

RESEARCH

Open Access



Architecture for determining the cleanliness in shared vehicles using an integrated machine vision and indoor air quality-monitoring system

Nilusha Jayawickrama^{1*}, Enric Perarnau Ollé², Jesse Pirhonen¹, Risto Ojala¹, Klaus Kivekäs^{1,3}, Jari Vepsäläinen¹ and Kari Tammi¹

*Correspondence:
nilusha.jayawickrama@aalto.fi

¹ Department of Mechanical Engineering, Aalto University, Espoo, Finland

² Department in Mechanical Engineering, Laboratory of Microsystems and Nanotechnology, Polytechnical University of Catalonia, Barcelona, Spain

³ Faculty of Agriculture and Forestry, Helsinki University, Helsinki, Finland

Abstract

In an attempt to mitigate emissions and road traffic, a significant interest has been recently noted in expanding the use of shared vehicles to replace private modes of transport. However, one outstanding issue has been the hesitancy of passengers to use shared vehicles due to the substandard levels of interior cleanliness, as a result of leftover items from previous users. The current research focuses on developing a novel prediction model using computer vision capable of detecting various types of trash and valuables from a vehicle interior in a timely manner to enhance ambience and passenger comfort. The interior state is captured by a stationary wide-angled camera unit located above the seating area. The acquired images are preprocessed to remove unwanted areas and subjected to a convolutional neural network (CNN) capable of predicting the type and location of leftover items. The algorithm was validated using data collected from two research vehicles under varying conditions of light and shadow levels. The experiments yielded an accuracy of 89% over distinct classes of leftover items and an accuracy of 91% among the general classes of trash and valuables. The average execution time was 65 s from image acquisition in the vehicle to displaying the results in a remote server. A custom dataset of 1379 raw images was also made publicly available for future development work. Additionally, an indoor air quality (IAQ) unit capable of detecting specific air pollutants inside the vehicle was implemented. Based on the pilots conducted for air quality monitoring within the vehicle cabin, an IAQ index was derived which corresponded to a 6-level scale in which each level was associated with the explicit state of interior odour. Future work will focus on integrating the two systems (item detection and air quality monitoring) explicitly to produce a discrete level of cleanliness. The current dataset will also be expanded by collecting data from real shared vehicles in operation.

Keywords: Prediction model, Computer vision, Convolutional neural network, Indoor air quality, Shared vehicles, Interior cleanliness

Introduction

The use of shared vehicles has notably increased over the recent past. In terms of market expansion, the estimated number of carsharing users is expected to reach 36 million by the year 2025 [1]. The annual growth rate is estimated to be at 45% by 2025 [2]. Although the increased use of shared vehicles will result in a sedate growth of vehicle sales, this will also provide highly valued opportunities for automobile manufacturers, suppliers as well as mobility services. Consequently, the growth of global vehicle-sales, although slower, will not be reversed [3]. Based on the statistics in [4], it was evident that within Europe, carsharing has comprised a total of up to 2 million users sharing approximately 26,000 vehicles. Carsharing also paves the way for current issues with respect to changes in consumer expectations as shared cars would provide users with effective choices to enhance the accessibility of vehicles as well as improve the reliability and comfort of travel [5].

To ensure continuous growth of carsharing, it is important to ensure that vehicles subjected to this service maintain elevated levels of interior cleanliness to maximise comfort for the travellers. A lack of maintenance witnessed the demise of Autolib's carsharing service which operated in the French capital between 2011 and 2018 [6]. Moreover, research conducted in Berlin [7] suggests that car sharing operators lack information on the condition of their fleet during operation. Although the users in this case had the option of rating the cleanliness of the vehicles, it does not prevent the exposure of dirty vehicle interiors to immediate new users. Negative customer reviews were generated with pictures of unclean interiors which mostly contained cigarette butts, food wrappings and other types of waste. Additionally, there had also been complaints about losing certain types of belongings as well as finding those of someone else. Overall, 38% of vehicle interiors within the city of Berlin were deemed to be unpleasant, messy or very dirty [7]. The importance of cleanliness in carsharing is further highlighted under the current Covid pandemic situation with users urged to always remove dirt and dust after use [8].

Moving on to the main research contributions in this paper, we propose a novel vision-based architecture to predict the interior cleanliness state of a vehicle cabin. The method is intended for shared vehicles but can also be applied for other passenger vehicles. Thus, the system proposed aims on ensuring the elimination of leftover passenger items, including trash and valuables, in a timely manner. Specifically, two modular sensor units will be designed and installed within designated locations of the vehicle interior: a camera module to capture images of the rear seating area and an indoor air quality monitoring (IAQ) unit. A vision-based prediction model (in the form of a EfficientDet convolutional neural network(CNN) [9]) will be developed to identify distinct types of leftover items in the seating area from each image received from the camera unit. Moreover, a dataset of custom images containing leftover items within vehicle interiors will be made available for public use. In the scope of the current study, we will not consider scenarios with dark external settings such as nighttime operation. The IAQ unit will determine the concentrations of specific air pollutants within the vehicle cabin and generate a corresponding air quality index. Zero detections from the vision-based prediction model and minimum yields of air pollutant concentrations from the IAQ unit will, together, correspond to the cleanest state of the vehicle interior. Note however that the current

research article has a higher weight-age (from the implementation and analysis points of view) on the vision-based detection system over the IAQ unit.

Furthermore, to ensure secure data transmission and communication, a suitable communication architecture will be implemented and tested to validate an efficient flow of information and display of the predicted results. In particular, a centralised server comprising the prediction model will be expected to link all shared cars within the system in terms of storage and execution.

State-of-the-art

Vision-based detection and prediction

To the best of our knowledge, the computer vision related work carried out here is the sole initiative of the use of a camera-based unit in shared vehicles to monitor the state of a car interior for the purposes of cleanliness. However, similar systems have been utilised for other purposes within vehicles such as driver and occupant-monitoring as evident in [10, 11].

With respect to vision-based solutions, it is important to initially create an image database when dealing with special purpose applications. For the current study, existing image databases including Trashnet [12] and Kaggle [13] could be utilised. The first database of the forementioned options correspond to trash items while the latter comprises images of consumer belongings (valuables).

Detection and classification of trash using computer vision has been studied in a multitude of previous research applications. Thung and Yang presented a fine-tuned CNN for the classification of garbage with respect to their Trashnet dataset [12]. The classification accuracy of their prediction model was 22% as a result of suboptimal hyperparameter training. Tharani et al. [14] presented a method of detecting trash items which could be visually identified on water surfaces. Their study proposed the use of a novel attention layer in conjunction with a series of state-of-the-art neural network architectures. The results indicated that the YOLO-v3 [15] model outperforms the remaining networks which were analysed with an average precision of 48.1% on their dataset. Liu and Jiang [16] presented an article which focused on autonomous identification and classification of commonly found garbage types. Their system consisted of a Raspberry Pi which comprised their prediction model in the form of a convolutional neural network (CNN). The paper presented the analysis of a total of four trained pre-built CNN architectures and concluded that the highest performance was shown by the Vgg16 [17] model with an accuracy of 74%. A similar approach of using a CNN was presented in the study conducted by Adedeji and Wang [18] in which a pretrained ResNet-50 [19] was used to develop a waste material classification system. The model was trained with the trash image database in [12] resulting in it successfully yielding a final test accuracy of 87% over four different categories of trash in the dataset. Zhihong et al. [20] also studied and demonstrated the use of CNNs to enable autonomous garbage sorting via a robotic grasping mechanism. Their study presented the application of a Vgg16 model for image classification and Region Proposed Generation (RPN) [21] for object detection. The results from the study recorded an average model execution time of 220 ms with missed and false detection rates as 3% and 9%, respectively.

