Uncompromising Operator Safety: A Standalone Device Approach for Threat Immunity and Malfunction Prevention through Visual Cognition

Mihai Penica, Eoin O'Connell, Reenu Mohandas, William O'Brien, Martin Hayes
University of Limerick

Abstract

The emergence of Industry 4.0 has generated significant attention regarding the mitigation of security breaches and ensuring operator safety within industrial settings. This scholarly article addresses the concept of cognitive interaction within Industry 4.0 frameworks and elucidates the pivotal role of a stand-alone system deployed in industrial environments. The system in question leverages an edge standalone architecture, incorporating a visual cognition AI model to proactively safeguard operators working in conjunction with robots. Moreover, this article delineates an end-to-end integration methodology encompassing visual detection, hardware integration, and the overarching objective of safeguarding operators in the event of a system compromise.

Keywords: Machine Vision, Interoperability,Iot, Cybersecurity, Industry 4.0

1 Introduction

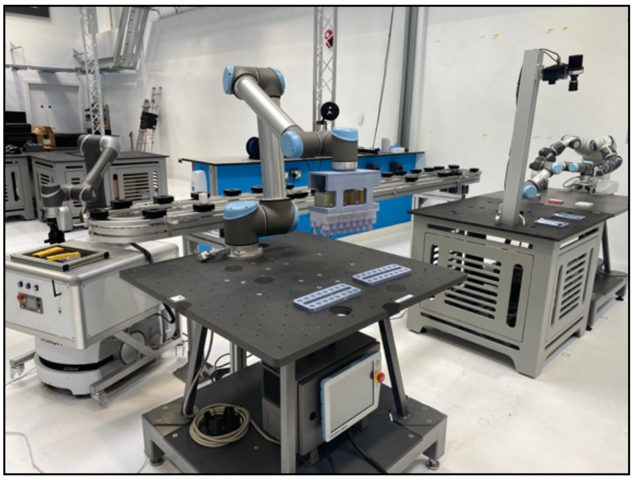

Figure 1: Equipment used in the experiment

The oversight of security considerations within the field of robotics frequently results in a state of heightened vulnerability for robots. This prevailing trend can be attributed to a multitude of factors. Primarily, the extant defensive security mechanisms designed for robots are still in the early stages of their development, lacking comprehensive coverage of the complete spectrum of potential threats. Consequently, these mechanisms prove to be inadequate in effectively safeguarding robots against malicious attacks. The intrinsic complexity inherent in robotic systems presents substantial challenges when attempting to implement robust security measures. The intricate nature of these systems exacerbates both the technical and economic costs associated with ensuring their protection. As a consequence, manufacturers frequently adopt suboptimal security practices due to the considerable expenses and technical intricacies entailed in fortifying these advanced systems [Mayoral-Vilches, 2022].Against this backdrop, this research paper introduces an autonomous system that leverages cost-effective components to mitigate the risks associated with robot compromise and potential harm resulting from malfunctions. The proposed system operates as a stand-alone entity shielded from external attacks while concurrently safeguarding the operator from potential dangers in scenarios involving robot hacking or malfunctions. Through the incorporation of visual cognition models, the system strives to ensure the operator's well-being and prevent injuries arising from compromised robot operations. This paper advocates for the implementation of an autonomous system fortified with visual cognition models, underscoring the importance of addressing security concerns and prioritizing the safety of operators engaged in robot-related scenarios. Figure 1 showcases a selection of robots on which our system has undergone testing and evaluation. The development of the proposed autonomous system aims to facilitate several key attributes, highlighting the importance of addressing security concerns and prioritizing the safety of operators engaged in robot-related scenarios. These attributes include:

- **Enhanced Security**: The system emphasizes the implementation of robust security measures, leveraging visual cognition models to detect potential threats and safeguard against likelihood of injuries or accidents.
- **Autonomous Functionality**: The system is designed to operate autonomously, reducing the reliance on manual intervention, and minimizing the potential for human error. This autonomous functionality enhances the efficiency and effectiveness of robot operations, while also mitigating the risks associated with operator involvement.
- **Real-time Threat Detection**: Through the integration of visual cognition models, the system is capable of real-time threat detection and response. By continuously monitoring the environment and identifying potential risks, the system can promptly react on anomalous activities, ensuring timely and appropriate actions are taken.
- **Operator Safety**: The system places a strong emphasis on operator safety, aiming to protect individuals from potential harm arising from robot compromise or malfunctions. By implementing proactive security measures and prompt response mechanisms, the system acts as a safety net, reducing the likelihood of injuries or accidents during robot operations.

