Detecting dangerous behaviours and promoting safety in manufacturing using computer vision

S. Colabianchi [a)], M. Bernabei [b)] and F. Costantino [c)]

a. Department of Computer, Control and Management Engineering, University of Rome "La Sapienza", Via Ariosto 25, 00185 Rome – Italy (silvia.colabianchi@uniroma1.it)

b. Department of Mechanical and Aerospace Engineering, University of Rome "La Sapienza", Via Eudossiana 18, 00184 Rome – Italy (margherita.bernabei@uniroma1.it)

c. Department of Computer, Control and Management Engineering, University of Rome "La Sapienza", Via Ariosto 25, 00185 Rome – Italy (francesco.costantino@uniroma1.it)

Abstract: The safe operation of forklifts in manufacturing environments is critical for the efficient transportation of goods. However, accidents can occur due to distraction, failure to use Personal Protective Equipment (PPE), and improper handling of the forklift. To improve safety, the authors propose a computer vision solution to monitor forklift operators and their compliance with safety regulations. The model is trained to detect behaviours that could lead to accidents and alert the operator in real time. The proposed solution can be integrated with the forklift's control system, providing immediate feedback to the operator to reduce the risk of accidents. The model uses transfer learning, a technique that leverages pre-trained models to improve the accuracy of the model with limited data. The PoseNet pre-trained model was fine-tuned on a dataset of annotated videos of forklift operators to improve its accuracy in classifying different behaviours. Future work can investigate the integration of the solution with other safety systems to provide a comprehensive safety solution in manufacturing environments.

Keywords: Object Detection Model; Artificial Intelligence; Warehouse; Safety; Digitalization

I. INTRODUCTION

These In warehouses, factories, and other workplaces, forklifts lift, stack, and transfer loads. Forklifts are a practical material-handling solution for many companies. However, they are associated with workplace accidents and injuries [1], generating significant human and financial costs for employees, industry, and society. To ensure safe forklift operations, the Occupational Safety and Health Administration provides guidelines [2]. Drivers must observe those concerning behaviours inside the vehicle and toward their surroundings. After all, the forklift driver does not always perform this regulation during its use. Enabling technologies represent a major opportunity in this regard, potentially increasing the health and safety of the operators working with them, reducing risks, and supporting operators' activities [3]. Many 4.0 technologies have been consolidated in this field, with different objectives and with different levels of operator involvement [4]. Among the different applications, there are those for the safe operation of vehicles, including the computer vision-based ones. In manufacturing, computer vision solutions have been developed and applied to forklifts only to prevent collisions with external objects or people [5,6]. Dangerous behaviour of drivers inside vehicles remains unaddressed, as is regularly covered in other industries, such as automotive [7–10]. Computer vision algorithms for monitoring driver behaviour through body detection and analysis outperform existing commercial automotive-grade devices [11]. For the abovementioned reasons, the following research presents a computer vision solution that leverages the use of internal cameras to infer information regarding the driver's behaviours and status, promoting the safe use of forklifts in manufacturing.

II. RELATED WORKS

The evolution of artificial intelligence technologies leads to the development of the so-called Advanced Driver Assistance Systems (ADAS), vehicle-based intelligent safety systems to support the driver, improving his safety and driving safety in general [11]. These solutions can rely on computer vision, to control information about the vehicle movements, the objects or entities detected in the same environment, and the driver's

behaviours [11,12]. Both internal and external vehicle events can cause accidents or collisions. Related works refer to the development of Computer vision solutions in automotive to monitor the behaviours of a vehicle driver and the surrounding environment, as well as to ensure the safe use of forklifts in industrial settings. In the manufacturing context, computer vision solutions emerge only to monitor external obstacles and environmental settings, not to supervise driver behaviours and status.

A. Computer vision solution in automotive

External obstacles and environmental settings

In [7] the authors present an intelligent driver assistance system based on computer vision. It is developed to prevent accidents with early detection of harmful objects signboards and road lines. An overview of the various pre-processing and detection techniques for creating effective vehicle-pedestrian collision avoidance systems by computer vision is provided in [13]. Here, also dynamic scenes of different environmental conditions, such as daytime, night-time, weather conditions, shadows, and anomaly detection, are considered. Pedestrian detection is addressed also by [14]. A deep-learning based framework is developed by [15] to detect vehicle and pedestrian in rural roads. In [16] several techniques concerning computer vision for detecting and recognising traffic road signs are discussed. The latter is focused on object detection to reach high precision in animated video or still image processing.