Monitoring the explicit cleanliness with the utilisation of computer vision has also been studied in previous research applications. Rad et al. [22] presented the application of CNNs to locate and identify various types of trash items on the streets. The system presented in their study was tested and validated by attaching a camera unit to a street sweeping machine to gather data which would subsequently be used to train the prediction models. The qualitative assessment of their concept presented results of false positive and missed detections of small trash sources such as leaves or, in some cases, cigarette butts, but additionally highlighted that their system had the capability of detecting multiple and overlapping trash items. A cleanliness analysis was also utilised in the studies conducted by Alfarrarjeh et al. [23] and Ghildiyal et al. [24]. The first research in [23] explored the application of the Caffe architecture in the form of a CNN which produced an F1 score [25] of 0.78. The latter research (in [24]) was comparatively a lower scale model which however once again utilised CNNs, but this time in the form of Vgg19 [17] and Inception-v3 [26] models. The models were tested to analyse their capability of categorising an image scene into one of three possible outcomes: poor, average, and good. The study concluded that the use of CNNs with properly tuned parameters for each feature could perform general classifications in such custom applications, but with noticeable misclassifications. Contrary to the approach of detection, Jayasinghe et al. [27] developed a study focused on classifying a given image (corresponding to a particular scene) to one of predetermined categories. The research was conducted to evaluate the real-time cleanliness of a public restroom. The prediction model in their system comprised the application of principle component analysis (PCA) together with selected CNN architectures. The research compared the performance of three CNNs: Vgg16 [17], ResNet-50 [19] and Inception-v3 [26]. The proposed method of classification in their research ultimately yielded an accuracy of 90.52% with an inference time of 2.87 s per image. Furthermore, Ojala et al. [28] developed a method to monitor the cleanliness in public transportation vehicles such as trams and metros. A laboratory prototype of a tram interior was used in their study for data collection during which images were captured using a stationary wide-angled camera unit. The results, presented in Fig. 1, depicted that the

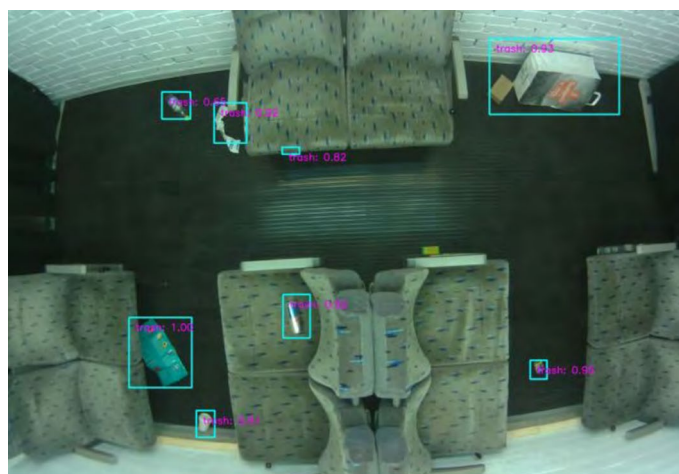


Fig. 1 Garbage detection in a lab replica of a metro. Source: Adapted from [28]

use of Single-Shot Multibox Detector [29] yielded an average accuracy of 95% on the test set of images.

Unlike certain related applications in previous studies, the current research must tackle environmental variations, such as changes of lighting and shadow levels in the images (which is due to the simple fact that light enters the vehicle during operation). Cuhadar et al. [30] theorized how the effect of lighting conditions on the performance of CNNs as they seem to corrupt the detection accuracy of the model. Fu et al. [31] studied on-road vehicle detection under varying lighting and weather conditions in which the concept was validated using a CNN model and a target-oriented scene classification module to co-adapt scene classifiers using a vehicle detector. The research concluded that the developed framework could be utilised in rainy, snowy, and foggy weather conditions.

Indoor air quality monitoring

During daily activities, more than 1 h is spent in enclosed vehicles with the relatively small interior of vehicles contributing to higher concentrations of pollutants and particles which may potentially compromise the health and comfort of the passengers [32]. Thus, in recent years, it is not surprising that the air quality within vehicle cabins has become a relatively hot topic for researchers, public authorities, and industry partners as evident in [33–35]. It is important to note that most odours related to a potential bad smell indoors are as a result of one or more VOCs [36]. In the context of the current research, where the subjective environment is a vehicle cabin, even a smell which maybe considered by certain individuals as a pleasant smell (such as the smell of certain food items) is also considered to be unacceptable. This is because VOCs is an indicator for both pleasant and unpleasant smells. Essentially, the sensor unit will not differentiate between good and bad smells. Thus, if the concentration of VOCs is high, we know that there is a change in odour and action needs to be taken.

In the scope of the current study, we focus on a holistic approach with respect to the two key indicators carbon dioxide (CO₂) and volatile organic compounds (VOCs). Previous studies have concluded that information regarding in-vehicle air quality can be successfully conveyed using the concentrations of CO₂ and VOCs together with indicators mapping the thermal conditions inside the vehicle (temperature and relative humidity) [37]. While CO₂ is directly related to the human occupation indoors, VOCs are a family of chemical compounds which can be emitted in the environment by multiple sources including interior materials, human metabolism and food [38]. These compounds are a main factor for pleasant and unpleasant odours in an enclosed environment and high concentrations of these pollutants can compromise the health of passengers. Based on the signals corresponding to the concentrations of CO₂ and VOCs, air quality can be classified into different categories ranging from good to poor air quality conditions as presented in [39] and [40]. Some previous research studies already conducted an evaluation of CO₂ concentration and other indoor pollutants, based on different ventilation settings [41]. Although there are studies that have evaluated the automotive suitability of an indoor air quality-monitoring system as evident in [42–44], there has not been a concrete implementation of the concept in shared vehicles.

Additionally, the monitoring of other gases (such as NO_x, CO, NO₂ and PM10) have traditionally been of importance to ensure an acceptable indoor air quality. However, sources of these pollutants are mainly found in the outdoor environment. Hence, they can be mitigated with proper air filtering and ventilation techniques. Due to recent advancements in car air filters such as [45] and [46], we decided not to consider the influence of exterior pollutants within the vehicle cabin in the current scope of the research. With this assumption, IAQ levels were determined solely by monitoring CO₂ and VOCs levels.

Research gap

A research gap is apparent in the lack of a system capable of detecting items within the cabins of passenger vehicles to determine the state of their interior cleanliness. Furthermore, it is essential that the system is capable of remotely and securely relaying information pertaining to the indices of cleanliness (leftover items and state of odour) from the vehicle. To carry this out, vision-based predictions should be formed via a suitable interface, and the results displayed to pursue corresponding maintenance routines in ensuring clean vehicle cabins. With respect to odour-monitoring, there is a lack of a holistic study evaluating the effect of different HVAC settings in the concentrations of both CO₂ and VOCs emitted by vehicle occupants.

Methods

Camera module

Based on previous studies in similar application domains of image recognition, it was evident that a camera-based system would best meet our requirements of system input with respect to reliability, accuracy and efficacy. Therefore, the methodology of the system could fundamentally be considered to be three-fold; a camera module to obtain the input status of the vehicle, algorithm development to create the prediction model which will determine the state of cleanliness and a system architecture for secure data transmission, execution and storage.

The research established specific requirements for the camera module to suit its scope of application. The detection area for the current study was limited to the rear seating area of the vehicle as indicated in Fig. 2, which depicts the boundaries of two obtained images. The location of the camera module had to be such that it avoided the top centre portion of the vehicle (as some vehicles have a sunroof). Consequently, it was decided to locate the camera on one side of the vehicle above the rear seat. Since the goal was to capture the entirety of the seating area (including the floor and backrest) using a single camera on one side of the rear interior, the lens had to be wide-angled. In addition, the camera module had to be compact (since its existence should not affect the ambience within the cabin), modular and durable for long term use. The unit also had to be optimised for effortless installation, component replacement, troubleshooting and cost. Additionally, in the possible event of a circuit failure or spark, the design had to consider fire safety with the unit being able to isolate itself and not spread the fire to the exterior of the housing.

