IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 May 2024, accepted 11 August 2024, date of publication 14 August 2024, date of current version 28 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3443638

RESEARCH ARTICLE

SCAN: Surveillance Camera Array Network for Enhanced Passenger Detection

PAVOL KUCHÁR[1], RASTISLAV PIRNÍK[1], JÚLIA KAFKOVÁ[1], TOMÁŠ TICHÝ[2], JANA ĎURIŠOVÁ[3], AND MICHAL SKUBA[1]

[1]Department of Control and Information Systems, Faculty of Electrical Engineering and Information Technology, University of Žilina, 010 26 Žilina, Slovak Republic
[2]Faculty of Transportation Sciences, Czech Technical University in Prague, 110 00 Prague, Czech Republic
[3]Department of Physics, Faculty of Electrical Engineering and Information Technology, University of Žilina, 010 26 Žilina, Slovak Republic

Corresponding author: Pavol Kuchár (pavol.kuchar@uniza.sk)

This work was supported by Mobile Robotic Systems During Crisis Situations through Vedecká grantová agentúra MŠVVaM SR a SAV (VEGA) project under Grant 1/0241/22.

ABSTRACT The automated detection of individuals within vehicles, with minimal to no human intervention, hold multifaceted implications in contemporary contexts. These applications span from aiding emergency responders and optimising transportation networks to facilitating automated crash response mechanisms and enforcing regulations concerning High Occupancy Vehicle and High Occupancy Toll lanes. In this paper, we introduce our camera system designed for passenger counting, leveraging five IP cameras equipped with a range of optical filters, employing image registration techniques, and integrating the YOLOv8 object detection model. The Surveillance Camera Array Network (SCAN) operates within the near-infrared domain of the electromagnetic spectrum in conjunction with the visible part. Four VIVOTEK IP cameras are outfitted with near-infrared, neutral density, polarising, and ultraviolet optical filters, while the final camera retains its stock lens. Our primary challenge lies in managing variable lighting conditions throughout the day. However, during nighttime, we achieve nearly perfect image capture of vehicles. To mitigate noise, glare, and other impediments, we initially apply camera calibration, image preprocessing, cropping, image registration, and finally, image fusion. Our findings demonstrate that our cost-effective SCAN system adeptly detects passengers in cars equipped with window tinting. The results obtained during testing conditions resulted in an 66% true positive rate, 8% false positive rate and 26% false negative rate within best dataset. Additionally, we provide created datasets displaying passengers inside Suzuki Vitara, Jaguar XF, and Honda CRV vehicles with various levels of window tinting, to facilitate future community endeavors in addressing this challenging task.

INDEX TERMS Camera calibration, digital image processing, HOV, image fusion, image registration, machine learning, near-infrared filter, occupancy estimation, passenger detection.

I. INTRODUCTION

Researchers have become increasingly interested in automatically counting people inside vehicles in recent years. Detecting occupants is a fascinating and relatively unexplored field for professionals in various industries, including transportation, surveillance, autonomous vehicles, emergency services, and the military. It plays a crucial role in enhancing security and safeguarding individuals, facilitating monitoring across different sectors. This holds particular significance in contexts such as road transport, traffic management, and crowded environments, where unauthorised access or suspicious behaviour can present substantial risks.

The information collected by passenger detection systems can also serve as valuable statistical data for urban planning, enhancing infrastructure, improving existing roads, establishing High-occupancy vehicle (HOV) lanes, and enforcing penalties for unauthorized lane usage. This contributes to mitigating vehicle emissions and provides additional advantages [1], [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

Additionally, to enhance emergency response on roads, one crucial piece of information is the number of individuals involved in a traffic accident or in perilous situations, such as within a road tunnel. For instance, during a fire in a tunnel, chaos ensues, compounded by the confined spaces and thick smoke. Hence, obtaining an accurate estimate of the number of people in the tunnel is vital for rescue services. Depending on this count, they can swiftly mobilize adequate equipment and personnel upon receiving the emergency call, ensuring efficient management of the situation and maximising life-saving efforts.

Lastly, the passenger detection system can prove highly beneficial for military or law enforcement purposes. Numerous areas stand to benefit from such a system, enhancing security and efficiency. Border crossings offer a prime example. Implementing a passenger detection system would enable personnel to inspect vehicles more swiftly and effectively, deterring or impeding criminals and smugglers from engaging in illicit activities.

In recent years, passenger car manufacturers have made concerted efforts to reduce vehicle heating in direct sunlight. They've explored new additives in glass production and experimented with window tinting [4], [5]. These measures aim to enhance thermal comfort in vehicles during extreme heat and protect internal components from overheating and damage. Additionally, it contributes to the fuel economy and reduces energy consumption [6]. Consequently, this study will primarily investigate the previously unexplored impact of automotive window tints on the detection of individuals. This analysis extends to window tints applied to the windows of passenger cars, offering a side non-invasive view of the rear seats, as well as the examination of new solar windshields. These additions to clear automotive glass might present new obstacles in existing systems. Over the past years, acquiring vehicle occupancy data from external cameras, typically positioned atop highways to oversee each lane, has been a common practice. Despite being a challenging area of study, recent progress in artificial intelligence and computer vision indicates the potential of utilising such cameras to enforce road regulations. However, the increasing use of window tinting and new solar windshields in cars may impact the effectiveness of these systems.

In this study, we propose a SCAN system consisting of five cameras with various optical filters and active near-infrared (NIR) illuminator aimed at tackling the challenges associated with passenger detection. This system leverages machine learning methods and image registration techniques to enhance vehicle occupancy detection.

The main contributions of our work can be summarised as follows:

- We propose a flexible design of camera array network which can better capture the number of passengers inside vehicles with tinted windows, solar windshields, or clear glass.
- Our proposed scheme with various stages of image preprocessing allows our SCAN system to adapt to different environments, making it suitable for various applications ranging from urban traffic management to border security.
- The SCAN system offers a cost-effective solution compared to traditional methods, as it eliminates the need for expensive infrastructure.
- Our experimental results demonstrate that our methods achieve innovative outcomes in the field of occupancy detection.
- With easy adaptability of optical filters to emerging automotive technologies such as advanced window tints and solar windshields, the SCAN system ensures continued effectiveness and relevance in the face of evolving vehicle designs and features.

II. RELATED WORK

Over the years, a multitude of techniques have been developed and utilised for accurately counting individuals across various environments and scenarios. These techniques can be broadly categorised into two distinct groups, each offering unique advantages and applications [7]. The first category encompasses invasive systems, which rely on sensors positioned inside vehicles. The second category comprises non-invasive systems (usually camera-based solutions), which utilise video surveillance cameras and sophisticated image processing algorithms to identify and track individuals within a given space.

A. INVASIVE PASSENGER DETECTION METHODS

A commonly adopted method for passenger detection in vehicles involves invasive sensing of individuals from within the vehicle using a variety of methods, principles, and sensor types. While this approach is typically simpler and more cost-effective, it necessitates forwarding the detection results to a centralised system or network for subsequent processing, which often lacks standardisation and is absent in many regions.

Previous papers have explored automated methods for detecting passengers in moving vehicles using a variety of technologies, including standard cameras, thermal cameras, radars, PIR sensors, infrared cameras, TOF cameras, or measuring the variation in impedance between electrodes by utilising dielectric dispersion within human tissue [8], [9], [10], [11], [12], [13].

The increasing popularity of vision systems in security applications underscores their significance. However, utilising computer vision poses considerable challenges, especially in environments with extreme lighting variations, ranging from intense brightness to dark nights. This issue extends to noninvasive detection, where the presence of shadows, both stationary and moving, further complicates matters. In an effort to enhance passenger safety and comfort, Gautama et al. introduced a stereo system designed to monitor the cockpit scene and refine airbag firing control [14]. The paper explores various techniques and assesses the impact of

random and systematic errors on critical parameters such as robustness and processing speed.

Faber introduced a stereo system for an advanced airbag system in [15] and [16]. This system categorises seats and endeavours to estimate the geometry and positioning of occupants' heads. The human head's shape is represented by an ellipsoid model, with two monochromatic cameras affixed to the windshield for data capture, eliminating the need for supplementary lighting. Stereo cameras have been used in [17] and [18].

In 2004, a camera with a 360° parabolic mirror and NIR LEDs, along with Viola's classifier cascade, was proposed [19]. Modifications were suggested by Wender and Loehlein to enhance Viola's cascade classifier for crash and occupant information, resembling the EU's e-Call system. Géczy et al. proposed small form-factor IR sensors [20], achieving promising results with the AMG8833 sensor from Panasonic, particularly in detecting front passengers. However, rear passenger detection proved challenging due to distance and low resolution, requiring additional sensor nodes.

The pressure characteristics when individuals ascend or descend stairs in public transportation settings are investigated in [21]. This method distinguishes the direction of passengers, addressing limitations in existing technologies, which often fail to discern passengers' forward direction and are cost-prohibitive. A passenger flow counting method based on human body kinematics and Support Vector Machine (SVM) is proposed in [22]. This method utilises four pressure sensors on the bus stair pedal for analog output. The walking process is segmented into the supporting and swing phases.

Invasive passenger detection methods in vehicles present a range of advantages and disadvantages. On the positive side, these methods are often simpler and more cost-effective to implement compared to non-invasive alternatives. They typically utilise straightforward sensor technologies, making them accessible for a wide range of applications. Moreover, invasive sensors provide direct measurements of passengers within the vehicle, enabling real-time monitoring and possible immediate response in emergency situations if equipped with eCall. Additionally, in some scenarios, invasive sensors may offer higher accuracy and reliability compared to non-invasive methods, particularly in environments with challenging conditions such as low visibility or extreme temperatures. However, these benefits come with significant drawbacks. One major concern is privacy, as invasive sensors directly intrude on the privacy of individuals within the vehicle, raising ethical and legal issues regarding surveillance and data privacy. Furthermore, retrofitting invasive sensor systems may require significant modifications to the vehicle's interior, increasing installation complexity and potentially affecting vehicle aesthetics and functionality. Compatibility issues may also arise, as invasive sensors may not be suitable for all vehicle types or models, limiting their applicability and scalability across different vehicle fleets. Moreover, invasive sensor systems may require regular maintenance and calibration to ensure accurate performance, adding to the overall operational costs and logistical challenges. Additionally, results obtained from invasive sensors often need to be processed by external systems for meaningful interpretation, leading to dependencies and potential interoperability issues with non-standardised processing systems.

B. NONINVASIVE PASSENGER DETECTION METHODS

Detecting passengers noninvasively holds significant importance in the advancement of safer and more efficient transportation infrastructure while respecting the privacy of individuals within vehicles.

Several studies suggested using additional equipment, like multiple cameras capturing images of vehicles from various angles, to automatically count passengers on the road instead of inside specific vehicles [23], [24].

