robotics to annotate dataset images. The experts were asked to classify images from two datasets (from the top and side cameras) into one of the three hazard classes using their description (Sect. 2.3). Next, all annotated images were checked for compliance with the established criteria and combined into a final dataset for model training. To make the annotation process more convenient for experts, we developed an application that simplifies the image distribution into the three classes (Fig. 6).

A dataset from the top camera consisted of 2520 images and a dataset from the side camera comprised 980 images. With these datasets, two YOLOv8n-cls models were trained to recognize three classes of images: "Safe", "Attention", and "Danger". Two separate models classified images from the top camera and the side camera. During dataset labelling, images of the operator in the safe zone were assigned to the "Safe" class, in the moderate hazard zone were assigned to the "Attention" class, and in the critically hazardous zone were assigned to the "Danger" class.

For training, we utilized the base architecture YOLOv8n-cls (Terven & Cordova-Esparza, 2023), employing the pre-trained on the ImageNet dataset (Deng et al., 2009) YOLOv8n-cls model. Each dataset for the top and side cameras was split in an 80/20 ratio for training and validation. The training was conducted over 100 epochs and involved the cross-entropy loss function. Optimization was performed using the Adam algorithm with an initial learning rate of $2.3632e^{-04}$, which was reduced to $2.1277e^{-05}$ by the 100th epoch. The batch size was set to 32.

The model trained on top camera images achieved a validation accuracy of 0.92, while the model trained on side

camera images achieved a validation accuracy of 0.93. After processing an image, each of the two trained models provided a set of three values corresponding to the current image's classification into one of the three classes: "Safe", "Attention", and "Danger". The safety system aggregated these data and assessed a current level of danger based on an algorithm described in Sect. 2.5.

Danger level evaluation

This section presents a method for determining a current level of danger based on the classification of images from the top and side cameras. In accordance with Table 1 we consider the following incompatible events:

$$H_1 = \{\textit{The operator is in the safe zone}\},$$

$$H_2 = \{\textit{The operator is in the increased attention zone}\},$$

$$H_3 = \{\textit{The operator is in the danger zone}\}.$$

Let $p_{top} = \left[p^1_{top}, p^2_{top}, p^3_{top}\right]$ be a vector of probability estimates from the top-view camera image at the given moment, where:

- $p^1_{top} = P_{top}(H_1)$ is a probability that the operator is in the safe zone in the top-view camera image;
- $p^2_{top} = P_{top}(H_2)$ is a probability that the operator is in the increased attention zone in the top-view camera image;
- $p^3_{top} = P_{top}(H_3)$ is a probability that the operator is in the danger zone in the top-view camera image.

Let $p_{side} = \left[p^1_{side}, p^2_{side}, p^3_{side}\right]$ be a vector of probability estimates from the side view camera image at a given time, where:

- $p^1_{side} = P_{side}(H_1)$ is a probability that the operator is in the safe zone in the side-view camera image;
- $p^2_{side} = P_{side}(H_2)$ is a probability that the operator is in the increased attention zone in the side-view camera image;
- $p^3_{side} = P_{side}(H_3)$ is a probability that the operator is in the danger zone in the side-view camera image.

Let $\bar{H}_i$ denote an event that is an opposite to event H_i, $i = \overline{1,3}$. If we consider $P_{top}(H_i)$ and $P_{side}(H_i)$ as probabilities that an operator's location in a specified zone were obtained using the top view and side view cameras, respectively, then $P_{top}\left(\bar{H}_i\right) P_{side}\left(\bar{H}_i\right)$ represents a probability that both assessments were incorrect. Then, a probability that at least one assessment was correct can be calculated using the following equation:

$$P(H_i) = 1 - P_{top}\left(\bar{H}_i\right) P_{side}\left(\bar{H}_i\right).$$

Thus, a probability that at a given moment the operator is in a particular zone, provided that at least one of the two assessments was correct, can be calculated using the following equation:

$$P(H_i) = 1 - \left(1 - p^i_{top}\right)\left(1 - p^i_{side}\right), i = \overline{1,3}. \quad (1)$$

As a result, a set of values $p_i = P(H_i)$, $i = \overline{1,3}$, combines the assessments from the top and side view images for each class of the danger level.

Next, the danger level estimate at the given moment is carried out as follows. Predefined threshold values of probability estimates are set as follows:

- T_1 for the safe zone,
- T_2 for the increased attention zone.

The T_1 and T_2 values reflect a level of confidence established through the expert assessment and depend on specifics of an implemented case. Experts are provided with a set of image pairs from the top and side cameras, where the operator is located in one of the three specified zones. Each pair of images is supplied with p_1, p_2 and p_3 values, which are calculated according to Eq. (1). We recommend using a set of at least 100 images with all three levels of danger being represented in similar proportions, i.e., at least 30% of images for each zone of Table 1. The experts are invited to choose one of the following confidence levels: 0.99, 0.95, 0.9, or to specify their own confidence level, which should be in the range from 0.5 to 1. In this study, we used the values: $T_1 = T_2 = 0.9$. Then, for a set of values p_1, p_2, p_3 a value of $p_{max} = max\{p_1, p_2, p_3\}$ and its index, i_{max}, are calculated. In cases where equal probability values are detected, a value with a greater ordinal index is considered a priority.

If $i_{max} = 1$ and $p_{max} \geq T_1$, then a zone in which the operator is located at the given moment is evaluated as safe. Next, if $i_{max} = 1$ and $p_{max} < T_1$, then an index of a greatest among the remaining two probability estimates is determined. If this index is 2, then an operator location is evaluated as the increased attention zone; otherwise, this zone is evaluated as the danger zone.

If $i_{max} = 2$ and $p_{max} \geq T_2$, then the zone in which the operator is located at that moment is evaluated as the increased attention zone. Otherwise, if $i_{max} = 2$ and $p_{max} < T_2$, then an index of a greatest among the remaining two probability estimates is determined. If this index is 3, then an operator location is evaluated as the danger zone; otherwise, this zone is evaluated as the increased attention zone.

Table 3 Technical specifications of the computing nodes

Node ID	Processor	RAM	SD storage	I/O
1	Intel Core i7 4 cores @ 3.2 GHz	8 GB	256 GB	USB 3.0 Type-A @ 5 Gbit/s Ethernet 1000 Mbit/s
2	Intel Core i7 4 cores @ 3.1 GHz	16 GB	2.5 TB	USB 3.0 Type-A @ 5 Gbit/s Ethernet 1000 Mbit/s

If $i_{max} = 3$, then the zone in which the operator is located at the given moment is evaluated as the danger zone.

This algorithm allows the safety system evaluating a present risk level and sending corresponding control commands to the robot. If one of the cameras malfunctions and thus its image assessment is impossible, the safety system immediately activates an emergency stop scenario.