With the requirements in mind, the housing of the camera module was designed using CAD software to optimise its dimensions for size as we intended the module to



Fig. 2 Scope of detection with respect to the camera location of the two vehicles: **a** SEAT Mii **b** Aalto Ford Focus

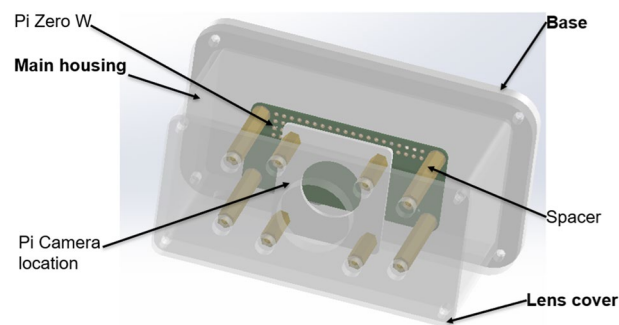


Fig. 3 Labelled CAD model of the camera unit housing

be simultaneously compact and capable of effortlessly replacing components if needed. The final unit consisted of three parts: main housing, base and the lens cover. The main housing and base were 3D printed using a fireproof material PA2241FR, while the lens cover was constructed by milling a clear acrylic sheet of 3 mm thickness. The labelled components of the assembly are shown in Fig. 3.

Figure 4 depicts the manufactured housing. The main housing comprises much of the circuit housing and is therefore primarily responsible for containing the circuit components. The lens cover protects the camera lens from possible damage by exterior sources during operation, which is essential as scratches or other forms of damage to the camera lens could result in hindering the camera's ability to obtain undistorted images. The base attaches to the roof of the vehicle interior on the top end, and to the main housing on the bottom end. With this design, the main housing could simply be unscrewed from the base (which is permanently attached to the roof of the vehicle) in the event of having to modify or replace the circuitry inside the housing (e.g., fuse replacement).

The goal of the electrical design was to use a minimum number of components for reliable operation, safety, and cost efficiency. Considering the proposed design together with the fore mentioned requirements of the scope of detection, a stationary

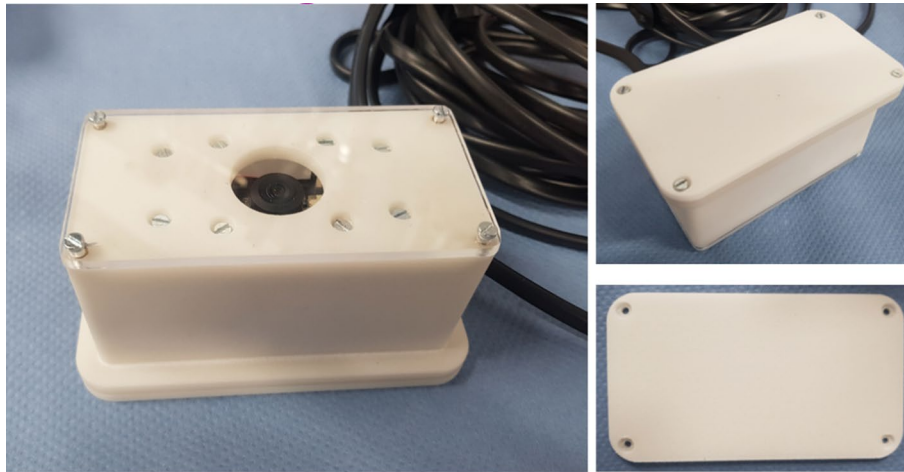


Fig. 4 Manufactured and assembled camera housing

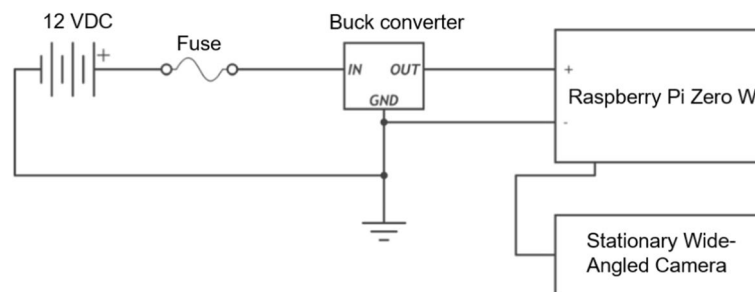


Fig. 5 Circuit schematic of the in-car camera module

wide-angled (160°) Raspberry Pi camera was used to capture images. It was paired with a single board computer, the Raspberry Pi Zero W, which had dimensions of 66 mm by 30.5 mm (the smallest in the Raspberry Pi series). The functions of the Raspberry Pi Zero W were to trigger the camera for image capture and temporarily store the images before transmission to the server for analysis. Since the input power to the camera module was provided by the 12 VDC outlet of the car (as a measure to prevent an additional inclusion of a power source), a buck converter was used as a step-down module to reduce the input voltage to 5 V for safe operation of the Raspberry Pi unit.

In addition to the protection that the single board computer receives from the buck converter, an external fuse was included as it would be easier and cheaper to replace this fuse in the case of a voltage surge. Figure 5 presents the final circuit schematic and Fig. 6 illustrates the internal assembly of the circuit with the base removed. Note that in this diagram the power chord outlet has been sealed using an internal knot of the power cable itself to ensure that the unit is dust-proof as well as to safeguard the circuit against unintended or sudden cable jerks from the exterior.

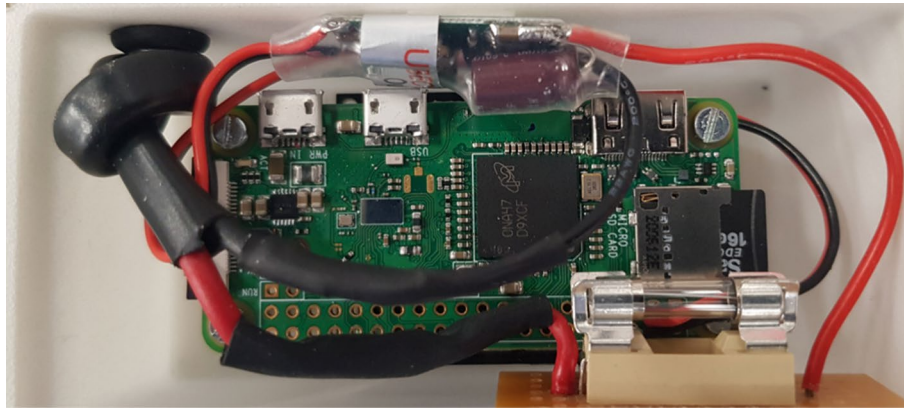


Fig. 6 Internal circuitry of the camera unit (as viewed from the top)



Fig. 7 Camera module installations in the SEAT concept car (top) and Aalto research vehicle (bottom)

Dataset for detection using computer vision

The research conducted in this paper focused on detecting trash and valuables within the vehicle cabin by utilising a self-obtained dataset, integrated with images imported from Kaggle [13] and the Trashnet [12]. Kaggle is an open-source platform which provides access to existing datasets that can be utilised to train custom prediction models. Trashnet is a dataset developed by Thung and Yang in Stanford [12] and aimed at training algorithmic models capable of trash detection.