Driver behaviours and status

Within this domain, several solutions are developed to identify whether the driver is using a cell phone [17–19]. Others, to monitor multiple behaviors together. For instance, general driver distraction is assessed in [15] by monitoring cell phone usage and the hands on steering wheel. In [10] a computer vision solution is presented to recognize the incorrect fastening of seat belts in passengers and drivers. Moreover, a deep-learning-based self-calibration strategy for the vehicular camera, and its practical application in the automotive domain, is presented in [9]. Also, drowsiness detection system solutions which indicate the status of the car driver using computer vision technology emerge [7,8]. While most computer vision systems to detect drivers' dangerous behaviors exploit in-vehicle cameras, in [18] a solution based on traffic monitoring cameras on the road is proposed.

B. Computer vision solution in industrial forklifts

External obstacles and environmental settings

Computer vision algorithms emerge to predict impending collisions, by distinguishing collisions with humans and objects [5]. Such diversification allows a different priorities-based warning system. In [6], the authors present a computer vision system to calculate possible collisions between forklifts and people. A multiple-object-tracking solution arises to detect objects or people in industrial warehouses, promoting greater safety in the facility [20]. In [1], the authors address the issue of pedestrian detection for forklift systems considering the complexity arising from environmental conditions and meteorological factors, such as night, rain or cloudiness.

III. TOOL DESIGN CONSIDERATION

This section entails an examination of the technical factors fundamental to the conceptualization of our tool's design, while the detailed design and development of the solution are presented in paragraph 4. Over the years, diverse methodologies have been suggested concerning monitoring driver behaviour, particularly concerning the identification of mobile phone usage while driving. These methodologies can be classified into two categories: non-vision-based and computer vision-based approaches [21]. Notable non-vision-based approaches include the utilization of a car's stereo system, a Bluetooth network, the incorporation of antennas at various locations within a vehicle, or integrated sensors. Nonetheless, our work is grounded on a computer vision-based approach, hence our focus is solely on vision-based techniques.

A. Safety critical driving events

Dangerous behaviours

As previously stated, the involvement of mobile phone usage as a source of distraction while driving has been recognized as a significant element in most recent truck accidents [18]. It is well understood that using a mobile phone while driving greatly impacts a driver's ability to concentrate. However, despite this, drivers still use their hand-held mobile phones while driving regularly. Besides this in our case study, another critical behaviour lies in not keeping both hands on the steering wheel. This behaviour is due to idleness, habit, the need to hold papers related to bills of orders, production details etc.

Critical behaviours

Next, we will focus on the behaviour of operators when using the reverse drive. Reversing is a critical movement within an industrial context due to both the speed and quantity of the operations performed and the poor visibility inside the forklift. Moreover, operators tend to carry out reversing operations in an automatic, repetitive manner without turning their torso and consequently without monitoring the presence of operators or obstacles. The developed system, once reverse is engaged, enters an alert state and monitors whether the operator makes a complete torso twist before starting the manoeuvre.

B. Real-time data and performance

The solution fundamentally relies on the real-time processing of a camera video stream. The camera was positioned at the front of the forklift in a high, central position. This made it possible to film all the operator's movements, from his entry into the cab to his exit, and to be able to identify landmarks of the face, torso, hands and objects such as the mobile phone. Real-time videos might require long processing time and processor availability to capture and analyse each frame.

IV. THE PROPOSED FRAMEWORK

The proposed framework is composed of four models: operator detection mode; mobile phone detection model, hands detection model; pose detection model. All models apply computer vision techniques. Computer vision is a set of techniques that attempt to simulate human vision and understand information from images or videos. In this solution, we will focus on object detection, a branch of computer vision that combines image classification and object recognition by trying to understand what (e.g. operator) and where the objects are in the image (e.g. the operator is in the forklift). For each model, the authors have selected a dataset and a pre-trained model. Subsequently, the model was refined and improved by adapting the bounding boxes, and confidence scores and by training an additional classification model. Figure 1 summarises the framework.