2 Literature review

Cyber Physical systems in the Engineering perspective can be explained as the monitoring and dynamic control of physical components of any environment using sensor data and actuators that are part of a distributed computing system [Nunes, 2015]. These Cyber Physical Systems involves controlled interactions among robotics, Internet of things, Wireless sensor networks, Edge Processing and Cloud computing technologies. Humans have been an essential part of these cyber physical systems especially in the central control, but for any cyber physical system, human is an external and unpredictable element in the environment. This generated the idea of human-in-the-loop concepts which consider the human input through sensory data available and most importantly safety and security of the human involved in the control loop. Safety of the human in human-in-the-loop environment is the focus of work in this paper.

Current research in the human-in-the-loop front is about the interaction between machines and humans for data collection for interactive machine learning, called human-in-the-loop machine learning [Mosqueira-Rey, 2023]. Hence, ensuring the safety of the human operator at all stages on the factory floor has become ever more important and urgent. With advanced sensors and camera systems, video surveillance has been deployed in various environments, public infrastructures, commercial buildings, manufacturing settings and factory floors. Real-time automated analysis and inference from the image or video sequences generated by these camera systems are enabled through Deep Learning. Different applications of real-time person detection using deep learning has been developed over the recent years, access/egress, people counting, person identification and tracking, and most importantly in threat detection and emergency response[Ahmed I. and Jeon, 2021]. Although it's been widely used, person detection is a challenging task in machine vision, with problem of occlusion when person and object overlap each other. Adding further to the complexity is the diverse human postures, gestures and actions [Bouafia, Y., Guezouli, L. and Lakhlef, 2022]. MobileNet-V1 is the model that is used as a backbone in the detection system, which was proposed by Google researchers in 2017 using Depthwise Separate Convolution making it very efficient and less energy consuming [Bouafia, Y., Guezouli, L. and Lakhlef, 2022]. In 2019, MobileNet-V2 was announced with Bottleneck Residual Block instead of Depthwise Separable Convolution blocks.

The model is trained using INRIA person dataset and transfer learning has been used to combat the data-hungry nature of deep learning model when training from scratch. Transfer learning are of two types: Transfer learning via feature extraction and Transfer Learning via fine-tuning. In the former case, the pre-trained network is used to extract features which could then be further used for classification or detailed data parameterization. In the latter case, the pre-trained network can be used fully by adding further layers and retraining the newly connected layers so that the currently stored weights could also be used to achieve good results.

Transfer learning is the best solution when the available dataset is small, compared to 14 million images in more than 21k categories in the ImageNet[J. Deng, 2009] dataset.

3 Proposed Solution

The methodology employed in this study adopts a qualitative research design, with a specific focus on qualitative content analysis. The primary objective of this research is to address real-world challenges related to operator safety in robot-related scenarios, particularly through the utilization of low-cost devices that are impervious to hacking. By emphasizing practicality and application, the study aims to identify and overcome the identified challenges in the field. The approach employed in this study involves practical experimentation and iterative development, allowing for the generation of tangible outcomes with practical applications that can be implemented in real-world settings. Through these iterative processes, the study aims to refine and enhance the proposed autonomous system, ensuring its effectiveness and reliability when addressing security concerns and prioritizing operator safety. By focusing on these key attributes, the system aims to mitigate potential risks and optimize the performance of robots in various operational contexts. By adopting a qualitative research methodology, this study seeks to provide a comprehensive understanding of the challenges and requirements associated with the development of the proposed autonomous system. Practical experimentation and iterative development serve as means to generate tangible outcomes that can be practically applied, thereby making notable contributions to the field of robotics and effectively addressing the identified challenges.