With respect to the self-gathered data, we obtained images from two camera modules. One of the modules (depicted earlier in Fig. 4) was installed in a concept car located at the SEAT facilities in Barcelona. The other module was locally installed inside one of Aalto University's research cars. Figure 7 presents the two installed camera modules. Note in the figure that the locally installed camera module has exterior dimensions which differ from the module installed in the SEAT concept car. The reason for the difference is due to the practical requirements with respect to the

installed locations. Despite this difference, the circuit schematic and the components used were the same for both camera modules.

A time interval for the measurements was applicable for taking empty images (images with no trash or other external items in the vehicle). Essentially, the process was to park the vehicle in different locations during the day and activate the raspberry pi to capture images periodically (every 30 min or every hour depending on the weather conditions) without the requirement of human intervention. On the other hand, for obtaining images with items inside the vehicle, there was no fixed time interval between the captures as the objective was to capture as many images (falling under the image classes considered within the scope of the current research) at different times of the day and varying external conditions (such as external lighting and shadows). Data quality control was conducted through visual inspection of the authors and members of the research consortium in Aalto (Espoo) and SEAT (Barcelona).

Figure 8 illustrates the breakdown of the classes in the training images. Note in this Figure that “trash” refers to miscellaneous trash items (such as plastic food cartons, food wrappers, trash bags and crushed paper) and “valuable” refers to different types of mobile phones and wallets. The following link corresponds to the current version of the self-gathered dataset of 1379 raw images which was made publicly available for future use in similar research applications: <https://github.com/nilushacj/Clean-Mobility.git>.

Furthermore, it was ensured that the images obtained from the two vehicles had varying levels of light intensity as well as shadows. The purpose here was to ensure that during the training process the accuracy of the predictions would not be hindered due to these external sources of variation in the images. To accomplish this, images were obtained at different times of the year under different weather conditions. It is also important to highlight that the purpose of integrating external images into the training dataset was to expand the dataset and enhance its ability to generalise better to different types of images in a particular class. Additionally, the quantities of images imported from Kaggle was managed in such a way that there was a quantitative balance across all image classes.

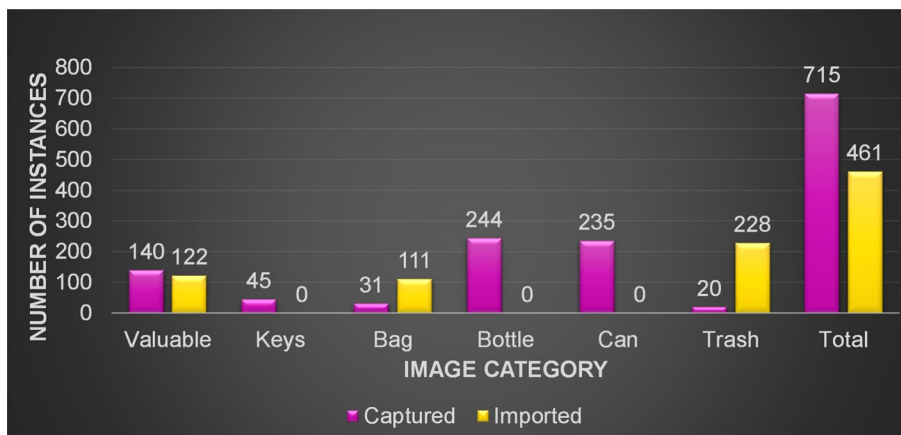


Fig. 8 Distribution of the image classes across the training dataset (quantities excluding augmentation, clean images and test data)

Once all the images for the training process were locally stored, each of the images were manually annotated with the respective image labels (corresponding to the classes in Fig. 8). The annotations were applied by using the LabelImg tool written in python. The bounding boxes were carefully drawn to ensure that the edges touched the outermost pixels of the labelled object and that the boundaries of the object were subsequently never outside the bounding box. At the end of all the annotations, each labelled image consisted of its own text file in Pascal format (.xml).

Vision-based algorithm

The implemented prediction model builds upon the Efficientdet object detection model. The main advantages of using this model is that its architecture comprises a weighted bi-directional feature pyramid network (BiFPN) that enables optimized feature fusion which allows easy and fast multiscale feature fusion. Additionally, it also provides the capability of uniformly scaling the resolution, depth, and width of the backbone, feature network and prediction network given in Figure 3 in [9] simultaneously. The performance metrics of the Efficientdet architecture (in comparison to other detection models) presented in Table 2 in [9] and the required application context of the current research together motivated us to choose this model to generate the vision-based predictions. In the current study, the model was developed in TensorFlow 2.5.0—Keras-nightly 2.5.0 [47] and open-source computer vision library OpenCV 4.5.3 [48]. The version of the Efficientdet object detection model used for the current study has been originally trained on the COCO dataset [49] with approximately 170 classes across 100,000 images. However, the classes we require in achieving the prediction goals of the current system are highly application specific as opposed to the more general classes in COCO. Therefore, it was not possible to establish a direct application of the Efficientdet detection model in our system. The execution speed of the model is registered as 54 ms and a COCO mean average precision (mAP) of 38.4, both of which are relatively similar to other commonly used prediction models such as SSD [29].

More importantly, our model uses input images of size 640 by 640 as input which, in comparison to other existing detection models, is high with respect to neural network processing. In the current application, it was important to ensure that the input resolution was high since a single camera located at one end of the interior (along the width of the vehicle) was intended to capture all the details of the rear seating area. With lower resolutions, certain details would be lost especially in low-light conditions as well as in scenarios in which the object lacked contrast with its neighbouring background (e.g., black mobile on black carpet located at the furthest end of the camera). With respect to the important parameter types of the model's training configuration developed for the current study, the predictions were intended across 6 image classes as shown in Fig. 8. The image pixels were mapped to the feature space using the bi-directional feature pyramid network technique as it optimises the traditional implementations of feature recognition in terms of computation and, in turn, time consumption. Regularization in the form of l2 [50] was utilised to reduce overfitting during the training process. A lack of regularization evidently resulted in a decline of the model's ability to generalise for new data. The model also uses state-of-the-art activation in the form of Swish activation as this function was reported to outperform the more commonly used ReLu activation [51].

Additionally, the model was trained in a GPU machine. Its architecture was of type Tesla V100 with a memory of up to 32 GB. Therefore, it was possible to train the prediction model with a relatively higher batch size of 16 images. The ability to enable larger batch sizes provides the advantage of applying high speeds during the training process while maintaining the stability of the learning process. Data augmentation was also enabled within the training configuration in the form of arbitrary image manipulation (such as scaling, cropping and flipping). Augmentation of the images helped to add more training data, thereby further reducing overfitting of the model. Furthermore, it was essential to include an optimizer capable of efficiently reducing the losses during the training process by iteratively updating the hyperparameters. This functionality was achieved using an optimizer of type Stochastic Gradient Descent (SGD) with momentum [52] with a cosine decay in learning rate. Finally, the model was trained for a total of 50,000 steps (as this was experimentally found to be a good value until the loss saturated and the model did not improve any further).

The algorithm was implemented within the TensorFlow object detection API. It was formerly stated (at the end of the previous subsection) that each training image had a corresponding text file to represent its labelled features. This was essentially the first step in optimising the training process as the input sizes were considerably reduced (from an image in jpg format being up to 4 MB, to an xml file which is less than 1 kB). In the next step, the data from all xml files were concatenated to two Tensorflow record files. The two files represented the training data and the testing data with sizes of 1.1 GB and 196.4 MB respectively. Therefore, the overall computation overhead of the model's training input was optimised in this manner.

Indoor air quality (IAQ) unit

The current work developed a new indoor air quality-monitoring system for shared vehicles. Compared to previous studies, our work created a holistic IAQ index, based on the evaluation of the concentrations of CO₂ and VOCs at different HVAC settings and occupation of the vehicle. SPG40 and SCD4x miniaturised sensing kits were purchased from SENSIRION for the monitoring of the concentrations of both VOCs and CO₂ respectively. Both sensors were included in a compact and modular housing unit as shown in Fig. 9 and included the elements presented in Fig. 10. Airflow was enabled within the housing by strategically placed inlets and exhaust outlets in combination with a small fan.