Image segmentation techniques for identification of windshield areas, and face detection is employed to count occupants in [25]). Birch et al. used color CCD cameras, converting images from RGB to HSV color space. After noise removal with a median filter and binary labeling, the largest area search is conducted for dilatation, erosion, and windscreen. The color mask undergoes post-processing. This method has shown reliability for approximately 80% of automobiles and trucks, although inconsistent lighting led to the discovery of only 38% of faces.

Capturing a comprehensive image of the human body inside a car from an external perspective is often challenging due to the obstruction or blurring of the face, which can vary depending on the viewing angle [26]. In 2008, Tyrer and Lobo explored passenger occupancy limitations, conducting night tests using IR and visible spectrum images [27]. They showcased reflectance spectra for various skin types and transmission/absorption spectra for typical windscreens.

Pérez-Jiménez et al. enhanced windshield detection by combining information from multiple classifiers [28]. Their approach involved searching for features like faces and safety belts using boosted classifiers, followed by refinement with a k nearest neighbor filter. This system achieved a remarkable success rate of nearly 90% with just a 2% false detection rate.

In noninvasive passenger estimation, windshield localisation is a crucial aspect. Yuan et al. explored this, using a maximum energy method to extract windshield regions and HOG descriptors for occupant detection within them [29]. However, windshield detection presents various challenges, including shape variations, low contrast, and diverse capture conditions. To address these, the authors proposed an integrated approach, combining shape, grayscale color, and complexity information for both color and NIR images. For occupant detection, they employed HOG descriptors to capture essential features succinctly.

Cornett et al. investigated the development of a multi-unit computational camera system to achieve consistent face recognition outcomes [30], [31]. Employing HDR (High

Dynamic Range) imaging, the system generates a dataset of through-windshield images. Overcoming challenges such as target distance, poor lighting conditions, intense glare, occupant poses, and vehicle speed constitute the primary objectives of this system.

Noninvasive passenger detection methods in vehicles offer several advantages but also come with certain limitations. On the positive side, noninvasive methods prioritise passenger privacy as they do not require direct intrusion into the vehicle's interior. This aspect addresses ethical and legal concerns surrounding surveillance and data privacy, making noninvasive solutions more socially acceptable. Additionally, the implementation of noninvasive sensors typically involves less complexity. Moreover, these methods tend to be more versatile and compatible across different vehicle types and models, facilitating broader applicability and scalability. However, noninvasive passenger detection methods also have limitations. They may be less accurate and reliable in certain scenarios, especially in challenging environments with factors like snow, fog or direct sunlight. Furthermore, the processing of data obtained from noninvasive sensors may still require external systems, leading to dependencies and potential interoperability issues with non-standardised processing systems. Also, the primary challenge lies in accommodating various vehicle types such as trucks, buses, passenger cars, and specialty vehicles, necessitating modifications to occupancy estimation systems for each scenario. Ensuring a robust and universally applicable system entails adjusting sensor height or position accordingly. Despite these drawbacks, noninvasive methods remain a promising approach for passenger detection in vehicles, balancing privacy considerations with functional effectiveness.

We strongly advise interested readers to explore our previously published review article, which comprehensively examines the existing literature in the domain of transportation safety and efficiency, particularly focusing on occupancy estimation within vehicles and passenger detection at public transport stations. The article also includes a comparative analysis of various approaches to passenger estimation [7].

Since the release of our review paper, further studies have been documented. Wasista et al. introduced a passenger face detection system based on webcams, employing the Single Shot Detector (SSD) method intended for implementation within buses [32]. The bus passenger detection is also investigated by Li et al. in paper [33]. Their lightweight bus passenger detection model based on YOLOv5 is customised to achieve better detection speed. They succeeded in boosting the speed by 6% without compromising accuracy. Furthermore, Pronello and Garzón Ruiz [34] employed the YOLOv5 object detection algorithm and a custom automatic passenger counting system (APC) based on a Raspberry Pi with a camera. Their findings showcased an equivalent or even superior level of accuracy compared to more costly commercial systems. A system for tallying the number of passengers at waiting points is introduced by Morozov et al. in a form of a single platform [35]. Their prototype enables passenger counting, monitoring of public transport, and management of public transportation services. Lastly, Hyun et al. [36] proposed a passenger monitoring scheme utilising a 60GHz FMCW radar for in-cabin applications with an average recognition rate at 96%.

Furthermore, our initial discoveries concerning the spectral transmittance of various car glass types, window tinting for cars, and solar windshields are detailed in our previous paper [37]. In this paper, we provide further spectral transmittance measurements in Chapter III-C.

III. MATERIALS AND METHODS

The presence of car window tinting presents a substantial impediment to the effective capture of passengers through exterior cameras for several reasons. Firstly, the attenuation of visible light due to the darkened tint reduces interior brightness, thereby impeding the cameras' ability to capture clear images. Moreover, tinted windows may introduce glare or reflections, further obscuring the occupants from external view. Additionally, disparities in tint darkness among vehicles, and even within the same vehicle, contribute to inconsistencies in image quality, complicating the task of reliably identifying and monitoring passengers. These challenges underscore the necessity for alternative methodologies or technological innovations to surmount such hindrances and ensure the dependable surveillance and monitoring of vehicle occupants. Additionally, window tint might cause a significant reduction of vision for automobile drivers [38]. Thus, the primary focus of this study will be on investigating the previously unexplored impact of automotive tints on the detection of individuals and objects affixed to the windows of passenger cars, particularly with regard to a side non-invasive view of the rear seats.

A. IMAGING SYSTEM OVERVIEW

In this study, to evaluate the number of people in a vehicle, we designed a custom camera system consisting of five VIVOTEK IP7361 IP cameras. It is a 2.0-megapixel network camera primarily designed for outdoor usage. Its high-definition video capabilities, day/night functionality, and other features render it an excellent option for various applications, such as monitoring parking lots, gas stations, or perimeter security. Furthermore, its ability to configure multiple video streams for simultaneous monitoring enables us to tailor high-definition video for local monitoring or recording, facilitating the creation of our own dataset. Similar to many surveillance cameras, the IP7361 exhibits sensitivity to the IR spectrum. Equipped with built-in IR lighting, it ensures clear images even during nighttime operations. Table 1 presents common camera parameters, while Table 2 showcases distinct settings for each camera.

Subsequently, four IP7361 cameras were outfitted with NIR, UV (ultraviolet), ND (neutral gray), and PL (polarising) filters manufactured by HOYA. The fifth camera remained

filter-free to provide a reference for undistorted and realistic imagery. Enhancing quality and ensuring optimal penetration, two active NIR illuminators were employed: the LIR-CH88 IR LAB, featuring a 130-meter range and a 25° illumination angle, and the 3N-80/60S2 with an 80-meter range and a 60° illumination angle. Both illuminators operate at an optical wavelength of 850 nm and incorporate a twilight sensor for automatic activation under insufficient lighting conditions. An illustrative example demonstrating the illumination level is depicted in Figure 1.

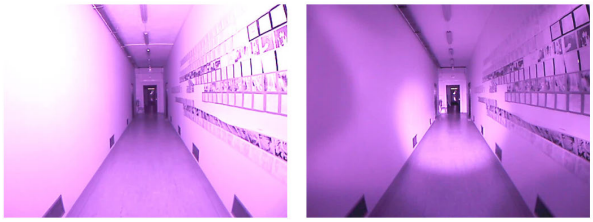


FIGURE 1. Images captured by a camera fitted with a NIR filter, showcasing active illumination from the 3N-80/60S2 (on the left) and the LIR-CH88 (on the right).

TABLE 1. Common camera system parameters.

Parameter	Value
Image resolution	800×600
Max. sampling frequency	30/s
Intra frame period	1/4
Constant bit rate	6 Mbps
Maximum exposure time	1/30s
Exposure level	4
Max. gain	4x
Enable BLC	True

TABLE 2. Individual camera system parameters.

Parameter	UV	PL	ND	Stock	NIR
Brightness	+3	+2	+5	+1	+0
Contrast	+2	+2	+0	+0	+0
Saturation	+1	+5	+0	+3	+5
Sharpness	+1	+2	+0	+0	+1
White balance	Keep value	Auto	Auto	Auto	Keep value
Orientation	-	-	Flip and mirror	Flip and mirror	Flip and mirror
Iris mode	Fixed	Fixed	Outdoor	Fixed	Fixed
Color	B/W	Color	Color	Color	Color

The powering of individual cameras adheres to the IEEE 802.3af PoE (Power over Ethernet) standard. These cameras draw power from an industrial switch through PoE. Furthermore, the system is supplemented by a router responsible for allocating IP addresses to the cameras and computing units such as PCs and notebooks, utilising DHCP (Dynamic Host Configuration Protocol). The cameras, along with their base, were positioned within a custom stand constructed from aluminium structural profiles. Subsequently, all pertinent accessories, including the power supply, external illuminators, switch, router, and computing device for displaying and storing captured data, were affixed to the stand. The configuration of this image recording system is depicted in Figure 2.

FIGURE 2. 3D model of proposed sensing system.

Given the use of five cameras, it is imperative to carefully select their orientation and relative positions to ensure consistency in the captured images for subsequent processing. Consequently, designing an appropriate method for anchoring the cameras becomes essential. Placing all cameras in a single line is not viable for seamless image registration, as it would lead to significant shifts in perspective, resulting in misalignment between the field of view of the edge cameras. Thus, we opted for a two-tiered arrangement, with three cameras positioned on one level and two cameras situated either above or below them. This configuration ensures adequate coverage of the cameras' field of view, facilitating accurate image registration in subsequent analyses.

Initially, the cameras were attached to a 3D-printed bracket. However, we later transitioned to a custom steel bracket to address concerns regarding robustness and minimise camera movement, aiming for optimal results during image registration.

For seamless image registration, precise positioning and alignment of the cameras relative to each other are paramount. Given the flexibility of adjusting the mutual rotation of the cameras through the bracket, we have proposed three configurations that appear to be the most suitable, as depicted in Figure 3:

(a) Two cameras aligned in parallel, with three directed towards a common point, and these groups aligned towards a shared focal point.
(b) Cameras placed in parallel.
(c) Two cameras aligned in parallel, with three others also in parallel, and these groups oriented towards a common focal point

Following an evaluation of the strengths and weaknesses of each configuration, we opted to proceed with configuration 3a. The common focal point was positioned at a distance of two meters from the edge of the optical filters, ensuring robustness for subsequent image registration.

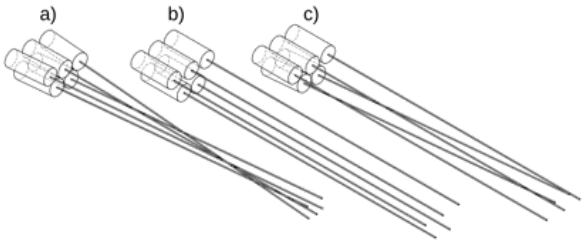


FIGURE 3. Possible mutual configuration of cameras.