To effectively tackle the challenges associated with safety in robotic and industrial machinery operations, the proposed solution integrates AI-trained models and camera systems to establish a robust and real-time virtual safety fence. By implementing the proposed approach, the system establishes a virtual boundary encompassing the designated robot work area. In the event of a breach where the operator crosses this virtual boundary, the system triggers the fail-safe switch promptly and automatically. This instantaneous response ensures the immediate termination of potentially hazardous operations, effectively safeguarding the operators involved. Figure 2 visually illustrates the pivotal role played by the designed system in ensuring the safety of the surrounding environment and operators during robotic operations.

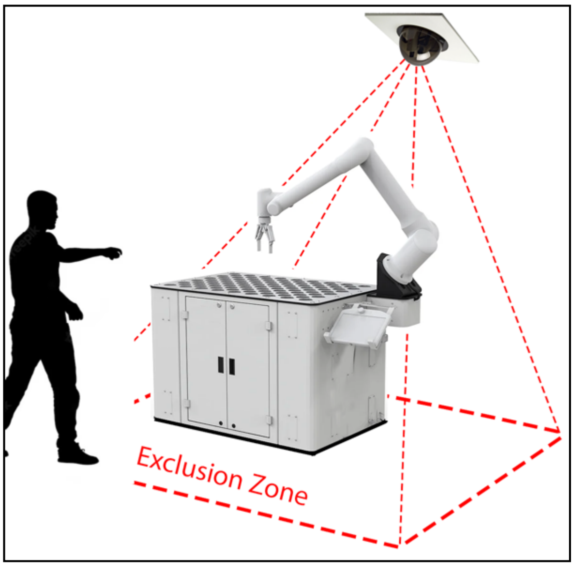

Figure 2 Virtual boundary

Because the system operates independently of any network connectivity, making it impervious to cyberattacks. The fail-safe switch, being a physical switch, cannot be overridden by any software intervention. This characteristic enhances the system's resilience and security, significantly reducing the risk of cyberattacks that could potentially harm the operators. The combination of an independent operation and a physical fail-safe switch enhances the system's suitability in preventing safety risks associated with cyber threats.

In summary, the proposed system creates a virtual boundary to delineate the robot work area, activating the fail-safe switch upon breaching this boundary. Figure 2 visually exemplifies the system's role in ensuring operator safety. Moreover, the system's independent operation and utilization of a physical fail-safe switch contribute to its robustness against cyberattacks, ultimately mitigating potential harm to the operators.

Figure 3 provides a visual representation of the sequential steps involved in the development and construction of the system proposed in this paper. The depicted process encompasses several key stages, including the design of the hardware solution, the implementation and testing of the hardware components, the development of the software, the integration of the software and hardware components, the design and training of deep learning networks, and the final stage of integrating and testing the software and hardware components together.