Six data outputs signals were considered from the IAQ module for each test: (i) raw VOC concentration (#1); (ii) VOCs relative index (#2); (iii) raw CO₂ concentration (#3); (iv) interior temperature (#4); (v) relative humidity (#5); and (vi) timestamp (#6). The raw concentration of VOCs was provided in 'ticks', an indicator used by the provider SENSIRION which is proportional to the electrical impedance of the sensor. The higher the number of ticks, the lower the concentration of VOCs. The magnitude is therefore given directly given by SENSIRION. The calibration curves of the VOCs sensor unit, which define the relationship between ticks and the actual concentration of VOCs (in ppm) can be found in Figure 2 of the datasheet of the sensor [53]. Thus, the VOC relative index is directly calculated by the sensor unit and it is given as an output parameter as presented in section 3.2 in [53]. The raw data signal

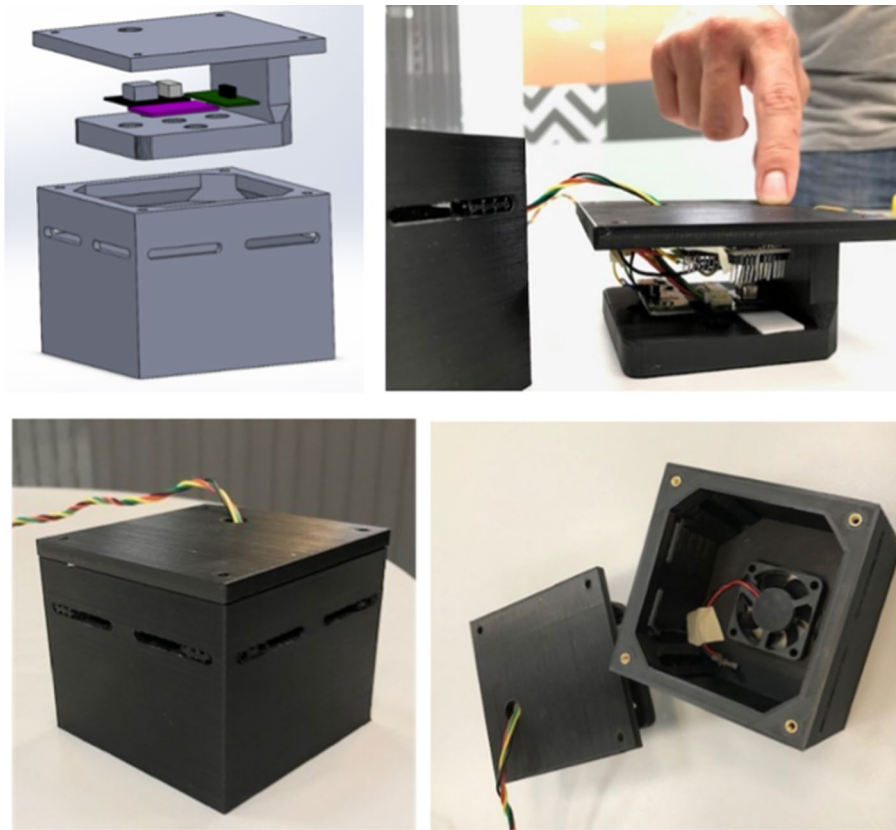


Fig. 9 Designed, implemented and assembled components of the IAQ unit

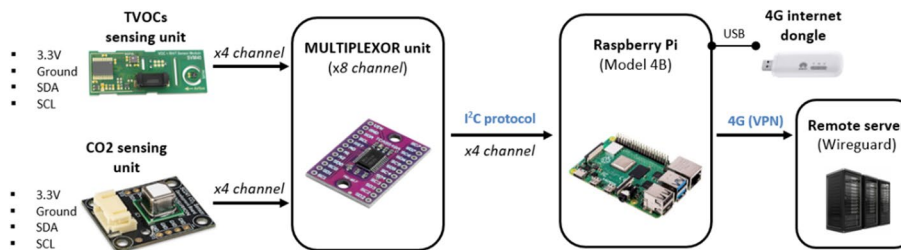


Fig. 10 High level diagram representing the connectivity of the IAQ monitoring system

of VOC sensor unit was used to analyse baseline interior air conditions before passenger occupation in the vehicle. Once occupation began, the VOCs index was used to capture relative changes in the concentration of VOCs, provided by passengers' activity, in comparison to baseline conditions. The raw concentration of CO₂ was displayed in parts per million (ppm) and was only considered during occupation of the vehicle. The VOCs index measures relative changes in the VOCs concentration compared to baseline conditions. Hence, it is well-appreciated that baseline conditions inside the vehicle will affect this relative measurement. The empirical study of the current research will define the reference baseline conditions (to which the concentrations of CO₂ and VOCs should be within prior to occupation of the vehicle and the

calculation of the relative VOCs index). The IAQ unit was installed in the SEAT MII concept car as shown in Fig. 11 (this vehicle had no sunroof). The position of installation shown in the figure was optimum to detect changes in air quality in both the front and rear seating areas.

System architecture for content visualization

To manage the overall functionality of the detection system with respect to maximum efficiency, no interference and ease of troubleshooting, it was crucial to ensure that all entities of the system architecture are conveniently linked in accordance with their roles. Subsequently, the architecture had to meet the requirements of data transmission, data storage, program execution, information access and visualization, all within a secure network. Thus, it was decided that there would be three main entities among which communication would occur: the sensor units, remote server, and interface client. Henceforth, these entities will be communicating within a VPN network (Wireguard) [54]. Figure 12 illustrates the communication architecture of the current research.

Starting from the front end of the detection system (the vehicle cabin), the Raspberry Pi Zero W is connected to an onboard Wi-Fi source enabled by a Huawei Wingle. Since it is not possible to guarantee the strength of the service provider's signal at all locations, the Raspberry Pi Zero W is capable of storing the images until it detects a signal sufficiently strong for the transmission of the image to the server. Additionally, in accordance with the privacy law corresponding to obtaining pictures with passengers, the unit will only process images which have no travellers. The remote server was purchased from Linode, a cloud computing service, which consisted of 2 CPU cores, 80 GB of storage and 4 GB of RAM. The specifications of the server were selected to best suit our application and balance the cost of usage. The server runs 24/7 and is a headless version of the Ubuntu operating system (no graphical user interface). The server contains the prediction model which will be executed on receiving an image from the camera unit in the vehicle cabin. Once the image is processed, the results (including the jpeg version of the image, detected objects and certainty level of detection) will be transferred to the interface client for visualization and adopting corresponding maintenance action. Additionally, bash scripts were written to enable client access to the server and the Raspberry Pi

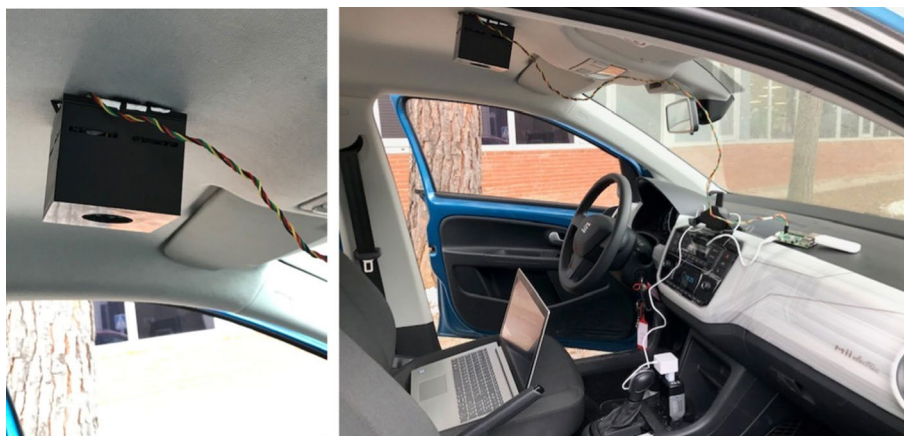


Fig. 11 Test setup of the IAQ module in the SEAT MII concept car



Fig. 12 Communication setup of the detection system

for system operation as well as for development purposes (such as updating software and modifying program contents). Communication protocols including secure copy protocol (SCP) and secure shell (SSH) were utilized here.