B. IMAGE ACQUISITION

Undoubtedly, the most critical and time-consuming phase of any machine learning endeavor is the collection of data to construct a model. Much akin to procuring materials for a construction project, this phase forms the bedrock upon which the entire project is built. The quality and quantity of gathered data significantly influence the performance and accuracy of the resultant model. Acquisition of imagery can prove to be a multifaceted, resource-intensive endeavor, demanding expertise and a profound comprehension of project objectives. Moreover, it is imperative to ensure that the obtained images are pristine, pertinent, and representative of the problem at hand, as any inaccuracies during this stage can proliferate throughout the entirety of the machine learning process. In essence, the successful collection of data stands as the cornerstone of a resilient and dependable machine learning model. Consequently, in this chapter, we delineate the methodologies through which we will procure images to construct our passenger detection model.

Considering the significant impact of image resolution on the performance of convolutional neural networks (CNN), we opted for a camera resolution setting of 800 × 600 pixels [39]. This configuration ensures a stable frame rate of 30 fps (frames per second). Increasing the resolution to 1600 × 1200 pixels would reduce the frame rate to 10 fps, significantly prolonging the neural network's training time and notably worsening the lag between synchronised image acquisition.

To ensure successful image registration, it is beneficial for the captured images to exhibit high similarity across different perspectives, encompassing varying cameras, times, or angles. Our previous chapters have outlined the selection process for cameras and the angles of individual shots (inter-camera configuration). However, the temporal dimension remains unexplored. In our practical scenario, simultaneous image capture is imperative. To streamline the system and facilitate remote control, we sought methods for software-based synchronised image capturing. While numerous options exist for displaying multiple camera feeds concurrently, the functionality for synchronised image and video saving was found to be hard to obtain. Through the utilisation of Threading (independently executing tasks) in Python, we successfully attained this objective.

The images are stored at the maximum quality allowed by the camera, maintaining a high frame rate (800×600@30fps). While it's feasible to enhance the image resolution, doing so would substantially decrease the camera's frame rate. This reduction could introduce substantial discrepancies between individual images, thereby compromising the quality of image registration.

The proposed methodology for passenger detection is illustrated in Figure 4. Following the acquisition of individual images from IP cameras, a correction process is initiated to mitigate distortion induced by optical filters. Subsequently, preprocessing is conducted, succeeded by the extraction of regions of interest (ROIs), encompassing views of the windshield or side windows, and the alignment of image pairs relative to a camera equipped with a NIR filter.

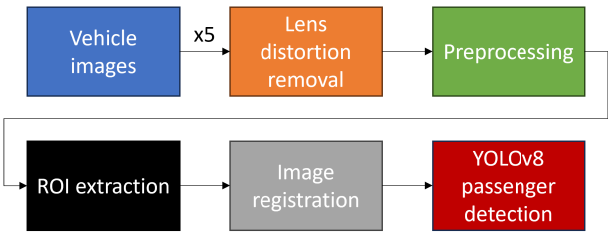


FIGURE 4. A schematic depicting passenger counting through computer vision techniques. Following the extraction of the region of interest (ROI) from side images of the vehicle, the images are registered, culminating in the determination of the passenger count.

For optimal assessment of the vehicle's interior, it is most advisable to position cameras on both sides of the vehicle. This setup mitigates the risk of occlusion, as capturing from a single side may result in the furthest passenger being obscured by intervening passengers or objects within the vehicle (refer to Figure 5). However, such a configuration is suitable only for single-lane scenarios. In instances involving multiple lanes, there's a possibility of other vehicles obstructing the cameras' field of view. Hence, the system is most conducive for applications where vehicles are stationary within designated lanes or move at a slow pace. Examples of such applications include border crossings or entrances to parking facilities.

Another aspect to consider in camera positioning is the height of the camera system. The varying heights of different

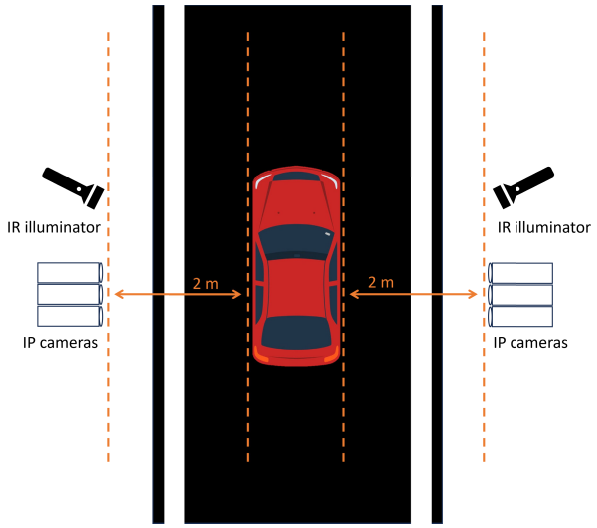


FIGURE 5. Possible configuration of cameras and active NIR lighting on the road for passenger detection.

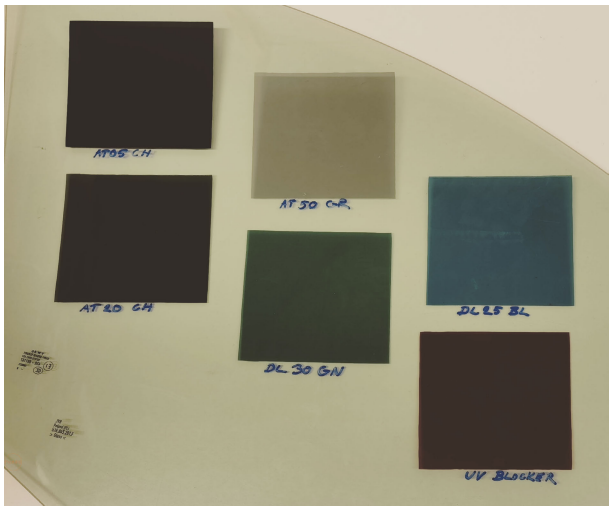


FIGURE 6. Analysed specimen with LLumar tinting applied to the Škoda Rapid front window.

vehicles pose a challenge in finding the optimal placement for cameras. For instance, sedans typically have a height around 1456 mm, while SUVs (Sports Utility Vehicles) range from approximately 1664 mm to well over 2000 mm. Given that our experiments involved vehicles like the Jaguar XF, Suzuki Vitara, and Honda CRV, we opted to position the center of the camera with NIR filter at a height of 1400 mm on the stand. However, it's worth noting that this height may not be suitable for capturing taller vehicles such as buses or trucks, necessitating a reassessment in future endeavors. Previous studies have delved into determining the ideal specifications for camera placements, heights, and angles, as referenced in [40], [41], and [42].

C. SPECTRAL TRANSMITTANCE OF WINDOW TINTING

In our experiments, we selected the most widely used automotive tinting options from the manufacturer LLumar: AT 05 CH SR HPR (very dark charcoal), AT 20 CH SR HPR (dark charcoal), DL 25 BL SR HPR (medium blue semi-reflective), DL 30 GN SR HPR (medium green semi-reflective), AT 50 CH SR HPR (light grey), and UV Blocker. These samples were applied to the windows of a 2017 Škoda Rapid (Figure 6). The experiment utilised incident light from a standard car halogen bulb, directed through the clear window of a 2017 Škoda Rapid, as the reference light source. Experimental results illustrating the spectral transmittance of six measured tints are depicted in Figure 8.

Apart from analysing the automotive tints themselves, we conducted an investigation into and comparison of the spectral transmittances across different manufacturers of automotive glass. Samples were selected from a range of car manufacturers including Volvo, Alfa Romeo, Škoda, Peugeot, and one sample of a solar windshield (see [37]).

Traditional tungsten and carbon filament lamps, along with more recent tungsten-halogen lamps, have proven to be highly dependable light sources in optical microscopy for many decades. These sources adequately serve our needs, generating a continuous spectrum of light spanning from near violet to deep infrared. With the ability to operate at elevated temperatures, the spectrum tends to shift towards shorter wavelengths, predominantly into the blue range, thereby enhancing energy efficiency. Consequently, we regard halogen lamps as an excellent choice of light source, offering a spectrum akin to blackbody radiation, reminiscent of that emitted by the Sun. Thus, a halogen bulb intended for passenger cars served as the source of electromagnetic radiation. It was powered by a constant voltage source, specifically set at 12.006 V throughout our experiments. Detailed images illustrating the setup, including the positioning of the source in the stand, the spectrometer, and the measurement workstation with the sample under evaluation, alongside the resulting spectrum, are presented below Figure 7.

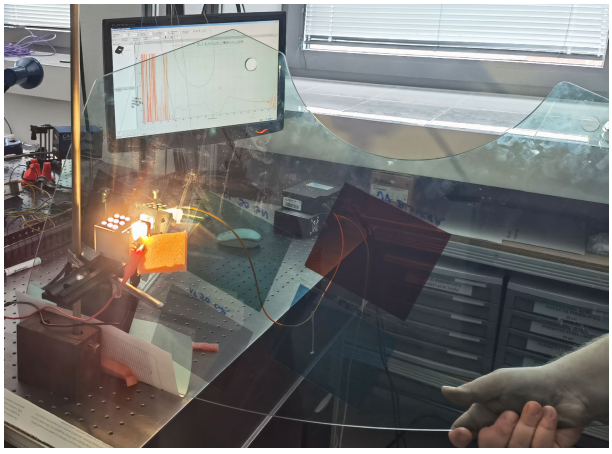


FIGURE 7. Experimental measurement of spectral transmittance for individual car tint samples, alongside the entire measurement apparatus.

As anticipated, the observed results from Figure 8 closely approximate the specifications outlined in the manufacturer's product sheet for the specified car tints. However, an intriguing observation arises regarding the spectral transmittance pattern in the proximity of the NIR region. Notably, the semi-reflective medium green and semi-reflective blue tints exhibit a spectral transmittance rate of approximately 50%, contrasting with the nearly 100% spectral transmittance rate observed in other tints. A tentative inference drawn from these measurements suggests that the NIR region of the electromagnetic spectrum holds promise for the detection of passengers and objects.

Attempting to scan passengers in the visible area with active illuminator could potentially deprive drivers of vision for a short period, posing a safety risk that is not present when using additional lighting in the NIR range.

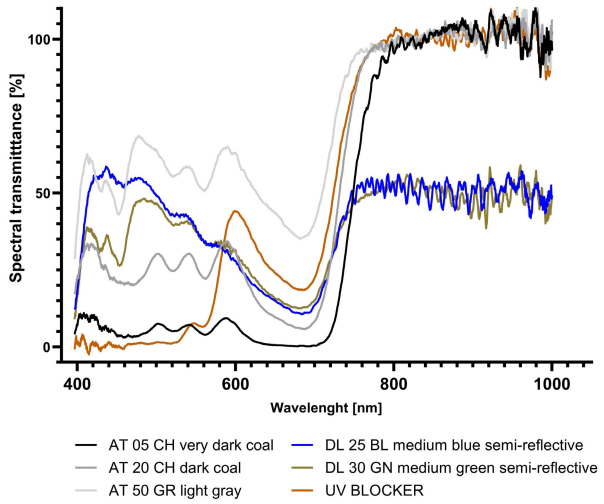


FIGURE 8. Comparison of LLumar window tint spectral transmittance.