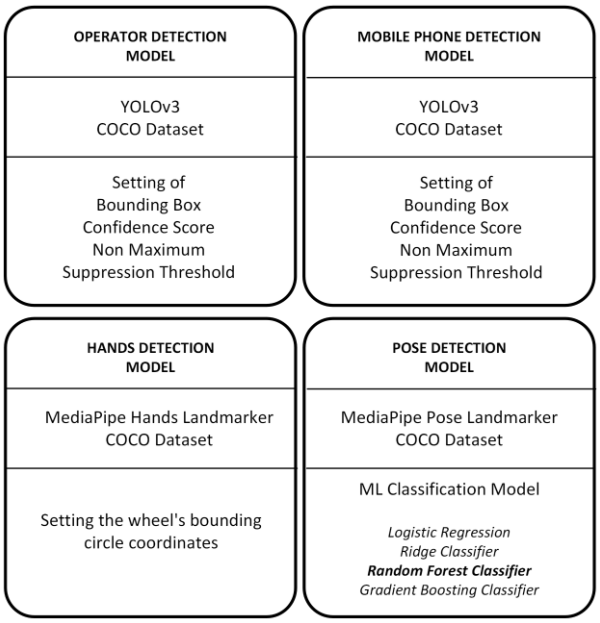


Figure 1. Proposed Framework

A. Operator detection model

In the literature, there are several Machine Learning (ML) models capable of performing the task of object detection. For each of them, several algorithms can be tested. In particular, the following aspects are critical in choosing the best model: the variability of objects contained in the same class, as well as a change in perspective, and illumination, which can create shadows or reflections and cause a significant loss of information. Another aspect to consider was the time efficiency, memory management and storage needed to train these detectors [22]. For this solution, the MobileNet SSD (Single Shot Detector) model and the YOLOv3 (You Only Look Once Version3) model were compared, both trained on Microsoft COCO (Common Objects in COntext), both known for Real Time Object Detection [23]. The MobileNet SSD is a Convolutional Neural Network that uses the Single Shot Detector (SSD) object detection technique and can achieve fast object detection optimised for mobile devices, reducing the computing capacity required. The SSD technique is based on a forward convolutional network that generates a collection of fixed-size bounding boxes and a confidence value for the presence of object class instances in those boxes [24]. YOLO is a deep convolutional neural network. Its algorithm applies a single neural network to the entire full image. Then this network divides that image into regions which provide the bounding boxes and predict probabilities for each region. These generated bounding boxes are weighted by the predicted

probabilities. The YOLO v3 is the real-time object detection algorithm version that identifies specific objects in videos or images. The COCO dataset is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of 328K images. The dataset can detect and labelling 91 object categories. In this first scenario, we tested the model in detecting the operator inside the forklift. Both solutions reported excellent results. The SSD was faster but less accurate, especially in identifying small objects, in line with previous work [25]. We, therefore, chose to use YOLOv3 for our detection models. Next, the confidence score was adapted to the identification of the operator so that it could be identified even in poor lighting circumstances. Finally, the Non Maximum Suppression (NMS) technique was implemented. When an object detection pipeline is applied, several proposals for classification are generated. The proposals are the candidate regions for the object of interest. However, processing many proposals in a real-time solution can be expensive. The NMS technique filters the proposals by calculating the Intersection Over Union and comparing it with the threshold value (Figure 2).