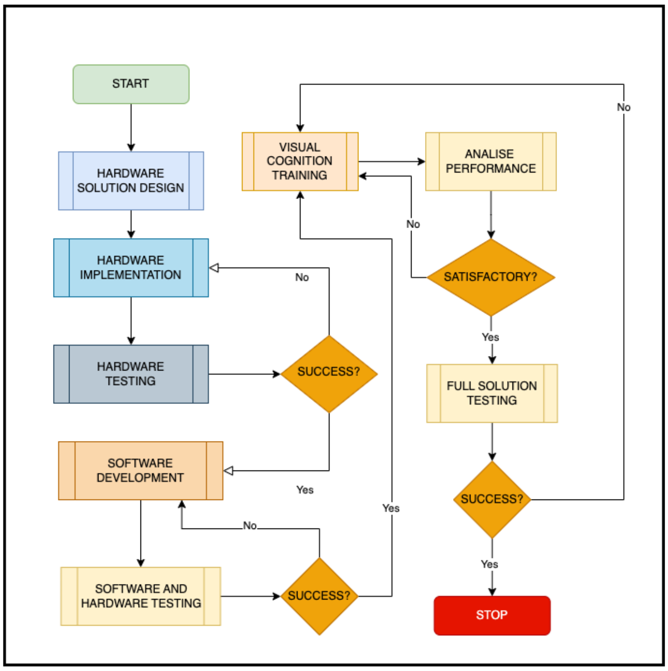

Figure 3 Solution development

The first stage, labelled as the hardware solution design, involves the conceptualization and design of the hardware components that constitute the system. This stage encompasses determining the necessary hardware elements and their specifications to meet the objectives of the proposed system. The hardware solution undergoes implementation and testing to ensure its functionality and compatibility with the intended objectives. Rigorous testing and evaluation are conducted to verify the performance and reliability of the hardware components. Parallel to the hardware development, the software components of the system are created. This entails the development of software modules and algorithms that enable the desired functionality of the system. The software development process includes coding, debugging, and optimization to ensure efficient and effective operation. Once the software and hardware components are individually developed, the integration process takes place. This phase involves combining the software and hardware elements to create a unified system. The integration process ensures seamless communication and coordination between the software and hardware components. The design and training of deep learning networks constitute a critical stage in the development process. Deep learning models are designed and trained to enable the system to perform complex tasks such as visual recognition or decision-making. This stage involves data collection, pre-processing, model design, training, and evaluation to achieve optimal performance.

Finally, the integrated software and hardware components undergo comprehensive testing to validate the functionality, performance, and reliability of the system as a whole. This testing phase ensures that the system operates as intended and meets the specified requirements.

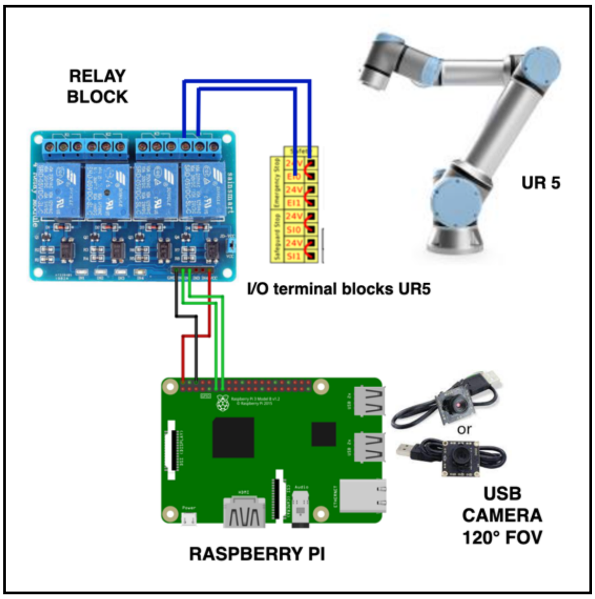

Figure 4 Hardware components

3.1 Hardware design

The hardware employed in this prototype encompasses the utilization of a Raspberry Pi 4 Model B/4G, renowned for its advanced capabilities. This device is equipped with an ARM Cortex-A72 CPU, which provides ample computational power necessary to sustain the execution of a dedicated mode within the system. This mode is specifically designed to facilitate deep learning detection, enabling the system to discern and identify various visual patterns or cues. Moreover, this powerful CPU facilitates the execution of subsequent actions required to activate safety procedures, ensuring the well-being and protection of operators in robot-related scenarios.

The Raspberry Pi 4 Model B/4G is purposefully employed in an autonomous manner, operating independently and without reliance on any external network connections. This autonomous functionality serves a crucial role in the system, allowing it to operate seamlessly and self-contained, irrespective of any external network dependencies. By leveraging the processing capabilities of the Raspberry Pi, the system can effectively execute visual cognition models, which are fundamental in detecting the presence of operators and evaluating potential risks associated with their interactions with the robot. The system is able to promptly trigger the robot's kill switch mechanism, ensuring a swift and appropriate response to mitigate any potential dangers or hazardous situations. The hardware connections and configurations are visually represented in Figure 3. This figure provides a comprehensive illustration of the interconnections and arrangements of the hardware components within the system.