D. CAMERA CALIBRATION

Geometric camera calibration, also known as camera resection, entails estimating the parameters of a camcorder's lens and image sensor. These parameters facilitate the correction of lens distortion, the measurement of object sizes in world units, and the determination of the camera's position in a given scene. Such tasks find utility across diverse applications, including machine vision for object detection and measurement, robotics, navigation systems, and 3D scene reconstruction [43], [44].

Calibration algorithms commonly address the pinhole camera model and the fisheye camera model, with the latter offering a field of view of up to 195 degrees. Given that VIVOTEK cameras do not possess such a wide field of view, they are adequately represented using the pinhole camera model. This model depicts a basic camera devoid of a lens, featuring a single small aperture. Light rays traverse through this aperture, projecting an inverted image onto the opposite side of the camera. This model serves as a straightforward depiction of camera functionality.

Simultaneous estimation of all parameters, encompassing distortion coefficients, is achieved through nonlinear least squares minimisation employing the Levenberg-Marquardt algorithm.

We initially presumed that calibrating two cameras simultaneously would yield superior results compared to calibrating them individually. However, upon examining Figure 13a and Figure 13b, it becomes evident that individual calibration yields better blending of samples. Notably, when calibrating pairs of cameras with respect to camera with NIR filter, larger deviations are observable in individual patterns, a phenomenon mirrored in the average reprojection error.

Figure 9 presents a comparison between distorted and undistorted images. Although we observe a loss of some pixels around the image edges, upon closer inspection of the building's windows, we notice that the window plane now appears geometrically correct following distortion removal. Subsequently, we applied distortion removal to all remaining cameras using the calibration coefficients obtained. With the distortion successfully removed, we can now advance to the next stage of processing the acquired images: image registration.

E. IMAGE REGISTRATION

Upon examining our images after distortion removal, it becomes evident that disparities between the images primarily arise from sensor displacement, slight rotation, scaling, and potentially skewing. Consequently, an Affine transformation is employed to characterise the spatial transformation. Control points, essential for identifying and comparing corresponding elements between reference and distorted images, can be manually or automatically selected using the Registration Estimator App from IPT (MATLAB's image processing toolbox).

Initially, we presumed that sensor alignment remains consistent over time, thanks to the robust base. Therefore, registering only one fundamental set of images would suffice, subsequently applicable to all captured images. However, this assumption proved inaccurate, as objects in the images were in close proximity to the cameras, resulting in significant disparities between images. This methodology, however, finds utility in scenarios such as aerial imagery, where salient landmarks are considerably distant, and image disparities are minimal [45].

Consequently, we found it necessary to employ techniques where registering each frame separately became imperative—seeking transformations for each pair of frames. It became evident that registering entire frames did not yield the desired level of quality and overlap, as initially envisaged. Subsequently, we embarked on exploring potential solutions to address this issue. One promising approach involved registering only specific parts of the image likely to contain passengers—namely, the windows. Consequently, we proceeded with further image processing, which entailed cropping the images before registration. To streamline the entire procedure, we endeavoured to automate it by training

(a) original image **(b)** undistorted image

FIGURE 9. Comparison between the original image (left) and the image after distortion removal (right) obtained from a camera equipped with a NIR filter.

the YOLOv8 model to detect various components, including the windshield, front and rear windows, and rear windshield.

F. DATASETS

For optimal passenger detection outcomes, we opted to construct multiple datasets, conduct comparative analyses, and select the most effective one based on passenger detection performance. Detailed distinctions between these datasets are provided later in this section. Initially, we captured 470 images from each camera, among which 53 images served as calibration checkerboards. Additionally, 78 frames captured the perspective through the heavily tinted windows of the Jaguar XF (characterised by a very dark black tint), while 65 frames depicted the windshield. Moreover, 60 images portrayed the view of the Honda CRV (featuring dark black tinting), exhibiting both front and rear passengers, alongside 42 images of its windshield. Furthermore, we gathered 52 windshield images and 119 side view images showcasing front an rear passengers for the Suzuki Vitara (equipped with low black tint). These configurations are illustrated in Figure 11 and a Figure 10 showing verification of our system at the University of Žilina.

The images were subsequently partitioned across all datasets as delineated below:

- Test set: Comprising 119 side images of the Suzuki Vitara (featuring 359 individuals within the images).
- Validation set: Consisting of 52 images portraying the Suzuki Vitara windshield (depicting 109 individuals within the images).
- Training set: Encompassing the remaining images sourced from Jaguar XF and Honda CRV vehicles (illustrating 556 individuals within the images).

Training set underwent augmentation via the following processes:

- Image rotation: Ranging from $-5°$ to $+5°$.
- Saturation adjustment: Varying from -15° to $+15°$.
- Brightness modification: Extending from $-15°$ to $+15°$.
- Exposure alteration: Spanning from $-10°$ to $+10°$.

Upon augmenting the training set, an additional 3 images were generated from each original image by implementing the aforementioned adjustments. Consequently, the initial pool of 243 images expanded to 729 images post-augmentation. The depicted individuals in the images are European men and women. Throughout the photoshoot, participants altered their positions within the vehicles, as well as their headgear and body postures, to facilitate a more diverse dataset.

FIGURE 10. The proposed sensing system capturing images of vehicles during the verification at the UNIZA (University of Žilina) parking lot.

Image processing methodologies are delineated in greater detail in the subsequent paragraphs and examples are shown in Figure 12:

a) The initial approach involved no pre-processing of the images; the neural network received the raw images directly from the cameras without any post-processing, including the removal of distortion.
b) The second dataset was formulated by initially rectifying distortion using acquired camera parameters and subsequently identifying vehicle windows likely to contain individuals.
c) Images from the preceding step (b) were initially aligned through similarity and subsequently through Affine

transformation, utilising a standard optimiser and fusion method allocating channels G to the first image (registered) and R and B to the second image—originating from a camera equipped with a NIR filter. This fusion method, recognised for its high contrast, is particularly suited for individuals with colour blindness (green-magenta).

d) RGB images were cropped and subjected to preprocessing by conversion into grayscale images, thereby eliminating hue and saturation information while retaining brightness. Subsequently, the contrast-limited adaptive histogram equalisation (CLAHE) algorithm was applied [46].

e) Images from the aforementioned step (d) underwent initial co-registration via similarity and subsequent co-registration through Affine transformation, utilising a standard optimiser and the green-magenta fusion method.

f) Frames from step (d) underwent initial co-registration via similarity and subsequent co-registration through Affine transformation, employing an optimiser with reduced step size and a custom fusion method allocating R channel to the first frame and G channel to the second frame.

g) Frames from step (d) underwent initial co-registration via similarity and subsequent co-registration through Affine transformation, employing an optimiser with reduced step size and a custom fusion method assigning R and G channels to the second frame and B channel to the first frame.

FIGURE 11. Various types of images captured by a camera with an NIR filter. Top row: Jaguar XF, middle: Honda CRV, bottom: Suzuki Vitara.

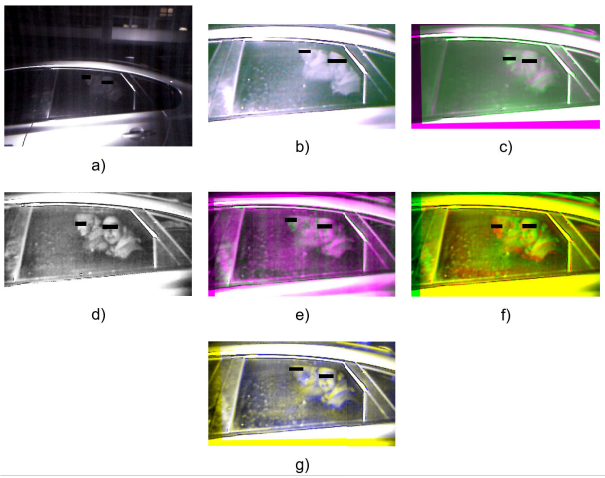


FIGURE 12. Illustration depicting image comparison showcasing various processing techniques across different datasets.

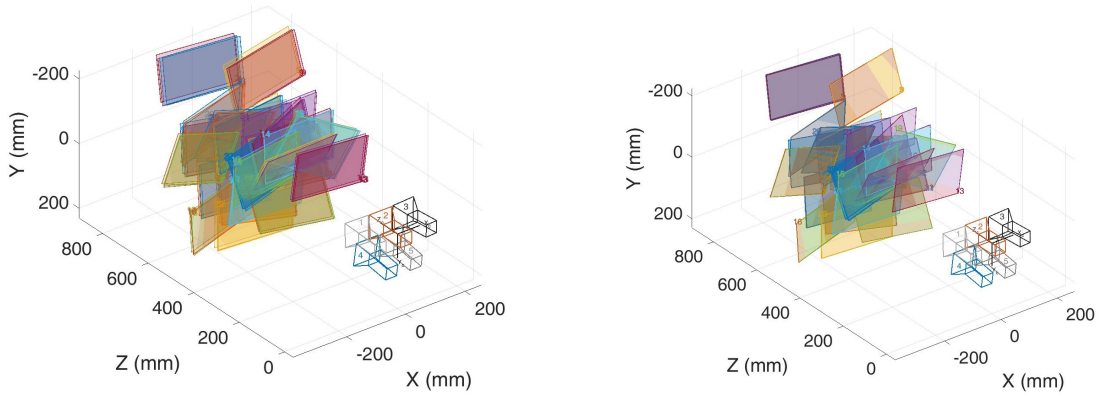
Upon visual comparison between datasets, it becomes evident that registering only the crucial segments of the images yields improved alignment of individual discrepancies between images, enhancing our ability to visually discern passengers. Nonetheless, an important question persists: whether such preprocessing methods would prove more conducive to training neural networks and ultimately enhancing training outcomes. The next chapter delves into addressing this inquiry by examining the obtained results.

IV. EXPERIMENTAL RESULTS

Following the successful acquisition and preprocessing of image data, our next step involves evaluating the presence of passengers in vehicles, regardless of whether they have tinted windows or not. In recent years, artificial neural networks have gained prominence in practical applications, revolutionising computer vision by enabling machines to interpret and comprehend visual information akin to the human brain. These networks, commonly known as convolutional neural networks (CNNs), are tailored specifically for image analysis. Various architectures such as ResNet, EfficientNet, or YOLO have showcased remarkable accuracy in detection tasks, rendering them indispensable tools across a spectrum of applications spanning from facial recognition and autonomous vehicles to medical image analysis and manufacturing quality control [47], [48].

Given the imperative of swift image processing in passenger detection, it is more suitable to employ one-stage detection models, which encompass models like SSD (Single Shot MultiBox Detector), RetinaNet, or YOLO [49], [50], [51].