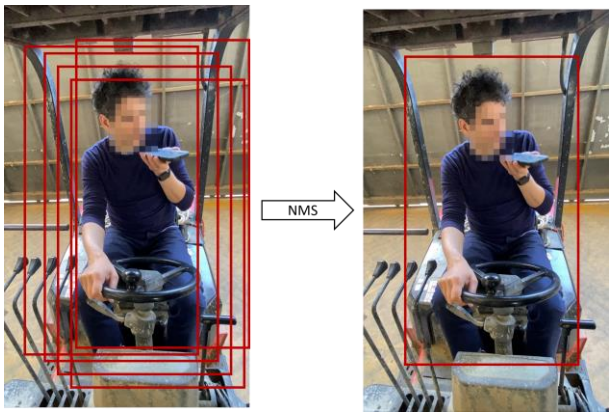


Figure 2. Non Maximum Suppression

B. *Mobile phone detection model*

YOLOV3 and the COCO dataset were also used for the mobile phone detection model. For this model, lower thresholds for the confidence score and NMS had to be chosen so that the phone could also be identified at different angles (frontal, lateral and oblique positions).

C. *Hands detection model*

Human pose estimation technology is the subject of active research worldwide in the fields of sports, surveillance, work monitoring, home care for the elderly and entertainment. In general, human pose estimation is categorized into 2D and 3D coordinate estimation methods, single-person and multi-person methods, based on the number of target subjects and the number of cameras and based on the type of input (video or image) [26]. In our case, we used a fixed camera with a video, a single person and 2D coordinates. Specifically, we have used MediaPipe, Google's open-source framework for building machine learning pipelines for processing time-series data like video, audio, and images. It was trained on COCO dataset and consists of multiple functions. For the hands' detection model, we have implemented the MediaPipe Hand Landmarker which detects the 20 landmarks of the hands in an image, as shown in Figure 3. The model outputs hand landmarks in image 2D coordinates and detects handedness (left/right hand). Along with this, once the camera was fixed, we detected the coordinates of the steering wheel by marking a circle on it. Next, a function was constructed that took as input the coordinates of the hands and the steering wheel and compared whether the former was within the area of the steering wheel. A threshold value of 30 per cent of the landmarks was set.

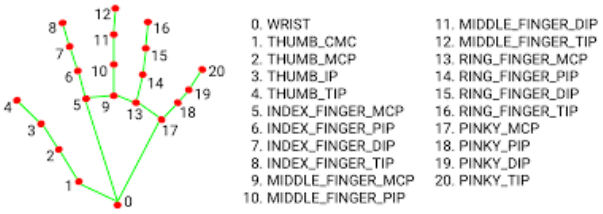
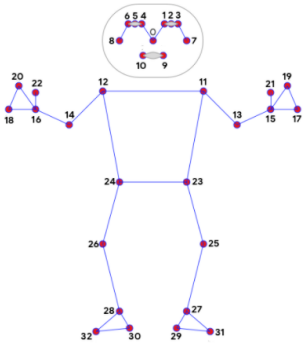


Figure 3. MediaPipe Hand Landmarker

D. *Pose detection model*

For this model, MediaPipe Pose Landmarker was used. MediaPipe Pose constructs pipelines and handles cognitive information in video format by leveraging machine learning (ML) techniques. MPP employs BlazePose, a computational model capable of identifying and extracting 33 two-dimensional landmarks on the human body, as depicted in Figure 4. BlazePose is a lightweight ML architecture designed specifically to deliver real-time processing capabilities on both mobile phones and PCs, utilizing CPU inference methods [26]. Of the estimated MPP landmarks, we have used 14 landmarks to detect torsion. In addition, a ML classification model was trained to identify the direction of the torsion. Specifically, we tested four algorithms, reporting the best results with the random forest classifier. The model was trained by passing as input both the videos of the operator on the forklift and the coordinates of the identified

landmarks, both previously labelled indicating the direction of the twist.