Results

Vision-based detection unit

To evaluate the performance and viability of the implemented algorithm and its execution platform, we primarily focused on the prediction accuracies from camera modules in both vehicles. In addition, the false predictions were further broken down into categories comprising misclassifications, false negatives, and false positives. The overall prediction accuracy of the model will be determined by the mean Average Precision (mAP), F1 score and a confusion matrix for the batch of test images. Furthermore, the execution speeds were analysed to identify the viability of implementing the proposed architecture in shared vehicles.

For the images acquired from the camera modules of the two vehicles, it was essential to have a preprocessing stage prior to inference. Preprocessing would occur in the form of rectangular crop. The parameters for the script which performs preprocessing for each image input were calibrated for each vehicle at installation. Preprocessing of this nature was required to eliminate visually redundant areas of the images such as exterior details of the car and unwanted corners. Existence of such features hindered the performance of the prediction model as it tended to detect unwanted features with incorrect class labels.

Figure 13 presents the breakdown of the results from a pilot conducted for the concept car in Seat. The testing included capturing images in five distinct parked locations at two times of the day to ensure varying light intensities and shadows in the images. The pilot was aimed at examining the execution speeds (from image capture to executing the prediction model and displaying the results). The observed execution speeds are presented

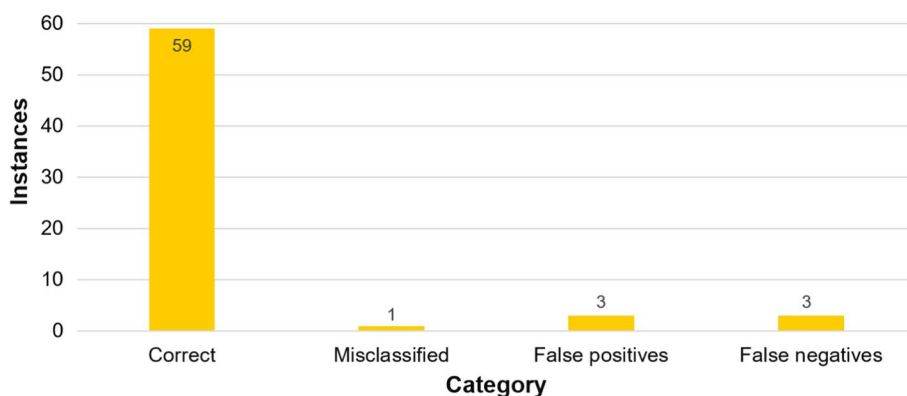


Fig. 13 Prediction breakdown from a pilot session conducted with the SEAT concept car yielded an accuracy of 89%

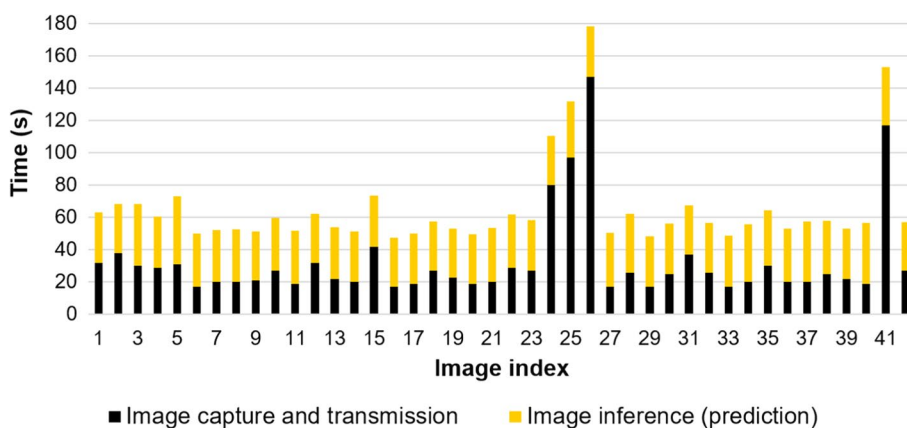


Fig. 14 Execution speeds for each test image from the pilot session conducted with the SEAT concept car (Mean: 65 s, Max: 178 s, Min: 47 s)

in Fig. 14. The inference time was precise in all trials and the main contributor of the variation in the total execution time was the image transfer from the camera unit to the server (which depended on the strength of the local network connection).

Figure 15 presents the precision and recall plots for the final trained prediction model. The graphs show that the mAP and recall reached 0.87 and 0.85, respectively. Figure 16 presents the confusion matrix for the test batch of images. The final entry of the matrix represents the empty images in the test set which were accurately detected. Therefore, with the exception of this entry, the last row and the last column corresponded to the false positives and false negatives respectively. Based on the confusion matrix, the accuracy of the detections from the test batch was 89%, with a corresponding F1 score of 0.96.

Indoor air quality-monitoring unit

A series of pilots were conducted with the SEAT concept car to establish IAQ indexes based on specific air pollutants. The initial pilots was conducted to establish baseline conditions, prior to passenger occupation and when there were no external sources

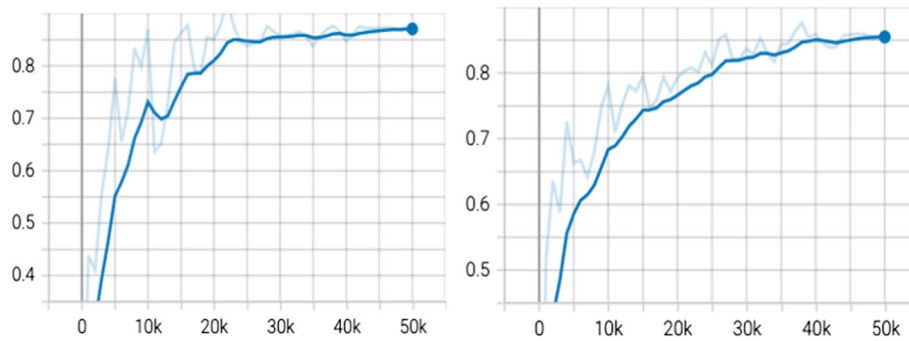


Fig. 15 Precision (left) and recall (right) curves for the batch of test images compiled from images taken with the two research vehicles. The faded curves presents the exact plot values while the darker colour correspond to the averaged (smoothed) curves