Safety system architecture and implementation

The algorithm was implemented using Ubuntu 20.04 operating system, Robot Operating System (ROS) Noetic framework and the Python3 programming language. The key libraries employed for the implementation include: RosPy for working with ROS and Python3; Ultralytics for image classification using the YOLOv8n-cls model; OpenCV, NumPy, and CVBridge for image processing; socket for data transmission between computing nodes.

The hardware setup included two Intel Core i7 computers: the first with 8 GB RAM and 256 GB storage, and the second with 16 GB RAM and 2.5 TB storage; both models support USB 3.0 Type-A at 5 Gbit/s and 1000 Mbit/s Ethernet connectivity (Table 3). Two Basler acA1300-200uc cameras were used for image capturing.

Figure 7 presents a safety system architecture. The system comprises a decision-making module based on visual control of two projections of the shared workspace from the top-view and side-view RGB cameras. Utilizing data from the two cameras, the safety controller evaluates a current level of danger and determines further actions in accordance with the algorithm (Sect. 2.5).

The safety controller transmits commands to a robot controller and a safety status reporter module. The later informs an operator about a current level of danger and provides instructions on further actions. Emergency stop can be activated in three ways: via the emergency stop button of the robot, using the UR-VC AR interface, and through the safety controller module. Furthermore, the safety controller prohibits the robot from restarting after emergency stop if the hazardous situation remains unresolved. The safety system involves the following computing nodes (Fig. 8):

- Computing Node #1: a central computing node is responsible for the safety system, the manipulator control system, and data acquiring from the top view camera.
- Computing Node #2: an auxiliary computing node that performs synchronized aggregation of data from the two cameras, processes data with computer vision methods, and logs data.
- Collaborative Robot Manipulator UR3e.

The computing nodes interact with each other to ensure safe and efficient control of the manipulator within a collaborative environment. The computing nodes exchange data through a local area network (LAN) with a bandwidth of 1000 Mbps employing ROS 1 Noetic via topics and services over TCP/UDP. The specifications of the computing nodes are detailed in Table 3.

The two Basler acA1300-200uc cameras have a resolution of 1280x720 pixels, a frame rate of 30 frames per second, and an image format of BGGR8. The cameras are linked to the computing nodes via USB cables. The top camera provides the most optimal view in terms of distance and visibility of the workspace, detecting foreign objects in the workspace and its close vicinity. The side camera detects foreign objects at a longer distance from the work cell.

To experimentally evaluate system effectiveness it is necessary to record calculated danger levels (Table 2). Experimental data were recorded in a CSV format log file. Each line of the file represents a separate entry containing the following fields:

1. **Time** denotes elapsed time since a system's launch, measured in seconds using ROS Time. This facilitates precise event timing during the experiment and enables temporal data analysis.
2. **Danger class** represents the hazard level class, which can take one of three values: "0: Safe"—indicates an absence of a danger or a low risk; "1: Attention"—indicates potentially hazardous situations, requiring an attention of an operator or the safety system; "2: Danger"—indicates a situation with a high level of a hazard, requiring urgent safety measures.
3. **First model assessment** is a probabilistic assessment that analyzes data from the top camera and draws conclusions about a current level of danger or safety.