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

Figure 4. MediaPipe Pose Landmarker

V. THE EXPERIMENT

This section reviews the final proposed framework. Figure 5 shows the overall flow diagram of the proposed dangerous behaviours detection system. Once the tool is started the real-time camera continuously sends frames to the tool that detects if the operator is in the forklift. Once the operator is detected (Figure 7), the tool continuously monitors the operator detecting dangerous behaviours. Once the dangerous behaviour is detected, the system sends an alert. Three scenarios are represented. In the first scenario, the tool extracts 2D pixel coordinates of 20 hands’ landmarks Figure 8, detects the wheel and evaluates if the hands’ landmarks are inside the wheel’s bounding circle. A threshold on the number of landmarks is set in this scenario. If the operator moves his fingers or turns his hand downwards taking some landmarks out of the wheel circle the alert is not sent. If the operator raises his hand bringing all landmarks out of the wheel circle the alert will be sent. In the second scenario, the use of the mobile phone is monitored. The mobile phone can be identified in various positions Figure 9. All positions send an alert. Finally, in the third scenario, reverse operations are monitored. The scenario is activated when the system receives a reverse gear activation signal. The system then identifies the 6 torse landmarks and assesses whether the operator is performing the twist Figure 10. Also in this scenario, there is a threshold on the number of landmarks. If at least 50% of the landmarks do not disappear, the torso twisting movement is incorrect and an alert is sent. After the alert is sent, the system continues its monitoring.