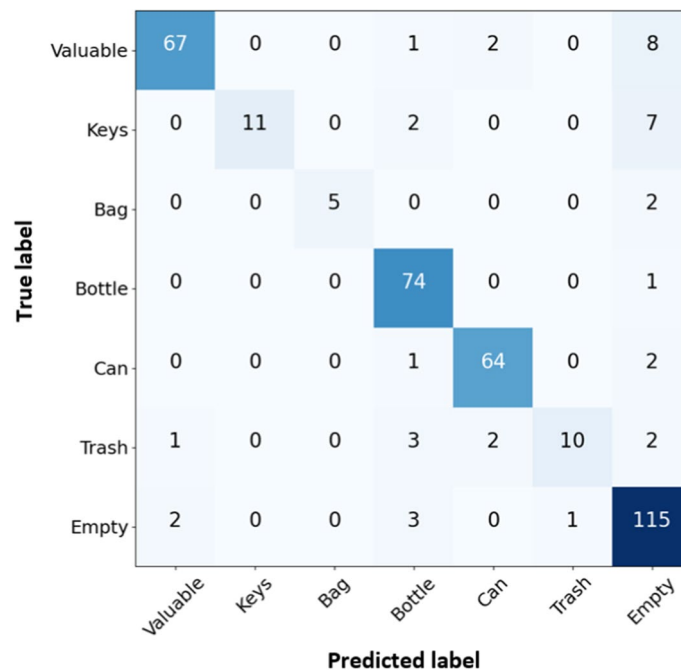
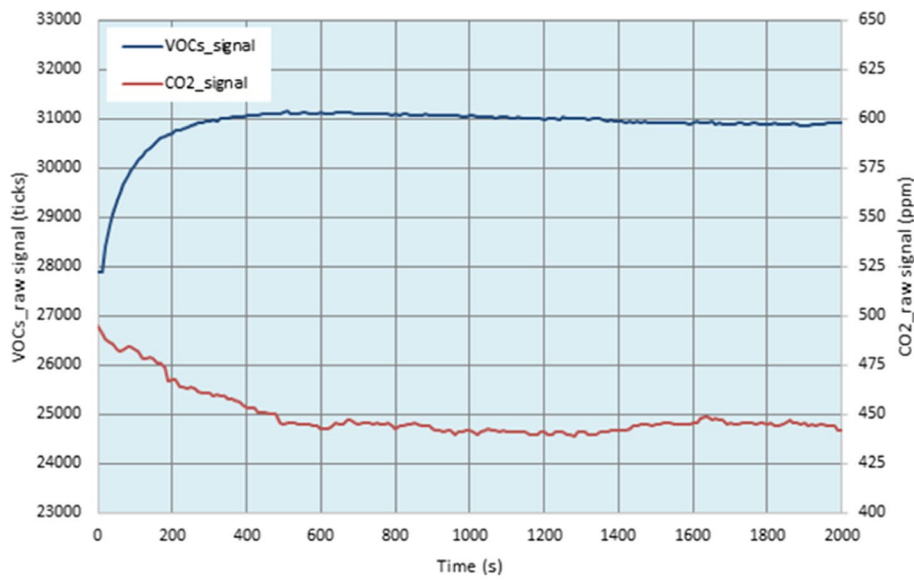


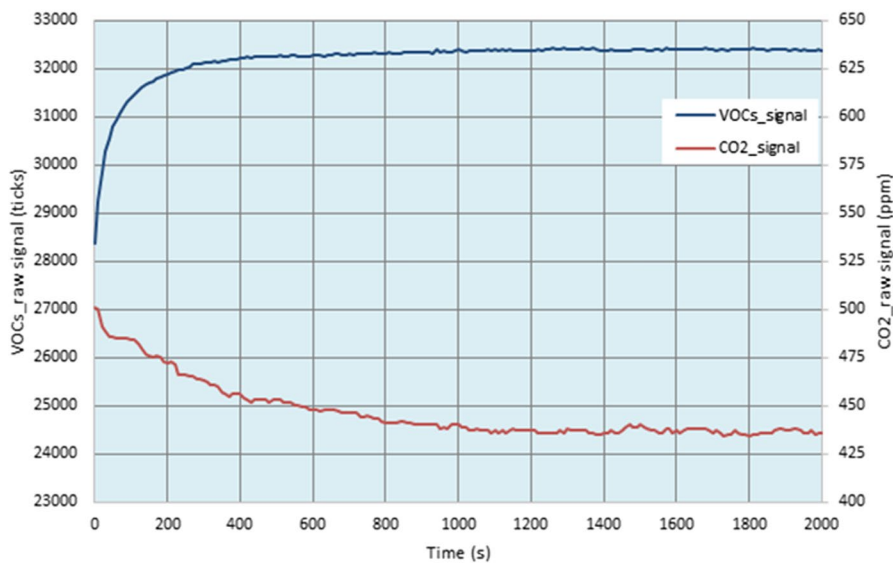
Fig. 16 Confusion matrix for the test batch of images captured with the two research vehicles

of odour compromising the environment. Trials were performed using different environmental conditions which were the only sources of signal variance for both CO₂ and VOCs. Hence, the tests were executed to define good or excellent indoor air quality conditions, which should be targeted by the IAQ module before passenger occupation. The temperature was maintained between 19 °C and 22 °C and relative humidity between 50% and 55%. The ventilation setting inside the vehicle was set to ‘auto’. As shown in Fig. 17, the concentration of VOCs remained between 30,500 ticks and 32,500 ticks and the concentration of CO₂ between 425 ppm and 450 ppm.

Variations of the VOC conditions were mainly caused by slight variations in environmental conditions (such as temperature and humidity). An increase in interior temperature also led to an increase in VOCs emitted by the surfaces within the



(a)



(b)

Fig. 17 IAQ output signals in empty vehicle conditions. Differences between VOC signals are due to slight variations in temperature or relative humidity

vehicle cabin. Thus, a smaller number of ticks is obtained with a higher temperature. Regarding CO₂ concentration, with no occupants in the vehicle this should be restrained below 450 ppm.

It was already presented that the ambient temperature can contribute to the rise of VOC levels inside the vehicle when empty. There is a direct correlation between temperature and VOCs emitted by the surfaces inside the vehicle: the higher the temperature, the higher the concentration of VOCs. This behaviour was observed by exposing the vehicle to sunlight for a prolonged period. The temperature inside the vehicle cabin increased from 23 °C to 35 °C after 2 h of exposure. Ventilation

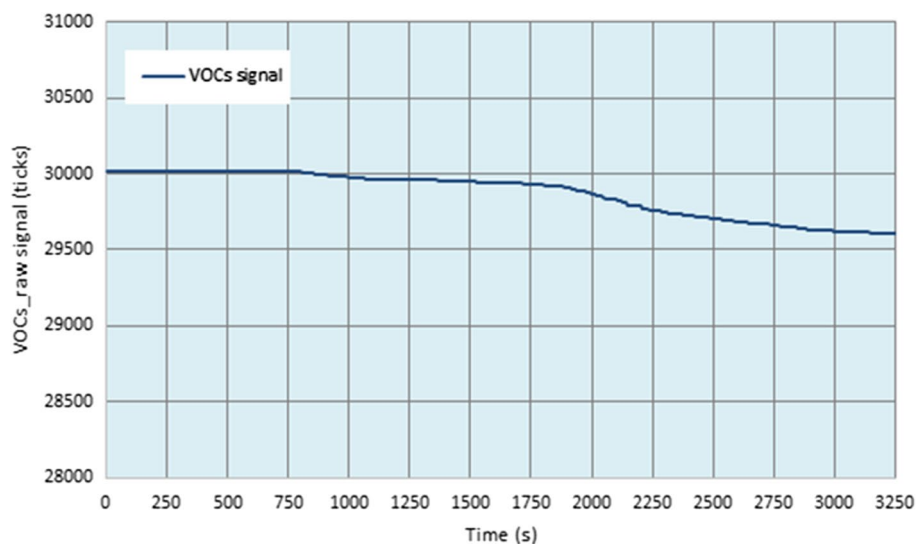
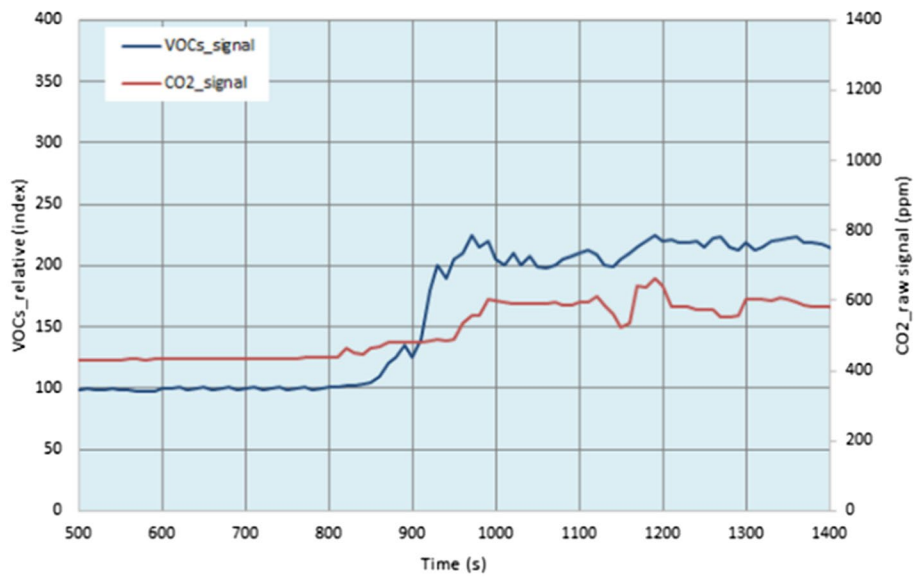