The YOLO models are renowned for their real-time object detection prowess. Their single-stage detection process, which involves processing the entire image in a single forward pass, renders YOLO models faster in comparison



(a) using couples of intrinsic parameters with respect to a camera with NIR filter

(b) using fixed intrinsic parameters

FIGURE 13. **Visualisation of external camera parameters during camera calibration with respect to a camera with NIR filter.**

to certain two-stage detectors. This speed advantage proves crucial in applications necessitating real-time performance, as exemplified in our application.

A. ROI DETECTION

Our registration experiments have indicated that registering cropped images leads to superior image registration results. Consequently, it is advisable to first detect the windows on the vehicle, crop these areas, and then apply registration before detecting passengers themselves. Once again, we opted for YOLOv8 for the ROI detection task. We utilised freely available images from the Car Damage Detection Computer Vision Project dataset [52] to train the model. This dataset comprises images of damaged vehicles involved in accidents across 29 classes, including various types of damage and vehicle body parts (e.g., windshield, hood, roof, rear doors, etc.). We filtered out irrelevant classes, retaining only the four crucial classes for our purposes: windshield and rear windshield, front and back windows. From the original 3226 frames, we obtained 829 usable frames. Given the dataset's primary focus on vehicle damage, many images lacked correctly labelled vehicle body part classes, necessitating manual adjustments for our requirements. Additionally, we utilised images captured by our camera system for training purposes. Ultimately, our dataset comprised 1707 training images (after data augmentation through rotation, brightness, or saturation), 163 validation images, and 97 test images. A potential future enhancement involves incorporating images of other vehicle types such as trucks, buses, trains, and others. The distribution of individual classes is illustrated in Figure 14.

Training of the model took place on a personal computer equipped with an NVIDIA GeForce RTX 4060 Ti, 64 GB RAM, and an AMD Ryzen 9 7950X processor. The YOLOv8n model underwent training using transfer learning methodology, wherein we initialised the model with pre-trained weights on the COCO (Common Objects in Context) dataset and subsequently fine-tuned it on our dataset [53].

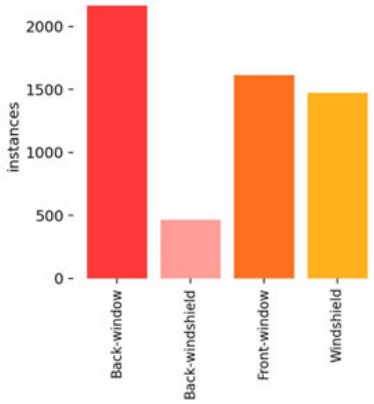


FIGURE 14. **The number of individual classes in the dataset.**

We employed predefined parameters and conducted model training for 350 epochs with a patience setting of 130. Training ceased after 180 epochs, as there was no improvement in loss over the last 130 epochs, thus preventing overfitting, and the best model from epoch 50 was saved. Table 3 delineates the configuration parameters of YOLO v8, while Figure 15 provides an overview of the trained model's overall results.

Based on the experimental findings, our model demonstrates the capability to detect vehicle parts with the following accuracies: 69.6% for the back window, 86.2% for the rear windshield, 88.1% for the front window, and 90.2% for the windshield, yielding a total mean Average Precision (mAP) of 67.9% across all four classes. Additionally, Figure 16 showcases the outcomes of body part detection using the trained model.

The trained model exhibits precise detection of crucial body parts. Furthermore, it demonstrates a good ability to generalize new data from standard cameras. However, for our specific objectives, it would be imperative to curate a

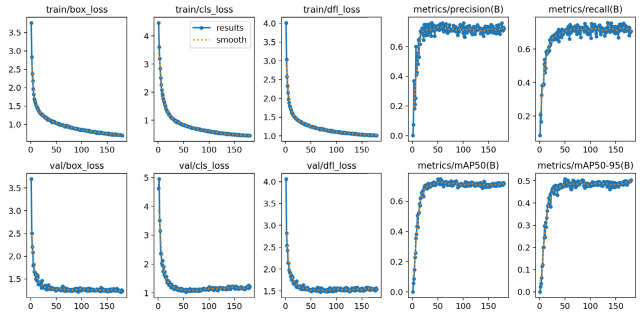


FIGURE 15. Overall results of the trained model for window detection.

FIGURE 16. ROI detection results.

significantly larger dataset to train a sufficiently robust model using images captured by our camera system. We anticipate that by assembling a diverse and extensive dataset and training exclusively on images from our camera system, we would attain markedly improved body part detection results.

TABLE 3. YOLOv8 configuration parameters for ROI detection.

Parameter	Values
Epoch	350
Learning rate	0.01
Image size	640
Batch size	16
Number of images	1967
Layers	168
Parameters	3,006,428

B. PASSENGER DETECTION

For passenger detection, we created thirty-one datasets with varying degrees of image preprocessing, subsequently comparing the learning outcomes of the YOLOv8n model across these datasets. These datasets can be categorised into the following groups: images without preprocessing, cropped images, cropped images preprocessed with histogram equalisation, registered images, and images with different combinations of image fusion, as elaborated in greater detail in previous chapters.

Each dataset comprised a partition of 729/119/52 images for training, testing, and validation, respectively, ensuring the prevention of data leakage, specifically train-test contamination. Figure 17 provides an example of images from the training set for one dataset.

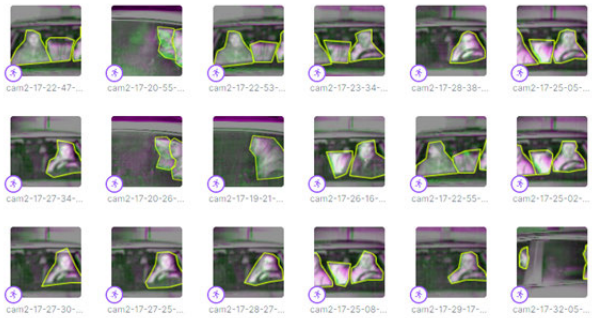


FIGURE 17. Sample images of the training set from dataset type e) in NIR and PL registration.

To ensure the most comprehensive comparison of learning results across individual datasets, we utilized tables for comparison. The metrics compared included the following:

1) Number of training epochs
2) Precision (P) - quantifying the proportion of true positive results among all positive predictions, evaluating the model's ability to avoid false positives.
3) Recall (R) - measuring the proportion of true positive results among all actual positives, assessing the model's ability to detect all instances of the desired class.
4) F1 score - representing the harmonic mean of precision and recall.
5) mAP50 - calculating the mean average accuracy at the intersection over union (IoU) with a threshold of 0.50, measuring the accuracy of the model for "easy" detections.
6) mAP50-95 - determining the mean of the average accuracy calculated at various IoU thresholds ranging from 0.50 to 0.95, offering a comprehensive view of the model's performance at different levels of detection difficulty.

These metrics are pivotal in assessing the precision and effectiveness of object detection models. They provide insights into the model's ability to accurately identify and locate objects in images while aiding in understanding its behaviour concerning false positives and false negatives. Such evaluations are essential for comprehensively evaluating the overall performance of object detection models.

In general, we advocate for utilizing both a validation set (a dataset employed to fine-tune hyperparameters) and a test set (a dataset used to assess the performance of a fully trained

model) to analyze the effectiveness of a model. However, for this specific application, our focus cannot solely be on achieving the highest precision or recall, as it is crucial to avoid instances where the model fabricates passengers or fails to detect them. Hence, the most suitable metric for comparison on the validation set would be mAP50.

On the other hand, if we maintain a neutral stance regarding false positives (FPs) and false negatives (FNs), it is recommended to utilise the F1 score for evaluating the best performing model. However, if false positives are deemed unacceptable to us, we should choose the model with higher precision. Conversely, if false negatives are unacceptable to us, we should prioritise selecting the model with higher recall.

Upon comparing the highest mAP50 values across validation sets among the datasets, we observe minimal differences. The lowest value is attained by models trained on images from a camera without any filter; however, this changes significantly when these images are registered with those from a camera with a NIR filter, resulting in the mAP50 metric surpassing 0.9. Notably, the maximum mAP50 value is achieved by images from dataset type e) (mutually registered images with the standard optimizer and the green-magenta fusion method) in a combination of images from NIR and ND filters. Surprisingly, the second-highest value is attained by images from dataset c) (registered with the standard optimizer and standard fusion method) in a combination of images from cameras without a filter and with a NIR filter. These findings are detailed in Table 10 and Table 4. The metrics with the highest achieved values are marked in bold.

However, upon comparing the test images, we observe other intriguing outcomes. The model trained on images from dataset a), featuring a camera equipped with a NIR filter, achieves the highest mAP50 value. Additionally, notable mAP50 values are observed in dataset type e), which combines images from NIR and ND filters, and dataset type f), which combines images from NIR and PL filters. These results are illustrated in Table 4, and detection results for one of the best dataset (dataset type e) in combination of NIR and ND) is depicted in Figure 18. It is evident that while the results during training are more than satisfactory, there is a noticeable deterioration in performance during testing. This decline in performance could be attributed to factors such as an inadequately varied and diverse dataset or insufficient image quantity. Additionally, concerning the F1 score metric, datasets with highest values are: dataset type a) utilising a camera with a polarising filter, dataset type c) employing NIR and PL filter registration, and dataset type e) utilising a combination of images from NIR and UV filters.

Based on the comparison of the mAP50 parameter across both the training and test sets, dataset type e) featuring the combination of images registered from cameras with NIR and ND filters emerges as the top performer along with dataset type f) featuring the registration of images from cameras with NIR and PL filters.

The results for the remaining datasets are presented in Tables 8- 13. Interestingly, a comparison of the number of epochs during training reveals notable variations. Certain combinations exhibit a considerably lower number of epochs compared to others. For instance, we achieved the lowest number of training epochs, 22, when training dataset type g), combining NIR and PL filters. Conversely, the dataset type c), combining cameras without a filter and with a NIR filter, required the highest number of epochs, reaching 139. It's worth noting that all parameters were consistent across training sessions, including the patience parameter, which determined the end of training and was set to 50 epochs.

TABLE 4. Results for datasets type e).

		P	R	mAP50	mAP50-95	F1
NIR + No filter	val	0,960	0,896	0,977	0,646	0,927
	test	0,685	0,677	0,741	0,388	0,681
NIR + ND	val	0,951	0,915	**0,979**	0,626	0,933
	test	0,826	0,701	**0,795**	0,415	**0,758**
NIR + UV	val	0,935	0,952	0,972	0,692	0,943
	test	0,838	0,592	0,718	0,356	0,694
NIR + PL	val	0,960	0,896	0,957	0,615	0,927
	test	0,773	0,556	0,653	0,290	0,647

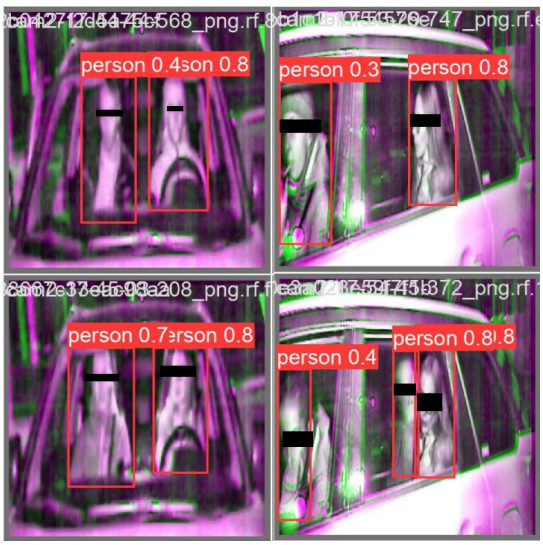


FIGURE 18. Passenger detection results from dataset e) with image registration from NIR and ND filter.