3.2 Software design

The software utilized in this prototype, which runs on the Raspberry Pi, was developed using Python, a prominent object-oriented high-level programming language. Python is renowned for its emphasis on code readability and simplicity, which has contributed to its extensive usage across diverse domains such as web development, data analysis, scientific computing, and artificial intelligence, among others. Python's support for object-oriented programming principles empowers developers to create and define objects that encapsulate both data and behaviour. This approach facilitates the construction of modular and reusable code, enhancing the overall organization and structure of software applications. By leveraging Python's object-oriented capabilities, developers can design software components that accurately model real-world entities and interactions, resulting in more efficient and maintainable codebases. The utilization of Python in this prototype ensures that the software implementation is coherent, understandable, and facilitates the integration of various functionalities required for the successful operation of the system. By harnessing Python's versatility and object-oriented paradigm, the software developed for the Raspberry Pi effectively accomplishes the intended goals and seamlessly interacts with the hardware components[Penica M., 2021]. To facilitate control over the relay responsible for triggering the robot's kill switch, the RPi.GPIO library was incorporated into the software. This library provides convenient and straightforward control over the relay, enabling the system to effectively manage the relay's activation. By utilizing the RPi.GPIO library, the software gains the ability to manipulate the relay with ease, ensuring precise control and timely response when necessary. The integration of the RPi.GPIO library into the software streamlines the process of controlling the relay, enhancing the overall functionality and reliability of the system. With this library's assistance, the system achieves efficient and precise triggering of the relay, thus enabling swift and accurate activation of the robot's kill switch mechanism as required for operator safety and protection.

3.3 Visual cognition

The person detection is achieved using TensorFlow object detection model with an artificial Neural Network as the backbone and Single Shot Detection as the detection method. The deep learning model is trained using publicly available benchmark dataset for person detection, the INRIA persons dataset [Dalal and Triggs 2005]. Transfer Learning is used to achieve higher accuracy in the detection, the model pretrained on ImageNet has been downloaded from the TensorFlow repository and fine-tuned for detection process. GPU GeForce RTX 2080 Ti is used to complete the model training and the TensorFlow-Lite graph is exported onto the Raspberry Pi.

Edge processing is vital in this application since the detection system will be component of(mounted on) a robotic arm. Added to that, the detection is real-time from the Raspberry Pi. SSD-MobileNetV2 is the model used in this prototype, achieving consistently high confidence score of detections, > 95% within detection time of <1ms, in all the scenarios tested. The SSD-MobileNetV2 framework is specifically designed for mobile and Edge applications, with Depthwise separable convolution modules making detections faster thereby reducing reaction time after the virtual fence has been breached.