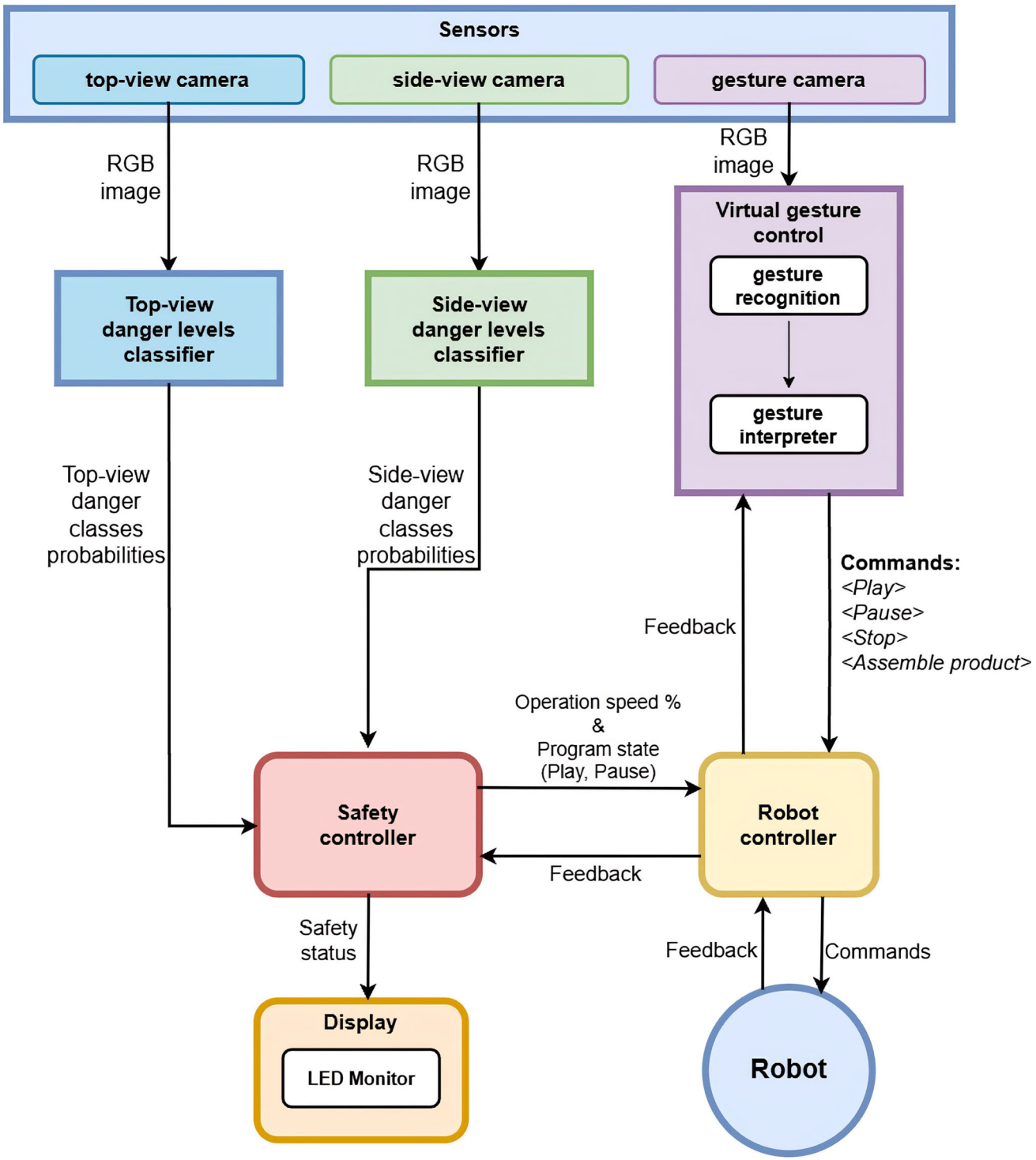

Fig. 7 Architecture of the vision-based safety system

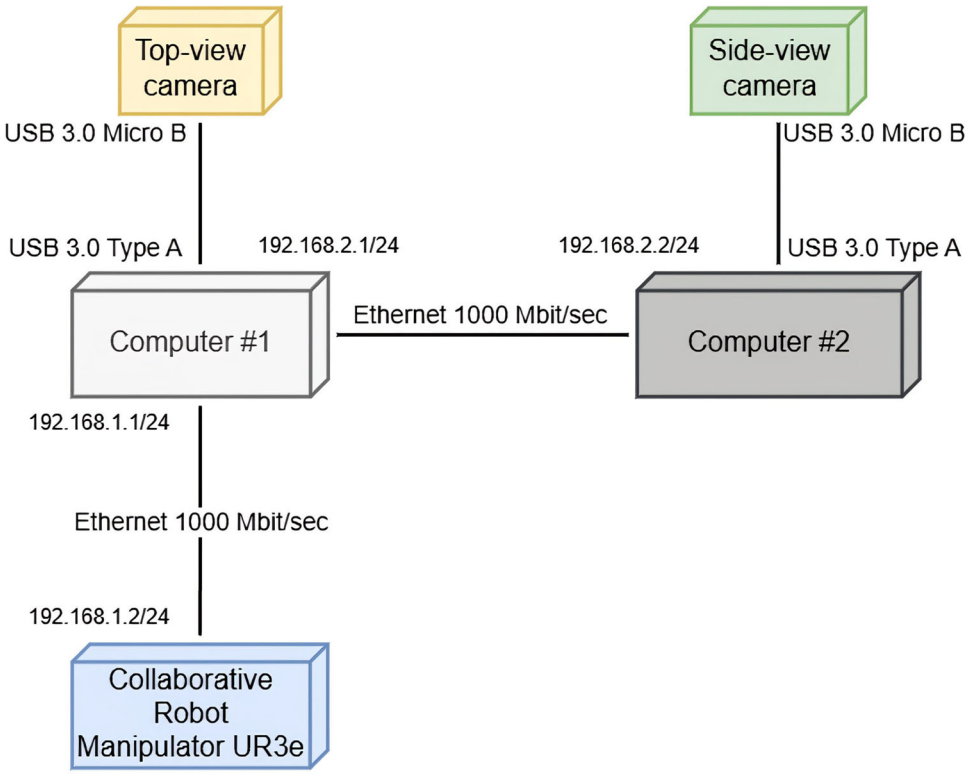

Fig. 8 Deployment diagram. Computing nodes and the manipulator are connected via Ethernet LAN. The top-view and side-view cameras are connected to the corresponding computing nodes via USB

4. **Second model assessment** is similar to the first model; it analyzes data from the cameras and independently draws conclusions about a current safety level.
5. **Overall assessment** is an overall probabilistic safety assessment that can be obtained by aggregating assessments from different models.

Data logging occurs at 1-s intervals preserving a system state and safety level information at each stage of the experiment for a subsequent analysis. Additional data logging may encompass details regarding a manipulator's state, environmental parameters (such as temperature, humidity, etc.), and other factors that could influence the system's safety. This approach facilitates a more comprehensive data analysis and an identification of potential risk factors.

An image frames synchronization is accomplished using the ROS Message Filter (*Time Synchronizer*). This method allows synchronizing data from different sources based on ROS timestamps and ensures alignment between frames from different cameras with a controlled level of accuracy. Aggregated and synchronized frames from the cameras are saved as a file with timestamp annotations. A resulting video has a frame rate of 25–30 frames per second. A temporal annotation of recorded data relies on the ROS Time API, ensuring a precise time synchronization between the system's components. This guarantees an unambiguous identification of time points for each video frame. Furthermore, the computing nodes are time-synchronized using the Network Time Protocol (NTP), ensuring overall time consistency among the nodes. This is crucial for a consistent operation of the safety system and an accurate temporal annotation of experimental data.

To ensure minimal latency of the safety system, we employed a simple optimization mechanism. Received from a camera images are stored in corresponding buffers of a Python3 *deque* format, which is a list-like object that supports fast append and pop operations with first and last elements.

Each received image I is labeled with a timestamp ts that is employed for further comparisons of images in terms of their relevance. The timestamp allows understanding which image contains newer data about the working cell state. First received image I_1 is stored in the buffer together with its timestamp $ts(I_1)$. Timestamp $ts(I_K)$ of new image I_K that arrives to the buffer is compared with $ts(I_1)$: if $ts(I_K) \leq ts(I_1)$ than image I_K is considered to be outdated relatively to I_1 and is deleted from the buffer; otherwise, the first image I_1 is replaced with I_K. Finally, only image I_1 (with the most relevant timestamp) is classified by the trained model.

The optimization eliminated an undesirable behavior of the algorithm when in the interval between receiving a next camera image and successfully classifying a previous image, the algorithm managed to classify other images in the buffer that had been already outdated, which caused a slowdown of the entire safety system. The relevant images' selection from the top-view and side-view cameras allowed achieving

a minimum delay between an action performed by a human in the work cell and processing a frame of this action by the system; the delay is defined by an image capturing speed (fps) of the camera.

The proposed optimization significantly increased the safety system speed achieving the average response time (which is required for the robot to adjust its velocity when a human enters the increased attention or danger zones) of 52.8 ms.

Experimental protocol

The experiments aimed to validate the developed system by comparing its assessments with those of human experts. We engaged specialists with several years of experience in robotics and industrial manipulator safety, who are familiar with the current safety standards in collaborative robotics and intelligent manufacturing. In our case, using expert assessment as a starting point was convenient for several reasons. Firstly, with properly trained experts, the expert assessment aggregates the requirements of the current standards and allows for a quick detection of dangerous situations. Secondly, by observing the collaborative assembly process, an expert could identify requiring attention aspects that are not covered by the standards due to particular features of the manufacturing process. It should be noted that the experts, which assessed the safety system, did not communicate with the experts that had annotated the images for the system training.

Before conducting the experiment, an operator received necessary instructions and underwent training, which included:

- Instructions on all stages of the assembly process,
- Instructions on all work operations and types of operator actions,
- Instructions on the relevant types of operations performed by the robot,
- Instructions on the specification of the chronological sequence of all types of actions,
- Safety training,
- Knowledge assessment.

The experiment comprised two stages. In the first stage, the operator (who had previously undergone the training) executed the task for 8 min. All operator's actions during this period were recorded using side and top cameras. Concurrently, our developed safety system conducted continuous monitoring and control of the work area. In the second stage, the safety level assessments of the side and top cameras' videos were independently conducted by two experts. Each expert was asked to review the recordings and indicate time codes corresponding to the following situations:

- The operator is near the robot to an extent where a robot's high velocity poses a potential risk;
- The operator is dangerously close to the robot, posing a high risk of a mechanical injury.

The experts were provided with video recordings that displayed views from both the top and side cameras simultaneously (Fig. 3). The experts were not limited in time and could pause the videos at any frame, scroll through to any frame, repeat multiple times, slow down and speed up the videos. The assessments were independent and the experts did not communicate with each other.