V. DISCUSSION

As evident from the detection results presented in the preceding chapters, it is apparent that the task of passenger detection in transportation systems, even when employing advanced computer vision techniques, is non-trivial and fraught with numerous challenges. In this chapter, we delve into a discussion of the attained results, outline the limitations of the system, propose potential enhancements, and explore prospective modifications to the system in the future.

At the outset, it's important to discuss the recommendations for constructing a dataset tailored for YOLO models. To attain optimal results, the authors suggest several prerequisites: a minimum of 1500 images for each class (in our scenario, representing a person class), 10,000 instances for each class, variation in shooting conditions (including different cameras, weather conditions, lighting scenarios, angles, and times of day), accurately tagged data, precise labeling, and inclusion of images devoid of the objects targeted for detection.

A study where the authors aimed to determine the number of occupants in vehicles using publicly accessible traffic camera images is presented in [10]. Their dataset comprised 568 images, partitioned into training, testing, and validation sets at ratios of 70%, 20%, and 10%, respectively. As part of their preprocessing steps, they converted images to grayscale and applied the CLAHE algorithm. They utilised the YOLOv3 model for training, attempting to detect two classes: passengers or empty seats. Their reported results were as follows: 37% false positive (FP) detections, 42% true positive (TP) detections, and 21% false negative (FN) detections. Notably, none of the 21 passenger examples in their validation set were correctly classified.

In comparison to one of our best results (dataset type e, comprising images registered from cameras with NIR and ND filters), our outcomes are as follows: 66% TP, 8% FP, and 26% FN detections. Consequently, our model achieves superior true positive and false positive detections, albeit exhibiting a slightly higher rate of false negative detections. From this comparison, it's evident that our system can provide more accurate estimations of passenger numbers in vehicles equipped with tinted and clear glass. However, it's worth noting that our results may be influenced by dataset imbalances, where there are considerably more instances of empty seats in dataset from [10], whereas in our datasets, the majority of seats are occupied. This bias in data might stem from real-world scenarios, as most passenger vehicle journeys typically involve only the driver without additional passengers [54]. Therefore, future iterations of our system development should aim to capture a diverse range of vehicle scenarios to mitigate this bias.

A. ALTERNATIVE OBJECT DETECTORS

In our attempt to achieve a thorough comprehension of object detection methodologies aimed specifically for passenger detection, we additionally integrated YOLOv9 and RetinaNet alongside our established YOLOv8 implementation to conduct a comparative analysis [49], [55]. By incorporating multiple state-of-the-art algorithms into our evaluation framework, we aimed to assess their performance in the context of passenger detection tasks.

Through this comparative analysis, we aimed to not only validate the effectiveness of our existing YOLOv8 implementation but also to explore paths for improvement and innovation. Through the strategic utilisation of the strengths inherent in various detection algorithms, our aim was to acquire an understanding of their capacity for generalisation and their appropriateness for practical applications in passenger detection within real-world contexts. This approach might help us to make informed decisions regarding algorithm selection, ultimately paving the way for the development of more reliable and efficient passenger detection systems.

1) YOLOv9

Following the introduction of YOLOv9 [55], which demonstrated exceptional model performance on the MS COCO 2017 [53] benchmark dataset, we resolved to conduct a comparative analysis between earlier used YOLOv8 and newest YOLOv9. The training parameters of YOLOv9 are shown in Tab. 5.Initially, we undergo a warm-up phase spanning 3 epochs, with the data augmentation configurations detailed in the bottom of Tab. 5. Experimental results of training with dataset e) with image registration from NIR and ND filter and utilised YOLOv9-c are shown in Fig. 19.

TABLE 5. YOLOv9 configuration parameters for passenger detection.

Parameter	Values
Epoch	160
Optimizer	SGD
Initial learning rate	0.01
Finish learning rate	0.01
Weight decay	0.0005
Image size	640
Batch size	4
Number of images	1967
Warm-up epochs	3
Warm-up momentum	0.8
Warm-up bias learning rate	0.1
Box loss gain	7.5
Class loss gain	0.5
HSV saturation augmentation	0.7
HSV value augmentation	0.4
Translation augmentation	0.1
Scale augmentation	0.9
Mosaic augmentation	1.0
MixUp augmentation	0.15
Copy & paste augmentation	0.3

Experimental results for YOLOv9 model are as follows: 66% TP, 10% FP, and 24% FN detections. This model achieves identical levels of true positive detections, with a slightly elevated rate of false positives but a reduced rate of false negative detections compared to previous version. This discrepancy may arise due to various factors such as the complexity of the dataset, the specific characteristics of the objects being detected, or the intricacies of the model architecture and training parameters.

FIGURE 19. Passenger detection results from dataset e) with image registration from NIR and ND filter with YOLOv9 model.

FIGURE 20. Passenger detection results from dataset e) with image registration from NIR and ND filter with RetinaNet model.

TABLE 6. RetinaNet hyperparameters for passenger detection.

Parameter	Values
Model backbone	ResNet34
Min size	800
Max size	1333
Optimiser	SGD
Learning rate	0.001
Weight decay	0.0005
Momentum	0.9
Horizontal flip	0.5
Shift scale rotate	0.5
Random brightness contrast	0.5

2) RetinaNet

RetinaNet is a popular object detection model introduced by Facebook AI Research group in [49]. It's known for its efficient and accurate detection of objects in images. One of its key features is the Focal Loss, which helps in addressing the class imbalance problem often encountered in object detection tasks. RetinaNet utilises a feature pyramid network (FPN) backbone coupled with a two-branch network for classification and bounding box regression, enabling it to detect objects at various scales in an image efficiently. RetinaNet employs anchor boxes for generating object proposals, resembling the approach of two-stage detection frameworks. However, it achieves a notable distinction by predicting object categories and locations through a single network, resulting in a more efficient inference process.

Similarly, we trained RetinaNet model to conduct a comparative analysis between previous models. Experimental results after training for 140 epoch are shown in Fig. 20 and hyperparameters used during training are shown in Tab. 6.

Experimental results for RetinaNet model are as follows: 61% TP, 9% FP, and 30% FN detections. RetinaNet achieves reduced levels of true positive detections compared to YOLOv8 and YOLOv9, with a similar rate of false positives as YOLOv8 but an elevated rate of false negative detections compared to YOLO models.

A direct comparison of F1 score, mAP50 and other metrics are displayed in Tab. 7. The results from these models are very close, making it challenging to draw distinct comparisons between them. The close similarity in results can be attributed to several factors, including the similarity in architecture, training data, hyperparameters, and optimisation techniques employed across the models, resulting in marginal differences in performance metrics. Additionally, the dataset used for evaluation might not effectively challenge the unique strengths of each model, contributing to the observed similarity in performance. Moreover, subtle variations in implementation details or random initialisation of parameters could further blur distinctions between the models' outcomes.

Further verification and training to address these discrepancies can be resource-intensive, often requiring significant computational resources. However, due to such constraints, our ability to conduct extensive validation and refinement may be limited. Nonetheless, we might explore these discrepancies and potential solutions in detail in future research attempts. Additional insights and findings might be reported in the future, providing a more comprehensive understanding of the model's performance and potential directions for improvement.

B. LIMITATIONS OF THE SYSTEM

Like many systems, whether technological, social, economic, or natural, our proposed system comes with its own set of disadvantages or limitations. One of the primary drawbacks is its susceptibility to changes in weather and lighting conditions. While the filters were intended to compensate for or eliminate lighting variations, experimental data revealed

TABLE 7. Comparison of different models.

Metric	YOLOv8	YOLOv9	RetinaNet
F1 score	0.796	0.793	0.758
mAP50	0.795	0.793	0.59
TP	66.1%	65.7%	61%
FP	8%	10%	8.5%
FN	25.8%	24.2%	30.4%

their limitations in mitigating glare from ambient or other light sources. Additionally, adverse weather phenomena like heavy rain, snow, fog, or frost on windows can adversely impact camera system performance, leading to inaccuracies in passenger counts and reduced system reliability under specific weather conditions.

Another limitation stems from the system's broad applicability across various industries with distinct configurations, necessitating adjustments and configurations tailored to each use case. Furthermore, privacy invasion is a significant concern, as continuous monitoring of individuals in vehicles raises privacy rights issues. Individuals may feel uncomfortable with the notion of being monitored, potentially leading to acts of vandalism or sabotage aimed at disabling or damaging the cameras, thereby compromising system functionality.

Gender or race bias represents another potential challenge. Poorly designed camera systems may inadvertently exhibit biases, resulting in inaccuracies in passenger counts, especially in recognising individuals of diverse racial or gender demographics. This could lead to skewed data and possible discrimination. Moreover, the current experiments primarily involved European participants, suggesting potential biases in detecting individuals with darker skin tones, resulting in fewer detected persons.

Finally, the costs associated with implementation and maintenance are significant considerations. Deploying and maintaining a camera system on roads entail substantial expenses for cleaning and upkeep to ensure the system operates at optimal efficiency.

C. POSSIBLE IMPROVEMENTS AND APPLICATION AREAS

There are several avenues for enhancing our developed system. While our experiments primarily focused on detecting passengers in cars, the techniques we proposed can also be adapted for use in public transportation (buses, trains, etc.) and freight transport. Additionally, consideration should be given to motorcyclists and drivers of other vehicles during warmer months as they are vulnerable road users [56], [57].

A significant potential improvement lies in the selection of cameras. To capture fast-moving vehicles, high-speed cameras would be necessary, albeit at a considerable cost. Our current system, while cost-effective, may not be suitable for capturing such vehicles.

In its current state, the system is well-suited for deployment in areas with dim or consistent lighting conditions, such as tunnels, border crossings during specific hours, and parking garages. The stringent lighting specifications typically found in tunnels should pose no significant obstacles to the implementation of our camera system.

Another promising application area is the detection of individuals in buildings.In the event of emergencies like fires, our camera system could non-invasively capture the situation in buildings equipped with solar tints by employing a camera system mounted on a drone or other mobile robotic system, leveraging the rapidly advancing technology of mobile robotics [58], [59].