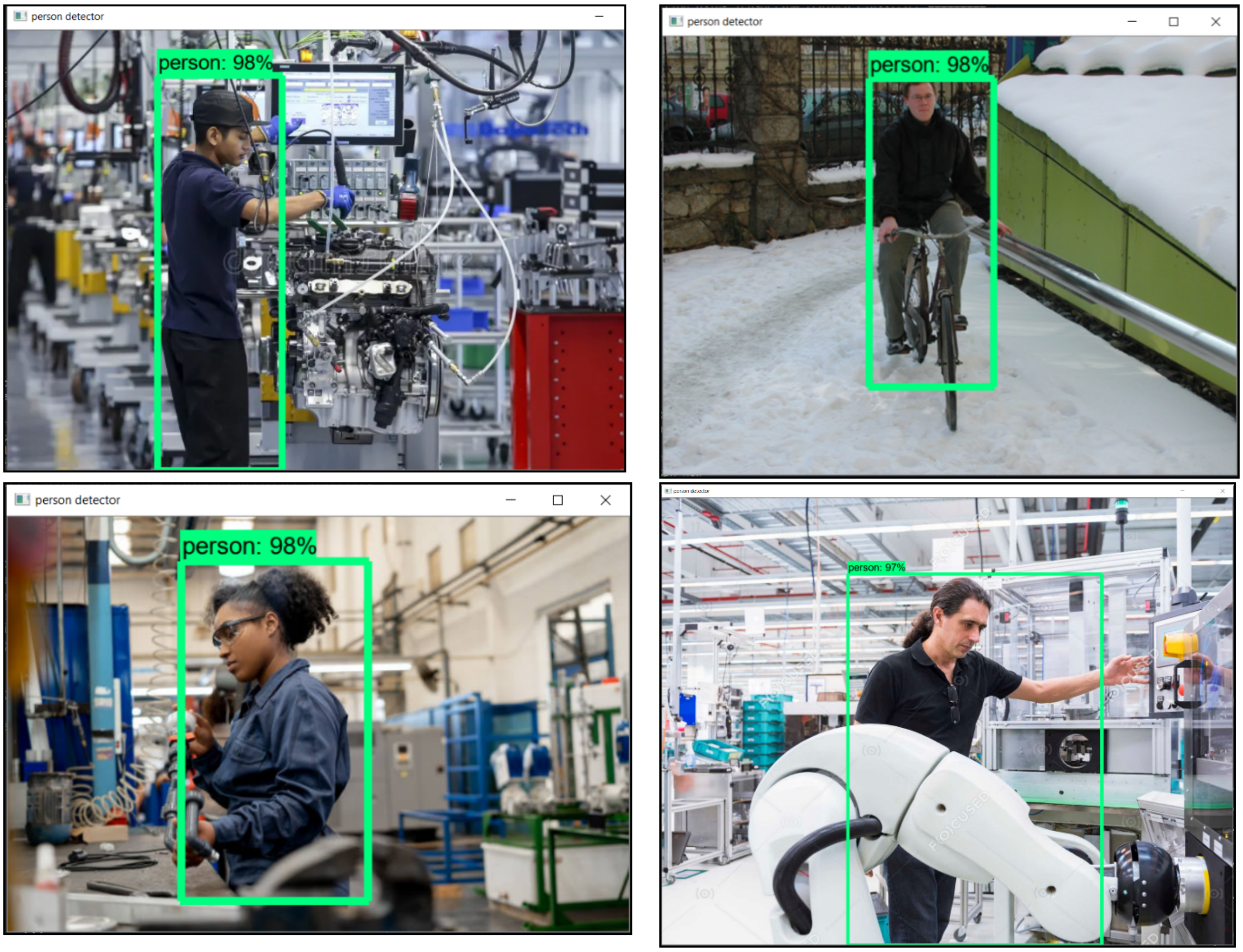

Figure 5: The above images shows high accuracy in person detection in factory settings and one of them in a random environment.

4 Results

In order to evaluate the performance of the system in detecting individuals and ensuring operator safety, a series of tests were conducted using the developed prototype and following the methodologies outlined in this paper. The tests involved ten volunteers with diverse appearances, who participated in ten different scenarios while wearing various accessories such as glasses. The objective of these tests was twofold: first, to determine the distance at which the system would effectively detect the presence of a person, and second, to measure the response time required to ensure the operator would be in a safe environment in the event of a malfunction or intentional hack aimed at causing harm to the operator. During the testing scenario, volunteers approached the robot while wearing different accessories and performed various actions. The system's performance was evaluated based on its ability to accurately detect the presence of the volunteers within a predefined distance and the timeliness of its response in ensuring operator safety. The conducted tests yielded valuable insights regarding the system's performance in real-world scenarios. Based on the collected data, optimal distances for detection and virtual fence placement were determined. For instance, it was found that the optimal distance for person detection was 3 meters away. Additionally, the optimal distance for placing the virtual fence varied depending on the type of robot used in the experiment.

Table 1 below presents the minimum distances and the reaction time required for the robots to come to a stop based on the specific robot types used, ensuring the operator is in a safe environment:

Robot Type	Minimum Safe Distance	Reaction time after the virtual fence has been breached(ms)
UR 3	43 cm	0.321ms
UR5	83.5 cm	0.191 ms
UR 10	122 cm	0.201 ms

Table 1: Minimum safe distance

Table 2 below presents the evaluation results for the person detection module used in this prototype. With finetuning only with INRIA dataset, the model has been able to achieve an IOU value > 80%. Intersection Over Union (IOU) value is used to determine the prediction accuracy of a model, it is the ratio of area of intersection of the predicted bounding box vs area of intersection of the ground truth bounding box. Precision value of a model represents the ratio of True Positives to All Positive Detections. Higher value of Precision value shows that the detections are very accurate and near negligible false detections. Recall is another metric to measure the number of false detections, it's the ratio of True positives against all ground truth instances, means, the number of objects detected from all the instances present in the scene. Higher value of recall indicates the system is highly sensitive to the trained object and detects all instances in range.