Table 4 Number of safe sequences, attention-requiring situations and dangerous situations identified by the safety system and experts

Danger level	Safety system	Expert 1	Expert 2
Safe	257	250	273
Attention	25	40	50
Danger	202	194	160

Experimental results

Table 4 describes numbers of safe sequences, attention-requiring situations (where the robot's velocity should be reduced) and dangerous situations (where the robot should be stopped) identified by the safety system and the experts. During the first stage of the experiment, the safety system identified 257 safe sequences, 25 cases requiring increased control and 202 dangerous situations. During the second stage of the experiment, the first expert identified 250 situations requiring increased attention and 40 dangerous situations. Meanwhile, the second expert identified 73 situations requiring increased attention and 50 dangerous situations.

Figure 9 shows examples of frames with safety system evaluations obtained during the first stage of the experiment. Figure 9a–c demonstrate successful identification of the danger level in accordance with Tables 1 and 2. Table 5 presents outcomes of images processing shown in Fig. 9a–c.

Figure 10a–c illustrate a timeline diagram of safety level assessments obtained from the system and two experts during the experiment. The vertical axis distinguish the three levels of a danger: a safe sequence (0), an attention-requiring situation (1) and a dangerous situation (2). The horizontal axis presents time of the experiment.

It should be noted that throughout the experiment, there were no actual collisions between the operator and the mov-

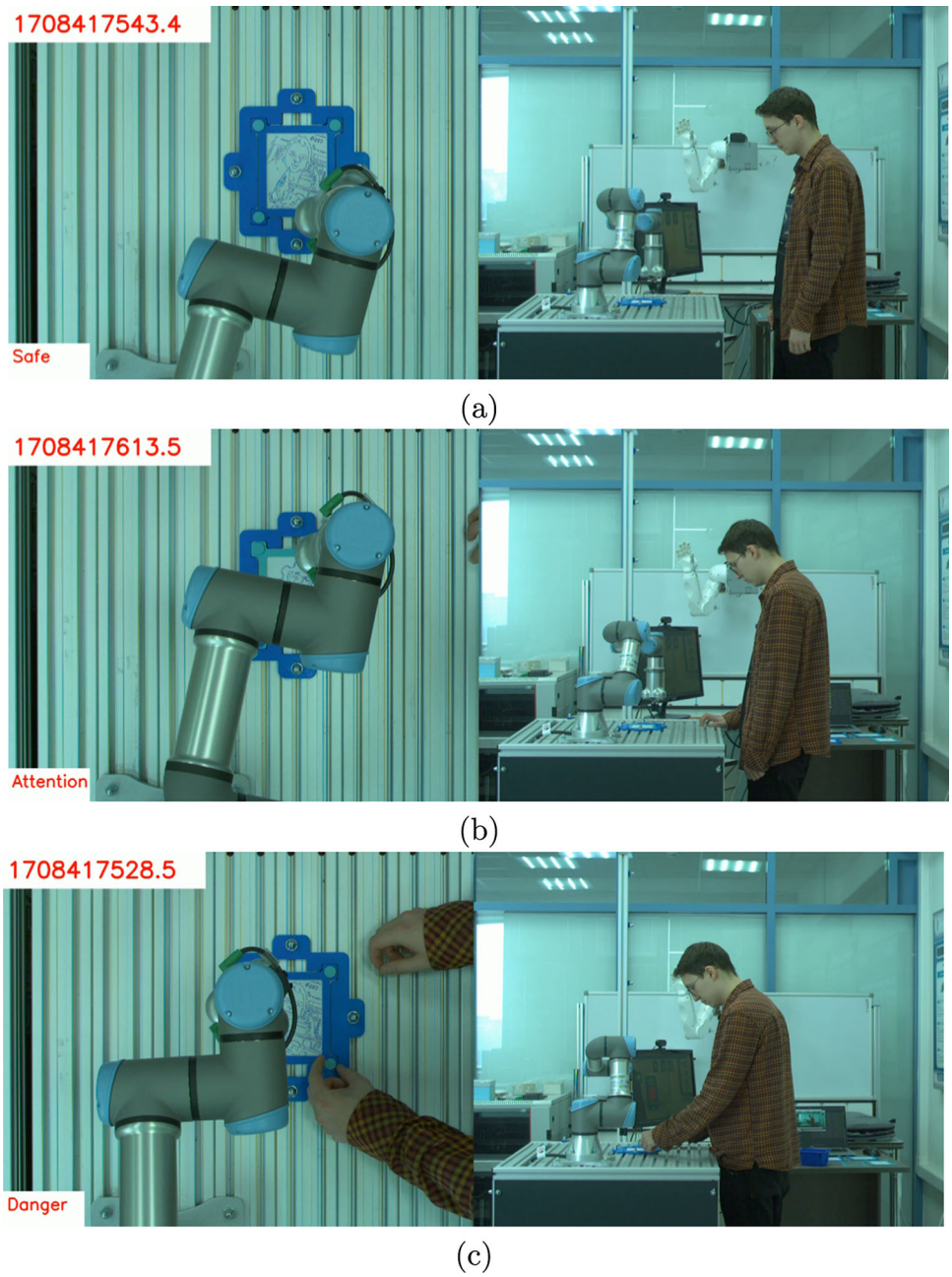

Fig. 9 Examples of 2-cameras combined frames with safety system evaluations obtained during the first stage of the experiment. **a** The operator is in the safe zone. **b** The operator is in the increased attention zone. **c** The operator is in the danger zone

Table 5 Outcomes of the image processing presented in Fig. 9

Estimation	Danger level		
	Low	Medium	High
p_{top}	[0.996986,0.002992, 0.000021]	[0.001931, 0.99408, 0.003989]	[0, 0.000002, 0.999998]
p_{side}	[0.9948, 0.001328, 0.003871]	[0.803264, 0.08751, 0.109225]	[0.000012, 0.00036, 0.999628]
p	[0.999984, 0.004317, 0.003892]	[0.803644, 0.994598, 0.112779]	[0.000012, 0.000361, 0.999999]