VI. CONCLUSION

The importance of passenger detection in modern transportation cannot be overstated. It not only enhances passenger safety but also contributes to efficiency and comfort across various transport systems. By potentially reducing accidents, optimizing resource allocation, and enhancing overall passenger experience, passenger detection emerges as a critical component of contemporary transportation infrastructure. As technology continues to evolve, it is vital for stakeholders within the transportation industry to embrace and invest in these innovations to foster safer, more efficient, and more enjoyable travel experiences for all.

This study addresses the challenge of detecting individuals within vehicles equipped with tinted windows, presenting a significant optical obstacle to visual detection and monitoring

TABLE 8. Results for datasets type a).

		P	R	mAP50	mAP50-95	F1
No filter	val	0,557	0,458	0,520	0,219	0,503
	test	0,061	0,312	0,051	0,011	0,102
NIR	val	0,901	0,841	0,933	0,578	0,870
	test	0,865	0,744	0,843	0,408	0,800
ND	val	0,977	0,879	0,940	0,577	0,925
	test	0,611	0,537	0,311	0,190	0,572
UV	val	0,989	0,859	0,930	0,600	0,919
	test	0,707	0,663	0,715	0,337	0,684
PL	val	0,989	0,935	0,948	0,495	0,961
	test	0,764	0,695	0,764	0,324	0,728

TABLE 9. Results for datasets type b).

		P	R	mAP50	mAP50-95	F1
No filter	val	0,562	0,453	0,483	0,149	0,502
	test	0,035	0,094	0,015	0,005	0,051
NIR	val	0,949	0,925	**0,966**	0,669	0,937
	test	0,713	0,631	0,672	0,341	0,669
ND	val	0,903	0,890	0,942	0,541	0,896
	test	0,760	0,657	0,671	0,250	0,705
UV	val	0,901	0,877	0,930	0,617	0,889
	test	0,822	0,760	0,744	0,343	**0,790**
PL	val	0,989	0,866	0,940	0,592	0,923
	test	0,783	0,693	**0,749**	0,347	0,735

of interior activities in a non-invasive manner. Introducing a novel approach to image data synthesis utilising five cameras for image acquisition, the detection outcomes hold promise for diverse applications across various sectors of our society, including border surveillance, access control in secured facilities, tunnel monitoring, parking facilities, and beyond.

The achieved results, particularly in our optimized configuration (dataset e), where images from cameras with NIR and ND filters are combined, demonstrate notably favourable outcomes. With a 66% true positive rate and a low false positive rate of only 8%, these findings underscore the efficacy of our approach. Although there is room for improvement, these results indicate significant promise for the application of our system in practical scenarios.

TABLE 10. Results for datasets type c).

		P	R	mAP50	mAP50-95	F1
NIR + No filter	val	0,968	0,943	**0,977**	0,666	0,955
	test	0,747	0,688	0,758	0,411	0,716
NIR + ND	val	0,952	0,937	0,967	0,548	0,944
	test	0,746	0,649	0,719	0,250	0,694
NIR + UV	val	0,933	0,925	0,974	0,665	0,929
	test	0,727	0,717	0,759	0,384	0,722
NIR + PL	val	0,990	0,921	0,974	0,664	0,954
	test	0,804	0,671	0,773	0,954	0,732

TABLE 11. Results for datasets type d).

		P	R	mAP50	mAP50-95	F1
No filter	val	0,310	0,472	0,304	0,107	0,374
	test	0,048	0,335	0,038	0,009	0,084
NIR	val	0,956	0,925	**0,958**	0,638	0,940
	test	0,622	0,565	0,593	0,295	0,592
ND	val	0,968	0,847	0,936	0,511	0,903
	test	0,782	0,614	0,681	0,300	0,688
UV	val	0,966	0,830	0,940	0,633	0,893
	test	0,758	0,606	0,719	0,283	0,674
PL	val	0,989	0,838	0,955	0,621	0,907
	test	0,760	0,734	0,762	0,353	**0,747**

TABLE 12. Results for datasets type f).

		P	R	mAP50	mAP50-95	F1
NIR + No filter	val	0,908	0,925	0,959	0,592	0,916
	test	0,727	0,565	0,678	0,283	0,636
NIR + ND	val	0,926	0,877	0,948	0,549	0,901
	test	0,719	0,662	0,724	0,399	0,689
NIR + UV	val	0,930	0,925	0,962	0,643	0,927
	test	0,802	0,613	0,727	0,342	0,695
NIR + PL	val	0,908	0,915	**0,966**	0,656	0,911
	test	0,775	0,758	**0,781**	0,389	**0,766**

APPENDIX SUPPLEMENTARY EXPERIMENTAL FINDINGS

See Tables 8–13.

TABLE 13. Results for datasets type g).

		P	R	mAP50	mAP50-95	F1
NIR + No filter	val	0,971	0,887	0,940	0,585	0,927
	test	0,755	0,637	0,727	0,356	0,691
NIR + ND	val	0,979	0,890	0,941	0,558	0,932
	test	0,719	0,680	**0,738**	0,410	**0,699**
NIR + UV	val	0,947	0,844	0,949	0,543	0,893
	test	0,746	0,648	0,687	0,311	0,694
NIR + PL	val	0,960	0,912	**0,962**	0,581	0,935
	test	0,817	0,550	0,711	0,382	0,657

ACKNOWLEDGMENT

The authors wish to express their profound gratitude to Tomáš Mizera for his invaluable assistance and unwavering support in facilitating the acquisition of spectral transmittance data across diverse specimens. They would also like to express their deepest gratitude for generous funding and support.

REFERENCES

[1] B. Xu, P. Paul, Y. Artan, and F. Perronnin, "A machine learning approach to vehicle occupancy detection," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Qingdao, China, Oct. 2014, pp. 1232–1237.

[2] L. Chu, K. Nesamani, and H. Benouar, "Priority based high occupancy vehicle lanes operation," in *Proc. 86th Annu. Meeting Transp. Res. Board*. Washington, DC, USA: Transportation Research Board, 2007, pp. 1–26.

[3] T. Fontes, P. Fernandes, H. Rodrigues, J. M. Bandeira, S. R. Pereira, A. J. Khattak, and M. C. Coelho, "Are HOV/eco-lanes a sustainable option to reducing emissions in a medium-sized European city?" *Transp. Res. A, Policy Pract.*, vol. 63, pp. 93–106, May 2014.

[4] A. Krasnov and L. Uzai, "Optical considerations for automotive windshields with improved thermal performance," *Opt. Mater.*, vol. 139, May 2023, Art. no. 113807.

[5] R. Levinson, H. Pan, G. Ban-Weiss, P. Rosado, R. Paolini, and H. Akbari, "Potential benefits of solar reflective car shells: Cooler cabins, fuel savings and emission reductions," *Appl. Energy*, vol. 88, no. 12, pp. 4343–4357, Dec. 2011.

[6] N. Gure and M. Yilmaz, "Car window filming, tinting and shading's fuel, emission reduction and economic analysis around WA, NY, NC, USA and Istanbul, Turkey," *J. Energy Power Sources*, vol. 2, no. 1, pp. 6–21, Jan. 2015.

[7] P. Kuchár, R. Pirník, A. Janota, B. Malobický, J. Kubík, and D. Šišmišová, "Passenger occupancy estimation in vehicles: A review of current methods and research challenges," *Sustainability*, vol. 15, no. 2, p. 1332, Jan. 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/2/1332

[8] A. Satz, D. Hammerschmidt, and D. Tumpold, "Capacitive passenger detection utilizing dielectric dispersion in human tissues," *Sens. Actuators A, Phys.*, vol. 152, no. 1, pp. 1–4, May 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924424709001253

[9] P. Kuchár, R. Pirník, T. Tichý, K. Rástočný, M. Skuba, and T. Tettamanti, "Noninvasive passenger detection comparison using thermal imager and IP cameras," *Sustainability*, vol. 13, no. 22, p. 12928, Nov. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/22/12928

[10] L. Branco, F. Qiao, and Y. Zhang, "Detecting number of passengers in a moving vehicle with publicly available data," in *Proc. Intell. Syst. Conf. (IntelliSys)*, vol. 2. Cham, Switzerland: Springer, 2021, pp. 536–548.

[11] A. Rahmatulloh, F. M. S. Nursuwars, I. Darmawan, and G. Febrizki, "Applied Internet of Things (IoT): The prototype bus passenger monitoring system using PIR sensor," in *Proc. 8th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Yogyakarta, Indonesia, Jun. 2020, pp. 1–6.

[12] I. J. Amin, A. J. Taylor, F. Junejo, A. Al-Habaibeh, and R. M. Parkin, "Automated people-counting by using low-resolution infrared and visual cameras," *Measurement*, vol. 41, no. 6, pp. 589–599, Jul. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224107001078

[13] D. Klauser, G. Bärwolff, and H. Schwandt, "A TOF-based automatic passenger counting approach in public transportation systems," *AIP Conf. Proc.*, vol. 1648, Mar. 2015, Art. no. 850113.

[14] S. Gautama, S. Lacroix, and M. Devy, "Evaluation of stereo matching algorithms for occupant detection," in *Proc. Int. Workshop Recognit., Anal., Tracking Faces Gestures Real-Time Systems. Conjunct (ICCV)*, Corfu, Greece, Sep. 1999, pp. 177–184.

[15] P. Faber, "Seat occupation detection inside vehicles," in *Proc. 4th IEEE Southwest Symp. Image Anal. Interpretation*, Austin, TX, USA, Apr. 2000, pp. 187–191.

[16] P. Faber, "Image-based passenger detection and localization inside vehicles," *Int. Arch. Photogramm. Remote Sens.*, vol. 33, no. B5, pp. 230–237, 2000.

[17] M. Devy, A. Giralt, and A. Marin-Hernandez, "Detection and classification of passenger seat occupancy using stereovision," in *Proc. IEEE Intell. Vehicles Symp.*, Dearborn, MI, USA, Oct. 2000, pp. 714–719.

[18] T. D. Schoenmackers and M. M. Trivedi, "Real-time stereo-based vehicle occupant posture determination for intelligent airbag deployment," in *Proc. IEEE IV Intell. Vehicles Symp.*, Columbus, OH, USA, Jun. 2003, pp. 570–574.

[19] S. Wender and O. Loehlein, "A cascade detector approach applied to vehicle occupant monitoring with an omni directional camera," in *Proc. IEEE Intell. Vehicles Symp.*, Parma, Italy, Jun. 2004, pp. 345–350.

[20] A. Géczy, R. De Jorge Melgar, A. Bonyár, and G. Harsányi, "Passenger detection in cars with small form-factor IR sensors (Grid-eye)," in *Proc. IEEE 8th Electron. System-Integration Technol. Conf. (ESTC)*, Tønsberg, Norway, Sep. 2020, pp. 1–6.

[21] F. Zhu, J. Song, R. Yang, and J. Gu, "Research on counting method of bus passenger flow based on kinematics of human body," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, Wuhan, China, 2008, pp. 201–204.