Detection Time(ms)	**IOU value**	**Precision**	**Recall**
0.813	0.8134	1	1

Table 2: Evaluation Metrics for Operator Detection

These findings provide valuable guidance for implementing the system in practical applications, aiding in the effective detection of individuals and the establishment of appropriate safety measures based on the specific robot being used. In summary, the tests conducted using the developed prototype and following the methodologies outlined in this paper provided insights into the system's performance in detecting individuals and ensuring operator safety. Optimal distances for person detection and virtual fence placement were determined, taking into account the type of robot utilized in the experiments. These findings contribute to the successful implementation of the system in real-world scenarios, promoting enhanced safety for operators engaged in robotic operations.

5 Conclusions

This paper provides compelling evidence for the practicality of constructing an affordable device capable of running deep learning models at the edge. By proficiently detecting operators, the device effectively safeguards against malfunctions and external malevolent attacks, ensuring operator safety. The detection capability of deep learning model could be further improved by adding more training instances from manufacturing settings which presents the model with more variable instances of human operators working with robots. The research findings underscore the efficacy of employing cost-effective devices to enforce operator safety. The proposed approach emphasizes the device's standalone nature, offering an independent solution that is impervious to hacking or tampering. The study demonstrates that the utilization of a low-power ARM device yields optimal error rates and cognition model performance, enabling fast decision-making capabilities. Furthermore, the research validates the feasibility of implementing lightweight models on embedded devices. In conclusion, this paper conclusively establishes the feasibility of developing a cost-effective device capable of executing deep learning models at the edge. Through precise operator detection, the device effectively guarantees operator safety, even in the presence of malfunctions or external threats. The study affirms the independent operation of the proposed approach and showcases the exceptional performance of the cognition model when utilizing a low-power ARM device. Additionally, the successful implementation of lightweight models on embedded devices further strengthens the paper's contributions to the field.

Acknowledgements

The authors would like to thank the staff of the Electronic and Computer Engineering Department at the University of Limerick for their assistance. The authors would also like to acknowledge resources made available from Science Foundation Ireland through the grant award (16/RC/3918) to Confirm Centre for Smart Manufacturing.

References

[Mayoral-Vilches, 2022] Mayoral-Vilches, V., 2022. Robot hacking manual (rhm). *arXiv preprint arXiv:2203.04765*.

[Nunes, 2015] Nunes, D.S., Zhang, P. and Silva, J.S., 2015. A survey on human-in-the-loop applications towards an internet of all. IEEE Communications Surveys & Tutorials, 17(2), pp.944-965.

[Mosqueira-Rey, 2023] Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J. and Fernández-Leal, Á., 2023. Human-in-the-loop machine learning: A state of the art. Artificial Intelligence Review, 56(4), pp.3005-3054.

[Ahmed I. and Jeon, 2021] Ahmed, I. and Jeon, G., 2021. A real-time person tracking system based on SiamMask network for intelligent video surveillance. Journal of Real-Time Image Processing, 18, pp.1803-1814.

[Bouafia, Y., Guezouli, L. and Lakhlef, 2022] Bouafia, Y., Guezouli, L. and Lakhlef, H., 2022. Human Detection in Surveillance Videos Based on Fine-Tuned MobileNetV2 for Effective Human Classification. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 46(4), pp.971-988.

[J. Deng, 2009] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[Penica M., 2021] Penica, M., Mohandas, R., Bhattacharya, M., Vancamp, K., Hayes, M. and O'Connell, E., 2021, June. A Covid-19 viral transmission prevention system for embedded devices utilising deep learning. In *2021 32nd Irish Signals and Systems Conference (ISSC)* (pp. 1-8). IEEE.

[Dalal and Triggs, 2005] Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. *Comput Vis Pattern Recogn CVPR 2005* IEEE 1:886–893