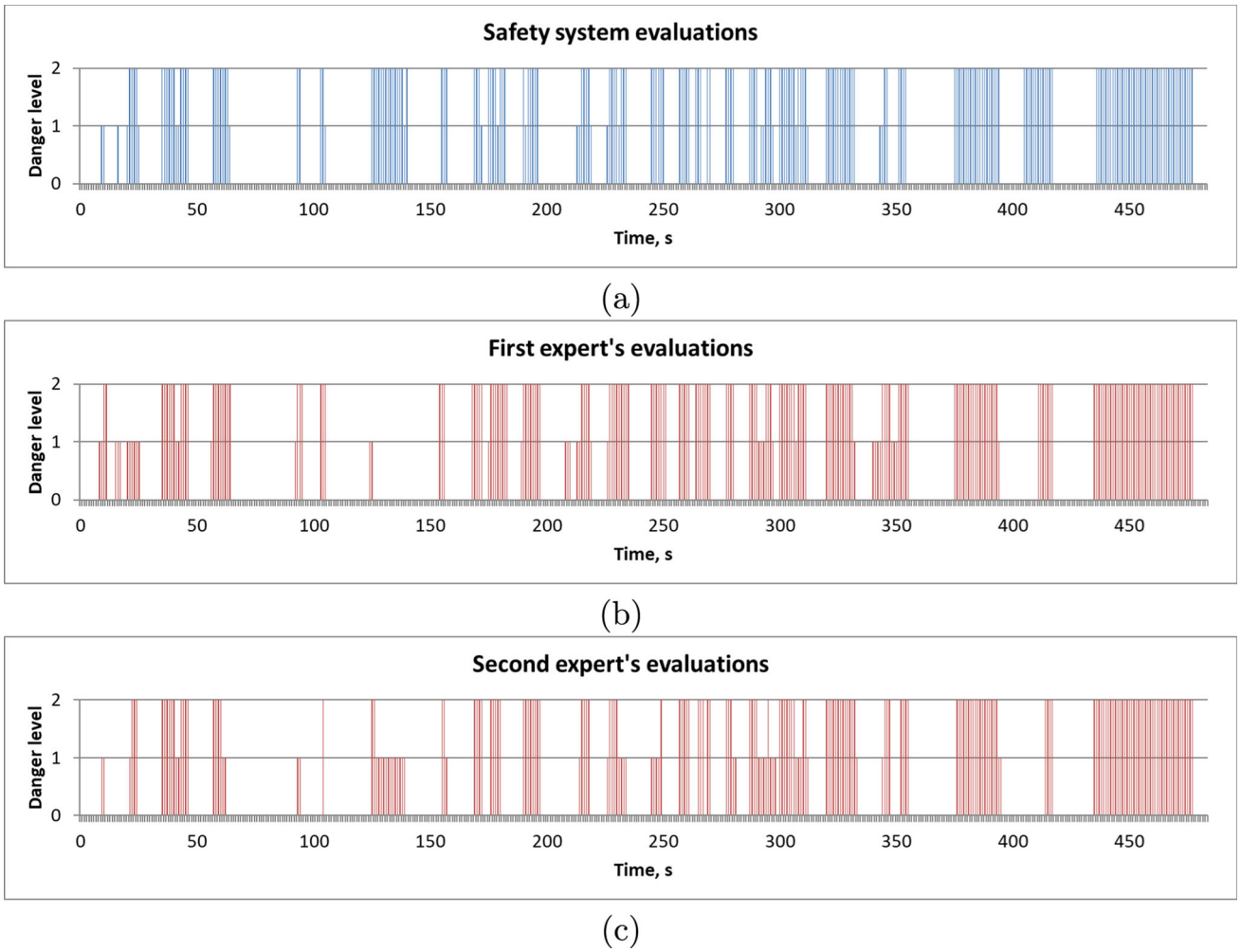

Fig. 10 Timeline diagram of safety level assessments obtained from the system and the two experts during the experiment. **a** Safety system evaluations. **b** Evaluations of the first expert. **c** Evaluations of the second expert

ing robot. The system slowed down the robot when the operator approached it, and stopped the robot if the operator approached the robot too closely. In addition, one of the authors continuously controlled the emergency stop button during the experiments.

Discussion

As a statistical measure of agreement between the evaluations from the safety system and the experts we used Cohen's kappa metric (Vieira et al., 2010). It measures a degree of agreement between two experts' assessments taking into account a chance agreement that could occur if the experts were giving assessments randomly. The Cohen's kappa coefficient is calculated as follows:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

, where P_o is a relative observed agreement among the experts, calculated as a ratio of a number of matching ratings to a total number of ratings, P_e is a hypothetical probability of a chance agreement. In our case, $P_e = \frac{1}{3}$ because with a random selection, the assessment would be chosen randomly from three possible danger levels: "Safe", "Attention", and "Danger".

Table 6 presents results of calculating the Cohen's kappa coefficient of comparing the safety system's ratings with those of each expert and comparing the ratings of the experts with each other. The safety system's vs. the first expert ratings achieved $\kappa = 0.78616$; the safety system's vs. the second expert ratings demonstrated $\kappa = 0.77996$. These values of the Cohen's kappa coefficient indicate a very good agreement between the safety system's ratings and the experts' ratings. However, when comparing the experts' own ratings, a slightly lower value of $\kappa = 0.72727$ was obtained, which still signifies a good agreement between the experts' ratings.

Table 6 The values of the Cohen's kappa coefficient for evaluations obtained from the safety system and experts during the experiment

Evaluations	Safety system, Expert 1	Safety system, Expert 2	Expert 1, Expert 2
Value of κ	0.78616	0.77996	0.72727
Level of agreement	Good	Good	Good

These discrepancies arose from situations, which the experts did not consider to require the increased control, whereas the safety system classified them as such. This is clearly visible in the diagrams presented in Fig. 10a–c. The observed discrepancies in the ratings are partially explained by peculiarities of the implemented algorithm: the danger level evaluation may be excessive in doubtful cases as the algorithm makes a choice in favor of the worst prediction.

The experiments have shown a potential applicability of the approach based on the joint use of convolutional neural networks for image classification in monitoring and preventing dangerous situations in the robotic cell's workspace during collaborative assembly. An advantage of this approach is its speed and ease of setup. Training two pre-trained models requires relatively small datasets, and a process of annotating such datasets takes significantly less time compared to annotating datasets for object detection, image segmentation, or human pose estimation with a keypoint indication.

To collect a dataset for our system, it is sufficient to capture video sequences in three types of situations:

- Situations where an operator is located in the safe zone;
- Situations where the operator is located in the heightened control zone that requires the robot reducing speed to avoid a collision;
- Situations where the operator is located too close to the robot, posing a high risk of injury.

When collecting images of hazardous situations, there is no need to expose the operator to a real danger. Instead, the industrial robot is positioned in various configurations, and videos of the operator approaching the stationary robot are captured. Due to the simplicity of the dataset collection process and a relatively short time required for model retraining, our proposed approach can be quickly tailored to specific production cases of intelligent manufacturing.

While exploring existing solutions, which could serve as alternatives to the proposed safety system, we noticed that often provided by authors data are rather brief and piece-wise. Therefore, for a comparative analysis in terms of practical applications, we selected only three computer vision-based systems designed to facilitate safe collaboration between a human and a robot; these three are most closely aligned with our development. The first system by Amaya-Mejía et al. (2022) utilized a single Kinect camera to monitor a UR3 robot and humans within a shared workspace. The second system by Forlini et al. (2024) employed three Intel RealSense D455 RGB-D cameras to observe a UR3 robot and a human. The third system by Katsampiris-Salgado et al. (2024a) was based on two Azure Kinect depth cameras and was specifically designed for industrial high payload collaborative robots. Most important characteristics of such systems primarily include accuracy, response time, cost and system setup; it should be noted that often improving one of these characteristics can lead to a deterioration of others.