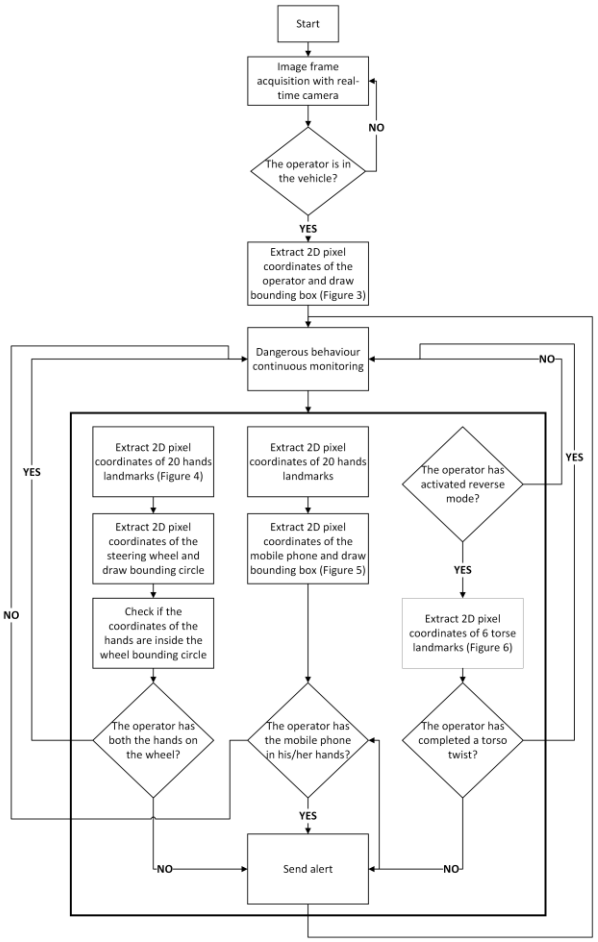


Figure 5. Experiment flow diagram

Figure 6. Forklift

Figure 7. Operator detected

Figure 8. Hands detected

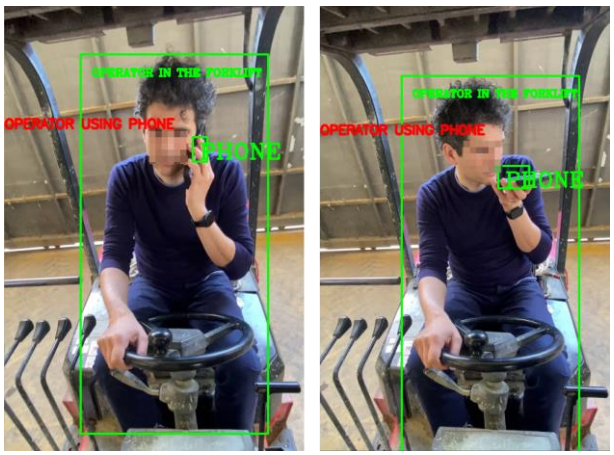


Figure 9. Mobile Phone detected

Figure 10. Reverse pose detected

VI. Discussion and Conclusion

This paper deals with the problem of detecting dangerous behaviours of forklift drivers, in industrial settings. Many incidents are recorded, and current research develops solutions only to monitor external forklift obstacles. For this reason., the paper presents an innovative computer vision-based approach to automatically detect driver behaviour, especially mobile phone usage detection while driving and dangerous behaviours in critical operations such as reverse driving. Although the developed system manages to classify the operator's behaviours with good accuracy, it presents some criticalities. Specifically, some frames show errors related to the detection of false positives and false negatives. These are linked to limitations of the model and problems related to the quality of the frame (e.g., positioning and lighting). The article introduces an innovative technological solution aimed at ensuring the safety of the operators. However, the proposed solution also introduces organizational, technical, and social aspects that should not be underestimated. From an organizational and technical point of view, the solution requires an initial investment and a model training and testing phase that management should accept. From a social point of view, we need to prepare employees for human-ai collaboration. Employees must understand, and preferably experience, the strengths, problems, and risks of the computer vision solution to successfully integrate it into their

job tasks. Data security and ethics are for example major risk areas. The solution proposed is completely anonymized. Moreover, from an ethical perspective, the solution is in line with the Ethics guidelines for trustworthy AI [27]. Future research foresees the inclusion of a z-coordinate useful for 3D coordinate recognition. A 3D scenario can change the identification of the key points of the steering wheel, which is currently represented in 2D as a circle. Furthermore, future steps include its testing in a real scenario with integration to audio sensors and forklift reverse gear sensors.

VII. REFERENCES

[1] E. SUN, X. MA, M. LI, *Improved SSD based pedestrian detection algorithm for forklift active warning system*, IMCEC 2022 - IEEE 5th Adv. Inf. Manag. Commun. Electron. Autom. Control Conf. (2022) pp. 1523–1528.

[2] OSHA, *Safe Forklift Operation*, (2019).

[3] S. COLABIANCHI, M. BERNABEI, F. COSTANTINO, *Chatbot for training and assisting operators in inspecting containers in seaports*, Transp. Res. Procedia. 64 (2022) pp. 6–13.

[4] F. COSTANTINO, A. FALEGNAMI, L. FEDELE, M. BERNABEI, S. STABILE, R. BENTIVENGA, *New and Emerging Hazards for Health and Safety within Digitalized Manufacturing Systems*, Sustain. 2021, Vol. 13, Page 10948. 13 (2021) pp. 10948.

[5] T. BANDA, A.A. FARID, C. LI, V.L. JAUW, C.S. LIM, *Application of machine vision for tool condition monitoring and tool performance optimization–a review*, Int. J. Adv. Manuf. Technol. 2022 12111. 121 (2022) pp. 7057–7086.

[6] A. LANG, W.A. GÜNTHNER, *Evaluation of the usage of support vector machines for people detection for a collision warning system on a forklift*, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 10293 LNCS (2017) pp. 322–337.

[7] N. DARAPANENI, M. RAVIKUMAR, S. SINGH, A. TIWARI, S. DAS, A.R. PADURI, A. BALARAMAN, G. PRATAP, *Computer Vision Application in Automobile Error Detection*, 2022 Int. Conf. Interdiscip. Res. Technol. Manag. IRTM 2022 - Proc. (2022).

[8] C.N. GIREESH BABU, K.T. CHANDRASHEKHARA, J. VERMA, M. THUNGAMANI, *Real time alert system to prevent Car Accident*, 2021 Int. Conf. Forensics, Anal. Big Data, Secur. FABS 2021. (2021).

[9] A. RACHMAN, J. SEILER, A. KAUP, *CAMERA SELF-CALIBRATION: DEEP LEARNING FROM DRIVING SCENES*, Proc. - Int. Conf. Image Process. ICIP. (2022) pp. 2836–2840.

[10] A.Ş. ŞENER, I.F. INCE, H.B. BAYDARGIL, I. GARIP, O. OZTURK, *Deep learning based automatic vertical height adjustment of incorrectly fastened seat belts for driver and passenger safety in fleet vehicles*, Https://Doi.Org/10.1177/09544070211025338. 236 (2021) pp. 639–654.

[11] S. BATTIATO, S. CONOCI, R. LEOTTA, A. ORTIS, F. RUNDO, F. TRENTA, *Benchmarking of computer vision algorithms for driver monitoring on automotive-grade devices*, 2020 AEIT Int. Conf. Electr. Electron. Technol. Automotive, AEIT Automot. 2020. (2020).