[22] F. Zhu, J. Gu, R. Yang, and Z. Zhao, "Research on counting method of bus passenger flow based on kinematics of human body and SVM," in *Proc. 2nd Int. Symp. Intell. Inf. Technol. Appl.*, Shanghai, China, Dec. 2008, pp. 14–18.

[23] Y. Artan, O. Bulan, R. P. Loce, and P. Paul, "Passenger compartment violation detection in HOV/HOT lanes," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 395–405, Feb. 2016.

[24] S. Wshah, B. Xu, O. Bulan, J. Kumar, and P. Paul, "Deep learning architectures for domain adaptation in HOV/HOT lane enforcement," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–7.

[25] P. M. Birch, "Automated vehicle occupancy monitoring," *Opt. Eng.*, vol. 43, no. 8, pp. 1828–1832, Aug. 2004, doi: 10.1117/1.1766300.

[26] Y. Gu and R. O. Sinnott, "Real-time vehicle passenger detection through deep learning," in *Proc. IEEE 19th Int. Conf. e-Science (e-Science)*, Oct. 2023, pp. 1–10.

[27] J. R. Tyrer and L. M. Lobo, "An optical method for automated roadside detection and counting of vehicle occupants," *Proc. Inst. Mech. Eng. D, J. Automobile Eng.*, vol. 222, no. 5, pp. 765–774, May 2008, doi: 10.1243/09544070jauto562.

[28] A. J. Pérez-Jiménez, J. L. Guardiola, and J. C. Pérez-Cortés, "High occupancy vehicle detection," in *Structural, Syntactic, and Statistical Pattern Recognition*, N. da Vitoria Lobo, T. Kasparis, F. Roli, J. T. Kwok, M. Georgiopoulos, G. C. Anagnostopoulos, and M. Loog, Eds., Berlin, Germany: Springer, 1007, pp. 782–789.

[29] X. Yuan, Y. Meng, and X. Wei, "A method of location the vehicle windshield region for vehicle occupant detection system," in *Proc. IEEE 11th Int. Conf. Signal Process.*, Oct. 2012, pp. 712–715.

[30] D. Cornett, A. Yen, G. Nayola, D. Montez, C. R. Johnson, S. T. Baird, H. Santos-Villalobos, and D. S. Bolme, "Through the windshield driver recognition," *Electron. Imag.*, vol. 31, no. 13, pp. 140-1–140-9, Jan. 2019.

[31] M. Ruby, D. S. Bolme, J. Brogan, D. Cornett III, B. Delgado, G. Jager, C. Johnson, J. Martinez-Mendoza, H. Santos-Villalobos, and N. Srinivas, "The Mertens unrolled network (MU-Net): A high dynamic range fusion neural network for through the windshield driver recognition," in *Auton. Syst., Sensors, Process., Security Vehicles Infrastruct.*, vol. 11415, M. C. Dudzik and S. M. Jameson, Eds., Bellingham, WA, USA: SPIE, 1117, pp. 72–83. [Online]. Available: https://doi.org/10.1117/12.2566765

[32] S. Wasista, "Webcam-based bus passenger detection system using single shot detector method," *IJCCS Indonesian J. Comput. Cybern. Syst.*, vol. 18, no. 1, p. 37, Jan. 2024.

[33] X. Li, Y. Wu, Y. Fu, L. Zhang, and R. Hong, "A lightweight bus passenger detection model based on YOLOv5," *IET Image Process.*, vol. 17, no. 14, pp. 3927–3937, Dec. 2023.

[34] C. Pronello and X. R. Garzón Ruiz, "Evaluating the performance of video-based automated passenger counting systems in real-world conditions: A comparative study," *Sensors*, vol. 23, no. 18, p. 7719, Sep. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/18/7719

[35] R. Morozov, V. Nicheporchuk, and I. Perevalov, "Prototype of urban transport passenger accounting system," *Transp. Res. Proc.*, vol. 68, pp. 468–474, Jan. 2023.

[36] E. Hyun, Y. Jin, J. Bae, and P. Chi-Ho, "Machine learning based in-cabin radar system for passenger monitoring system," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2023, pp. 1–4.

[37] P. Kuchár, R. Pirník, J. Ďurišová, M. Skuba, T. Mizera, and J. Kafková, "Effect of window tinting on passenger detection and enforcement in road transport," *Transp. Res. Proc.*, vol. 74, pp. 938–945, Jan. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352146523005252

[38] J. LaMotte, W. Ridder, K. Yeung, and P. De Land, "Effect of aftermarket automobile window tinting films on driver vision," *Human Factors, J. Human Factors Ergonom. Soc.*, vol. 42, no. 2, pp. 327–336, Jun. 2000.

[39] V. Thambawita, I. Strümke, S. A. Hicks, P. Halvorsen, S. Parasa, and M. A. Riegler, "Impact of image resolution on deep learning performance in endoscopy image classification: An experimental study using a large dataset of endoscopic images," *Diagnostics*, vol. 11, no. 12, p. 2183, Nov. 2021. [Online]. Available: https://www.mdpi.com/2075-4418/11/12/2183

[40] J. Lee, J. Byun, J. Lim, and J. Lee, "A framework for detecting vehicle occupancy based on the occupant labeling method," *J. Adv. Transp.*, vol. 2020, pp. 1–8, Dec. 2020.

[41] X. Hao, H. Chen, Y. Yang, C. Yao, H. Yang, and N. Yang, "Occupant detection through near-infrared imaging," *J. Appl. Sci. Eng.*, vol. 14, pp. 275–283, Sep. 2011, doi: 10.6180/jase.2011.14.3.11.

[42] G. T. S. Ho, Y. P. Tsang, C. H. Wu, W. H. Wong, and K. L. Choy, "A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities," *Sensors*, vol. 19, no. 8, p. 1796, Apr. 2019.

[43] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 1106–1112.

[44] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 965–980, Oct. 1992.

[45] G. D. Hines, Z.-U. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-image registration for an enhanced vision system," *Proc. SPIE*, vol. 5108, pp. 231–241, Aug. 2003.

[46] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*. San Diego, CA, USA: Academic, Aug. 1994, p. 514.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

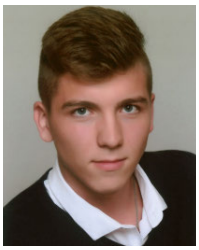
[48] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[49] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[50] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.

[51] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLOv8*. [Online]. Available: https://github.com/ultralytics/ultralytics

[52] CAPSTONE. (Aug. 2023). *Car Damage Detection Dataset*. [Online]. Available: https://universe.roboflow.com/capstone-nh0nc/car-damage-detection-t0g92
[53] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland. Cham, Switzerland: Springer, Sep. 2014, pp. 740–755.
[54] F. Benita, "Carpool to work: Determinants at the county-level in the United States," *J. Transp. Geography*, vol. 87, Jul. 2020, Art. no. 102791.
[55] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.
[56] J. E. Espinosa, S. A. Velastín, and J. W. Branch, "Detection of motorcycles in urban traffic using video analysis: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6115–6130, Oct. 2021.
[57] N. K. Kanhere, S. T. Birchfield, W. A. Sarasua, and S. Khoeini, "Traffic monitoring of motorcycles during special events using video detection," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2160, no. 1, pp. 69–76, Jan. 2010.
[58] M. Mihálik, M. Hruboš, P. Vestenický, P. Holeč ko, D. Nemec, B. Malobický, and J. Mihálik, "A method for detecting dynamic objects using 2D LiDAR based on scan matching," *Appl. Sci.*, vol. 12, no. 11, p. 5641, Jun. 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/11/5641
[59] M. Mihálik, B. Malobický, P. Peniak, and P. Vestenický, "The new method of active SLAM for mapping using LiDAR," *Electronics*, vol. 11, no. 7, p. 1082, Mar. 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/7/1082



PAVOL KUCHÁR received the B.S. degree in automation and the M.S. degree in process control from the University of Žilina, Slovakia, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the Department of Control and Information Systems. His research interests include machine learning, telematics, and transportation systems, with a primary emphasis on passenger detection across diverse transportation environments.

RASTISLAV PIRNÍK was born in Humenné, Slovakia, in 1978. He received the M.S. and Ph.D. degrees in telecommunication engineering from the Faculty of Electrical Engineering, University of Žilina, in 2001 and 2010, respectively. He is currently an Associate Professor with the Department of Control and Information Systems, University of Žilina. He has been involved in numerous research projects, particularly those supported by the APVV and VEGA agencies. In addition, he has served as a project manager for two projects funded by the ASFEU agency and has contributed to the National Traffic Information System project. His research interests include data transmission in transport telematic systems and legislative matters.

JÚLIA KAFKOVÁ received the B.S. and M.S. degrees in biomedical engineering from the University of Žilina, Slovakia, in 2021 and 2023, respectively, where she is currently pursuing the Ph.D. degree with the Department of Control and Information Systems. Her research interests include machine learning, signal processing, and electronics, with the main focus on ECG.

TOMÁŠ TICHÝ was born in Prague, in 1974. He received the degree from the Faculty of Transportation Sciences, Czech Technical University in Prague, in 2000, the Ph.D. degree in the engineering informatics, in 2004, and the degree from the University of Economics in Prague, in 2017. In 2009, he was appointed as an Associate Professor in engineering informatics (transport and telecommunications). He is currently the Head of the Laboratory of Traffic Control and Modelling, Faculty of Transportation Sciences, Czech Technical University in Prague. He teaches courses on ITS control systems, system safety and reliability, tunnel systems, and transport telematics, as a Lecturer at Czech Technical University in Prague. Since 2000, he has been with the ELTODO Group in various positions, from a Traffic Engineer to the Head of the Traffic Management Department. He has experience in the implementation and design of ITS systems, traffic control algorithms, tunnel systems, and public lighting. He is also the Director of the Strategic Development and is responsible for the development of the entire ELTODO Group (Energy, Transport, ICT, Lighting, and Industry), solves grant projects, and collaborates with the private and academic sectors in research and development. He is responsible for ISO and TQM in the ELTODO Group and manages as many as 50 employees with a turnover of more than CZK 50 million. He has solved science and research projects.

JANA ĎURIŠOVÁ was born in Žilina, in 1979. She received the degree from the Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, in 2003, specializing in physics–optics and optoelectronics, and the Ph.D. degree in measuring techniques from the Slovak University of Technology in Bratislava. Since 2011, she has been teaching courses in physics, mechanics, electricity and magnetism, optics, and applied optics, at the University of Žilina. She has extensive experience in photonics and biophotonics, with expertise in atomic force microscopy and AFM lithography, laser lithography, nanostructures and nanocrystals, photonic crystals, measurements of optical properties, LOC devices, optical tweezers, interferometry, sensors, near-field investigation, and other related areas.

MICHAL SKUBA received the bachelor's degree in automation engineering and the master's degree in process control from the University of Žilina, Slovakia, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the Department of Control and Information Systems. His research interests include machine learning, telematics, and transportation systems, with a focus on predominantly on smart traffic control, particularly in the domain of traffic light optimization.