System accuracy directly depends on a number and a type of employed sensors. The single Kinect camera system in Amaya-Mejía et al. (2022) demonstrated detection accuracy varying from 97% in a danger zone to 57.4% in a safe zone; the authors explained this discrepancy by a disparity effect of the Kinect, which reduces an effective field of view. Due to the employed RGB cameras, our system did not suffer from this effect and succeeded providing consistent accuracy across all zones.

As it was noted in Sect. 2.4, human detection based methods carry the risks of false negatives, when a vision system fails detecting an operator despite their actual presence in a danger zone. This is particularly true for human skeleton detection based methods, which were used in Amaya-Mejía et al. (2022) and Forlini et al. (2024). Moreover, these methods significantly increase the system's response time as a number of people in a frame increases. As shown in Amaya-Mejía et al. (2022), the time required for the robot to stop increased on average by 12.5 ms when two people appeared in a scene instead of a single human. At the same time, while image classification methods can also lead to increased processing time as a scene becomes more complex, they are generally less affected by a number of objects compared to methods that focus on detecting specific objects. As a part of our future work, we plan to conduct a thorough analysis of our system's sensitivity to a number of people in a frame.

An important advantage of our method is the relatively short time (52.8 ms in average) required for the robot to adjust its speed when a human enters the zone of increased attention or the danger zone. Moreover, this value includes both the time needed to detect a human in the danger zone and the time required to stop (or slow down) the robot. For a comparison, Amaya-Mejía et al. (2022) reported that the time required for the robot to stop moving after the system sends information about a human detection in a high-risk zone averaged 63.7 ms for one person in the scene. In study Forlini et al. (2024), coordinates of human joints obtained

from the skeleton detection system were sent to the obstacle avoidance algorithm at a frequency of 18 Hz, which hints that the time to send a signal to the robot could be around 55.56 ms, not including the time for processing images from the cameras. In study Katsampiris-Salgado et al. (2024a), it was stated that the estimated time for both communication and prediction was about 65 ms.

From the perspective of setup time, our proposed system may require more time for dataset preparation and model training compared to pre-trained solutions based on MediaPipe (Amaya-Mejía et al., 2022) and Azure Kinect Body Tracking SDK (Katsampiris-Salgado et al., 2024a). However, by utilizing a pre-trained YOLOv8 model and a simple automated annotation process, this time was significantly reduced. Like all machine learning-based methods, our algorithm requires updating the dataset annotation in the case of significant changes in the assembly process technology and/or robotic cell operating conditions. When collecting a dataset, it is essential to consider a variety of environmental factors, such as changes in lighting throughout a day and under different weather conditions, as well as a presence of permissible objects and people in a frame. It is important to note that updating the dataset in response to significant changes in a scene is a necessary requirement for most machine learning methods.

Like any computer vision based method, our system may face a number of typical challenges in real-world industrial settings that could be negotiated by further extensions of the proposed solution; potential extensions are briefly discussed in the next paragraph and provide a fruitful source for our future work. RGB cameras can be sensitive to changes in lighting conditions depending on the time of day, weather conditions, and fluctuations in artificial lighting (Ceccarelli & Secci, 2022; Atif et al., 2023). RGB cameras based methods may suffer from a loss of details in high-contrast scenes. Glare from reflective surfaces or bright light sources could create artifacts that degrade image quality. Vibrations from heavy machinery could cause image blurriness and lead to camera misalignment. Additionally, electronic components of RGB cameras could be sensitive to interference caused by increased electromagnetic radiation (Li et al., 2017; Wu et al., 2019).

To enhance our method reliability in real-world industrial conditions, adaptive image processing algorithms [e.g., (Wang et al., 2024; Shi & He, 2022)], stabilizing devices and polarizing filters [e.g., (Rodriguez et al., 2022b)] could be additionally employed. However, a need for such measures would be determined by specifics of a particular production process and environmental conditions. In our future work, we plan to test the proposed system in typical production environments and develop recommendations that would help mitigating the aforementioned negative factors.

In conclusion to the discussion section, we emphasize several key aspects that set our vision-based safety system apart within the landscape of existing collaborative assembly technologies. First and foremost, we developed an original risk assessment algorithm that utilizes data from two RGB cameras and does not require additional depth sensors. Currently, the overall cost of our system is comparable to that of depth camera based systems due to the employed high-quality RGB cameras; yet, we believe that our approach can be adapted for less expensive cameras, which is a part of our future work. Furthermore, our approach is based on simple image classification methods using the YOLOv8-cls model (Ultralytics, 2024b), which (unlike other methods that rely on human recognition) provides faster data processing. Finally, the proposed vision-based safety system is seamlessly integrated with the contactless robot control system through gestures, which is another unique aspect of the proposed solution.

Conclusions

The study introduces a novel visual control system designed to ensure safe collaborative assembly processes for intelligent manufacturing. Unlike many existing systems, this setup relies solely on data from two RGB cameras without additional depth information from other sensors. The cameras capture top and side projections of the collaborative workspace. The safety system evaluates a current danger level using two YOLOv8-cls neural network models, initially pre-trained on the ImageNet dataset and further fine-tuned on our data. Upon detecting potential hazards, the system triggers decisions to slow down or halt the robot based on the proposed algorithm. To facilitate assembly process management the system is integrated with a visual control system, which is governed by operator gestures.

Our validation experiments revealed good consistency between the danger level assessments in collaborative assembly obtained by the system and those provided by the human experts. These results underline a power of the proposed approach in managing hazardous situations that occur during collaborative assembly and preventing operator injuries. The proposed system stands out for its accessibility for intelligent manufacturing (due to 2D cameras) and could be employed by small and medium-sized enterprises seeking to enhance safety of production processes that involve industrial robots in collaborative assembly.

Acknowledgements This research was funded by the Kazan Federal University Strategic Academic Leadership Program ("PRIORITY-2030").

Data availability The video that was used for the risks evaluation by human experts and by the proposed system is publicly available at https://rutube.ru/plst/655575/. The file containing the evaluaion is publicly available at https://gitlab.com/lirs-kfu/public-data-for-papers.

Any additional data related to the paper are available for academic use on request; please send an email with a particular request to the first author or the corresponding author. Please cite this paper if you use any aforementioned data.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Research involving human participants All procedures involving human participants adhered to the ethical standards established by the Ethics Committee of the Institute of Information Technology and Intelligent Systems, Kazan Federal University, and met the safety requirements of ISO/TS 15066:2016, ISO 10218-1 and ISO 10218-2.

Informed consent All the participants involved in the experiments signed an informed consent form to participate in the study.