[12] A. MISHRA, J. PUROHIT, M. NIZAM, S.K. GAWRE, *Recent Advancement in Autonomous Vehicle and Driver Assistance Systems*, 2023 IEEE Int. Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2023. (2023).

[13] B. LOUNGANI, J. AGRAWAL, L. JACOB, *Vision Based Vehicle-Pedestrian Detection and Warning System*, Proc. - 2022 4th Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2022. (2022) pp. 712–717.

[14] H. NI, F. LI, *Fast pedestrian detection using T-CENTRIST in infrared image*, Infrared Laser Eng. Vol. 49, Issue S2, Pp. 20200423 - -1. 49 (2020) pp. 20200423.

[15] T.H.N. LE, Y. ZHENG, C. ZHU, K. LUU, M. SAVVIDES, *Multiple Scale Faster-RCNN Approach to Driver's Cell-Phone Usage and Hands on Steering Wheel Detection*, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. (2016) pp. 46–53.

[16] A. BARODI, A. BAJIT, T. EL HARROUTI, A. TAMTAOUI, M. BENBRAHIM, *An enhanced artificial intelligence-based approach applied to vehicular traffic signs detection and road safety enhancement*, Adv. Sci. Technol. Eng. Syst. 6 (2021) pp. 672–683.

[17] K. SESHADRI, F. JUEFEI-XU, D.K. PAL, M. SAVVIDES, C.P. THOR, *Driver cell phone usage detection on Strategic Highway Research Program (SHRP2) face view videos*, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. 2015-October (2015) pp. 35–43.

[18] A. ULHAQ, J. HE, Y. ZHANG, *Deep actionlet proposals for driver's behavior monitoring*, Int. Conf. Image Vis. Comput. New Zeal. 2017-December (2018) pp. 1–6.

[19] H. YASAR, *Detection of Driver's mobile phone usage*, HNICEM 2017 - 9th Int. Conf. Humanoid, Nanotechnology, Inf. Technol. Commun. Control. Environ. Manag. 2018-January (2018) pp. 1–4.

[20] S. FOUZIA, R. KLETTE, *Comparing Trackers for Multiple Targets in Warehouses*, Int. J. Fuzzy Log. Intell. Syst. 19 (2019) pp. 147–157.

[21] C.W. YOU, N.D. LANE, F. CHEN, R. WANG, Z. CHEN, T.J. BAO, M. MONTES-DE-OCA, Y. CHENG, M. LINT, L. TORRESANI, A.T. CAMPBELL, *CarSafe App: Alerting drowsy and distracted drivers using dual cameras on smartphones*, MobiSys 2013 - Proc. 11th Annu. Int. Conf. Mob. Syst. Appl. Serv. (2013) pp. 13–26.

[22] L. BARBA-GUAMAN, J.E. NARANJO, A. ORTIZ, *Deep learning framework for vehicle and pedestrian detection in rural roads on an embedded GPU*, Electron. 9 (2020).

[23] L. ZHAO, S. LI, *Object detection algorithm based on improved YOLOv3*, Electron. 9 (2020).

[24] A. HOWARD, M. SANDLER, B. CHEN, W. WANG, L.C. CHEN, M. TAN, G. CHU, V. VASUDEVAN, Y. ZHU, R. PANG, Q. LE, H. ADAM, *Searching for mobileNetV3*, Proc. IEEE Int. Conf. Comput. Vis. 2019-October (2019) pp. 1314–1324.

[25] A.C. RIOS, D.H. DOS REIS, R.M. DA SILVA, M.A. DE SOUZA LEITE CUADROS, D.F. TELLO GAMARRA, *Comparison of the YOLOv3 and SSD MobileNet v2 algorithms for identifying objects in images from an indoor robotics dataset*, 2021 14th IEEE Int. Conf. Ind. Appl. INDUSCON 2021 - Proc. (2021) pp. 96–101.

[26] J.W. KIM, J.Y. CHOI, E.J. HA, J.H. CHOI, *Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model*, Appl. Sci. 13 (2023).

[27] EUROPEAN COMMISSION, Ethics Guidelines for Trustworthy AI, 2019.