References

Amaya-Mejía, L. M., Duque-Suárez, N., Jaramillo-Ramírez, D., & Martinez, C. (2022). Vision-based safety system for barrierless human–robot collaboration. In *2022 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, (pp. 7331–7336). https://doi.org/10.1109/IROS47612.2022.9981689

Atif, M., Ceccarelli, A., Zoppi, T., & Bondavalli, A. (2023). Tolerate failures of the visual camera with robust image classifiers. *IEEE Access, 11*, 5132–5143. https://doi.org/10.1109/ACCESS.2023.3237394

Barari, A., de Sales Guerra Tsuzuki, M., Cohen, Y., & Macchi, M. (2021). Intelligent manufacturing systems towards industry 4.0 era. *Journal of Intelligent Manufacturing, 32*, 1793–1796. https://doi.org/10.1007/s10845-021-01769-0

Bdiwi, M., Naser, I., Halim, J., Bauer, S., Eichler, P., & Ihlenfeldt, S. (2022). Towards safety 4.0: A novel approach for flexible human–robot-interaction based on safety-related dynamic finite-state machine with multilayer operation modes. *Frontiers in Robotics and AI, 9*, 1002226. https://doi.org/10.3389/frobt.2022.1002226

Ceccarelli, A., & Secci, F. (2022). RGB cameras failures and their effects in autonomous driving applications. *IEEE Transactions on Dependable and Secure Computing, 20*(4), 2731–2745. https://doi.org/10.1109/TDSC.2022.3156941

Cencen, A., Verlinden, J. C., & Geraedts, J. M. P. (2018). Design methodology to improve human–robot coproduction in small-and medium-sized enterprises. *IEEE/ASME Transactions on Mechatronics, 23*(3), 1092–1102. https://doi.org/10.1109/TMECH.2018.2839357

Cherubini, A., & Navarro-Alarcon, D. (2021). Sensor-based control for collaborative robots: Fundamentals, challenges, and opportunities. *Frontiers in Neurorobotics, 14*, 576846. https://doi.org/10.3389/fnbot.2020.576846

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848

Faccio, M., Granata, I., Menini, A., Milanese, M., Rossato, C., Bottin, M., et al. (2023). Human factors in cobot era: A review of modern production systems features. *Journal of Intelligent Manufacturing, 34*(1), 85–106. https://doi.org/10.1007/s10845-022-01953-w

Forlini, M., Neri, F., Ciccarelli, M., Palmieri, G., & Callegari, M. (2024). Experimental implementation of skeleton tracking for collision avoidance in collaborative robotics. *The International Journal of Advanced Manufacturing Technology, 134*(1), 57–73. https://doi.org/10.1007/s00170-024-14104-7

Galin, R., & Meshcheryakov, R. (2019). Review on human–robot interaction during collaboration in a shared workspace. In A. Ronzhin, G. Rigoll, & R. Meshcheryakov (Eds.), *Interactive collaborative robotics* (pp. 63–74). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-26118-4_7

Gkournelos, C., Konstantinou, C., Angelakis, P., Tzavara, E., & Makris, S. (2023). Praxis: A framework for AI-driven human action recognition in assembly. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-023-02228-8

Gualtieri, L., Palomba, I., Wehrle, E. J., & Vidoni, R. (2020). *The opportunities and challenges of SME manufacturing automation: Safety and ergonomics in human–robot collaboration*. Springer International Publishing (pp. 105–144). https://doi.org/10.1007/978-3-030-25425-4_4

Hanna, A., Bengtsson, K., Dahl, M., Erős, E., Götvall, P.-L., & Ekström, M. (2019). Industrial challenges when planning and preparing collaborative and intelligent automation systems for final assembly stations. In *2019 24th IEEE international conference on emerging technologies and factory automation (ETFA)* (pp. 400–406). https://doi.org/10.1109/ETFA.2019.8869014

International Organization for Standardization. (2011a). *ISO 10218-1: 2011. Robots and robotic devices—Safety requirements for industrial robots—Part 1: Robots*. International Organization for Standardization.

International Organization for Standardization. (2011b). *ISO 10218-2: 2011. Robots and robotic devices—Safety requirements for industrial robots—Part 2: Robot systems and integration*. International Organization for Standardization.

International Organization for Standardization. (2016). *ISO/TS 15066: 2016. International Organization for Standardization: Robots and robotic devices—Collaborative robots*.

Kamezaki, M., Wada, T., & Sugano, S. (2024). Dynamic collaborative workspace based on human interference estimation for safe and productive human–robot collaboration. *IEEE Robotics and Automation Letters, 9*(7), 6568–6575. https://doi.org/10.1109/LRA.2024.3405352

Kanazawa, A., Kinugawa, J., & Kosuge, K. (2021). Motion planning for human–robot collaboration using an objective-switching strategy. *IEEE Transactions on Human-Machine Systems, 51*(6), 590–600. https://doi.org/10.1109/THMS.2021.3112953

Karagiannis, P., Kousi, N., Michalos, G., Dimoulas, K., Mparis, K., Dimosthenopoulos, D., Tokçalar, Ö., Guasch, T., Gerio, G. P., & Makris, S. (2022). Adaptive speed and separation monitoring based on switching of safety zones for effective human–robot collaboration. *Robotics and Computer-Integrated Manufacturing, 77*, 102361. https://doi.org/10.1016/j.rcim.2022.102361

Katsampiris-Salgado, K., Dimitropoulos, N., Gkrizis, C., Michalos, G., & Makris, S. (2024a). Advancing human-robot collaboration: Predicting operator trajectories through AI and infrared imaging. *Journal of Manufacturing Systems, 74*, 980–994. https://doi.org/10.1016/j.jmsy.2024.05.015

Katsampiris-Salgado, K., Haninger, K., Gkrizis, C., Dimitropoulos, N., Krüger, J., Michalos, G., & Makris, S. (2024b). Collision detection for collaborative assembly operations on high-payload robots. *Robotics and Computer-Integrated Manufacturing, 87*, 102708. https://doi.org/10.1016/j.rcim.2023.102708

Keshvarparast, A., Battini, D., Battaia, O., & Pirayesh, A. (2024). Collaborative robots in manufacturing and assembly systems: Literature review and future research agenda. *Journal of Intelligent Manufacturing, 35*(5), 2065–2118. https://doi.org/10.1007/s10845-023-02137-w

Kozamernik, N., Zaletelj, J., Kosir, A., Šuligoj, F., & Bracun, D. (2023). Visual quality and safety monitoring system for human–robot cooperation. *The International Journal of Advanced Manufac-*

turing Technology, 128(1–2), 685–701. https://doi.org/10.1007/s00170-023-11698-2

Li, X., Wu, P., Meng, C., Liu, Y., & Jin, H. (2017). Experimental study on probability threshold of electromagnetic effect of electronic equipment. In *2017 Asia–Pacific international symposium on electromagnetic compatibility (APEMC)*, Seoul, Korea (South) (pp. 347–349). https://doi.org/10.1109/APEMC.2017.7975502

Li, W., Hu, Y., Zhou, Y., & Pham, D. (2023). Safe human–robot collaboration for industrial settings: A survey. *Journal of Intelligent Manufacturing, 35*(5), 2235–2261. https://doi.org/10.1007/s10845-023-02159-4

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., et al. (2019). Mediapipe: A framework for perceiving and processing reality. In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*.

Makris, S., Michalos, G., Dimitropoulos, N., Krueger, J., & Haninger, K. (2024). Seamless human–robot collaboration in industrial applications. In *CIRP novel topics in production engineering: Volume 1. Lecture notes in mechanical engineering* (pp. 39–73). https://doi.org/10.1007/978-3-031-54034-9_2

Malm, T., Salmi, T., Marstio, I., & Montonen, J. (2019). Dynamic safety system for collaboration of operators and industrial robots. *Open Engineering, 9*(1), 61–71. https://doi.org/10.1515/eng-2019-0011

Maric, B., Jurican, F., Orsag, M., & Kovacic, Z. (2021). Vision based collision detection for a safe collaborative industrial manipulator. In *2021 IEEE international conference on intelligence and safety for robotics (ISR)* (pp. 334–337). https://doi.org/10.1109/ISR50024.2021.9419493

Mustafin, M., Chebotareva, E., Li, H., & Magid, E. (2023a). Experimental validation of an interface for a human–robot interaction within a collaborative task. In A. Ronzhin, A. Sadigov, & R. Meshcheryakov (Eds.), *In interactive collaborative robotics* (pp. 23–35). Bern: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43111-1_3

Mustafin, M., Chebotareva, E., Li, H., Martínez-Garcia, E. A., & Magid, E. (2023b). Features of interaction between a human and a gestures-controlled collaborative robot in an assembly task: Pilot experiments. In *Proceedings of international conference on artificial life and robotics (ALife robotics)* (pp. 162–165). https://doi.org/10.5954/ICAROB.2023.OS6-4

Petzoldt, C., Harms, M., & Freitag, M. (2023). Review of task allocation for human–robot collaboration in assembly. *International Journal of Computer Integrated Manufacturing, 36*(11), 1675–1715. https://doi.org/10.1080/0951192X.2023.2204467

Proia, S., Carli, R., Cavone, G., & Dotoli, M. (2022). Control techniques for safe, ergonomic, and efficient human–robot collaboration in the digital industry: A survey. *IEEE Transactions on Automation Science and Engineering, 19*(3), 1798–1819. https://doi.org/10.1109/TASE.2021.3131011

Rodrigues, I. R., Barbosa, G., de Oliveira Filho, A. T., Cani, C., Dantas, M., Sadok, D. H., et al. (2022a). Modeling and assessing an intelligent system for safety in human–robot collaboration using deep and machine learning techniques. *Multimedia Tools and Applications, 81*, 2213–2239. https://doi.org/10.1007/s11042-021-11643-z

Rodrigues, I. R., Dantas, M., de Oliveira Filho, A. T., Barbosa, G., Bezerra, D., Souza, R., et al. (2023). A framework for robotic arm pose estimation and movement prediction based on deep and extreme learning models. *The Journal of Supercomputing, 79*(7), 7176–7205. https://doi.org/10.1007/s11227-022-04936-z

Rodriguez, J., Lew-Yan-Voon, L., Martins, R., & Morel, O. (2022b). A practical calibration method for RGB micro-grid polarimetric cameras. *IEEE Robotics and Automation Letters, 7*(4), 9921–9928. https://doi.org/10.1109/LRA.2022.3192655

Saleem, Z., Gustafsson, F., Furey, E., McAfee, M., & Huq, S. (2024). A review of external sensors for human detection in a human–robot collaborative environment. *The International Journal of Advanced Manufacturing Technology, 134*, 1–17. https://doi.org/10.1007/s00170-024-14104-7

Scimmi, L. S., Melchiorre, M., Mauro, S., & Pastorelli, S. P. (2019). Implementing a vision-based collision avoidance algorithm on a UR3 robot. In *23rd International conference on mechatronics technology (ICMT)* (pp. 1–6). https://doi.org/10.1109/ICMECT.2019.8932105

Selvaraj, S. B., Canale, R., Piriyatharawet, T., Xiao, R., Vyas, P., & Horng, C. S. (2023). Towards safe and efficient human-robot collaboration: Motion planning design in handling dynamic obstacles. In *IECON 2023-49th annual conference of the IEEE industrial electronics society*, Singapore, Singapore (pp. 1–5). https://doi.org/10.1109/IECON51785.2023.10311798

Shi, L. & He, H. (2022). A review and comparison on video stabilization algorithms. In *2022 5th world conference on mechanical engineering and intelligent manufacturing (WCMEIM)*, Ma'anshan, China (pp. 1093–1097). https://doi.org/10.1109/WCMEIM56910.2022.10021453

Terven, J. R., & Cordova-Esparza, D. M. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction, 5*(4), 1680–1716. https://doi.org/10.3390/make5040083

Ultralytics, Inc. (2024a). Pose—Ultralytics YOLOv8 Docs. https://docs.ultralytics.com/ru/tasks/pose/

Ultralytics, Inc. (2024b). Classify—Ultralytics YOLOv8 Docs. https://docs.ultralytics.com/tasks/classify/

Vieira, S. M., Kaymak, U., & Sousa, J. M. (2010). Cohen's kappa coefficient as a performance measure for feature selection. In *International conference on fuzzy systems* (pp. 1–8). https://doi.org/10.1109/FUZZY.2010.5584447

Wang, K.-J., Lin, C., Tadesse, A., & Woldegiorgis, B. (2023). Modeling of human–robot collaboration for flexible assembly—A hidden semi-Markov-based simulation approach. *The International Journal of Advanced Manufacturing Technology, 126*(11), 5371–5389. https://doi.org/10.1007/s00170-023-11404-2

Wang, Y., Huang, Q., Liu, J., Jiang, C., & Shang, M. (2024). Adaptive video stabilization based on feature point detection and full-reference stability assessment. *Multimedia Tools and Applications, 83*(11), 32497–32524. https://doi.org/10.1007/s11042-023-16607-z

Wong, C. Y., Vergez, L., & Suleiman, W. (2024). Vision-and tactile-based continuous multimodal intention and attention recognition for safer physical human–robot interaction. *IEEE Transactions on Automation Science and Engineering, 21*(3), 3205–3215. https://doi.org/10.1109/TASE.2023.3276856

Wu, P., Xu, Z., Meng, C., Wen, L., & Guo, Q. (2019). The experiment study of effects on ADC chip against radiation and electromagnetic environment. In *2019 12th international workshop on the electromagnetic compatibility of integrated circuits (EMC Compo)*, Hangzhou, China (pp. 207–209). https://doi.org/10.1109/EMCCompo.2019.8919802

Zhang, S., Li, S., Li, X., Xiong, Y., & Xie, Z. (2022). A human–robot dynamic fusion safety algorithm for collaborative operations of cobots. *Journal of Intelligent Robotic Systems, 104*(1), 18. https://doi.org/10.1007/s10846-021-01534-8