Fig. 18 Concentration of VOCs due to variations in interior vehicle temperature (23–35 °C)

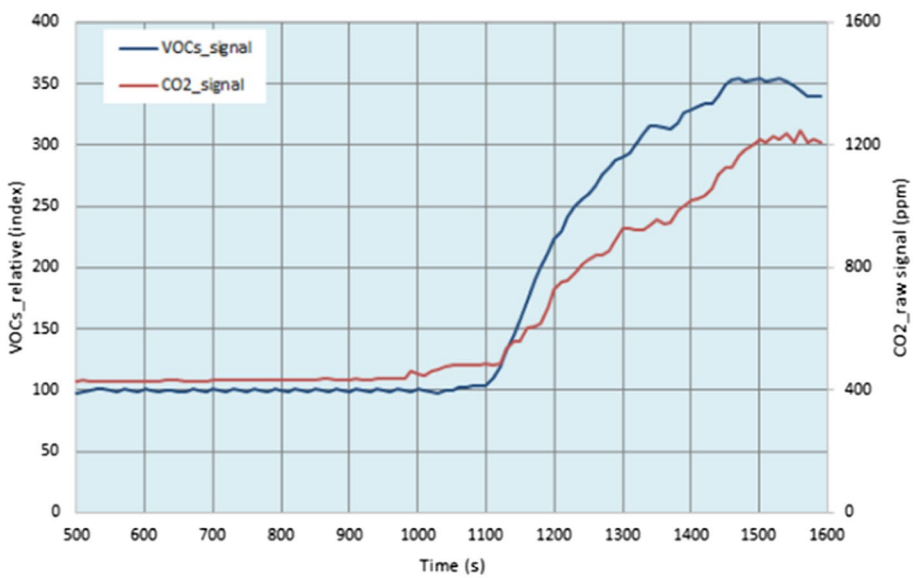
and air conditioning systems inside the vehicle were switched off during the trials. As depicted in Fig. 18, changes in VOC levels due to variations in interior temperature were depicted by only analysing the raw concentration of VOCs over time. The increase in temperature inside the vehicle caused an increase in VOC concentration and a decrease in the number of ticks below acceptable reference conditions (< 30,000 ticks), showing a slight deterioration of air quality over time. One observation gained from this first set of tests was the importance of controlling the temperature inside the vehicle cabin. This parameter impacted the air quality measurements inside the vehicle cabin and needed to be restrained. To achieve this, a recommended solution was to activate ventilation and climatic actuators until environmental conditions reached an optimum point (19–22 °C). Only after this, can the baseline conditions be established, and one can ensure that the condition of air is good for the occupants.

After establishing reference baseline conditions, indoor air quality was monitored according to passenger occupation. The vehicle was empty until good or excellent indoor air quality conditions were met after which occupation was initiated and maintained for a period of 5–10 min. Relevant output signals in this case were the raw concentration of CO₂ and the relative change in VOC levels (VOCs index) triggered by occupants. An increase in this index represented a worsening of the interior air quality. In the high-occupation scenario, the vehicle was occupied with 4 passengers and ventilation was kept to a minimum. In the low-occupation scenario, only 2 occupants were inside the vehicle and the cabin was well ventilated for the whole period. The results are shown in Fig. 19.

With high occupation, there was a period before occupation during which both VOC and CO₂ levels remained stable at reference conditions (VOCs index was approximately 100 and the CO₂ levels were less than 450 ppm). After occupation, both signals significantly increased, due to low ventilation and high occupation, revealing a worsening of indoor air quality over time. At low occupation, there was an



(a)



(b)

Fig. 19 IAQ output signals at (a) low occupation and (b) high occupation of the vehicle

increase in CO₂ and VOC levels, but the increase was more subtle and sustained over time due to good ventilation of the interior (CO₂ < 600 ppm and VOCs index < 250).

The final pilot was devoted to a scenario in which there was an external source of odour generated by sources such as food remains, smoking, care products, vomit and animal hairs. For the trials, Ethyl Acetate was selected as a representative compound since it has a strong and pungent odour. Drops of Ethyl Acetate (2 ml) were poured into a small recipient, placed in the vehicle and left for a certain time period at nominal ventilation conditions. Temperature and relative humidity were maintained within the optimum range during the trials. Figure 20 shows that the CO₂ levels remained constant and

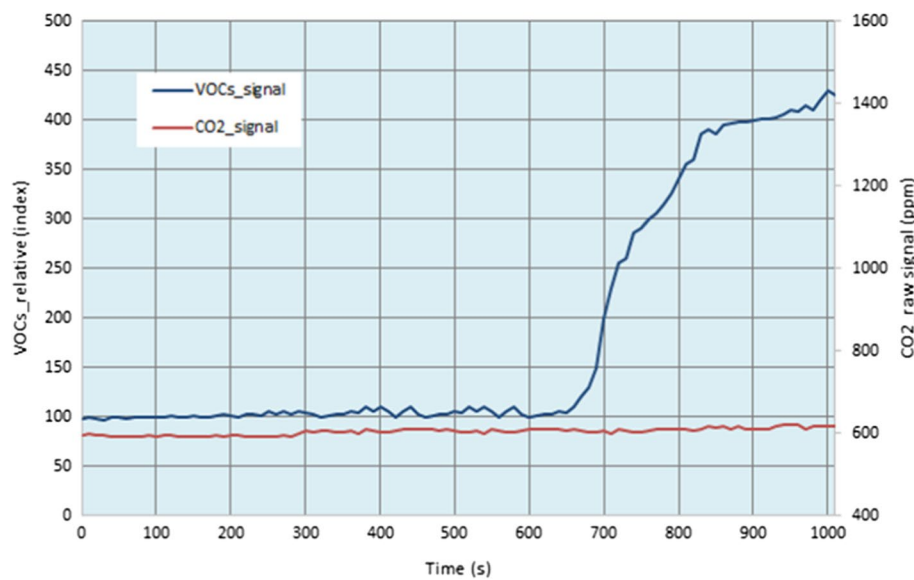


Fig. 20 Pilot 4: IAQ output signals of a malodour event

at acceptable levels. Only VOC levels were recorded to dramatically increase, indicating a serious worsening of interior air quality caused by the external source. This situation should be reverted with corresponding supervisory action.

Discussion

Based on the results, it is evident that the implemented vision-based system was able to detect the presence of items in the rear seating area of the vehicle. Figure 21 presents a set of sample images with successful detections. However, it is interesting to dive into the specific accuracies with respect to the trained classes of objects, external sources of variation including light levels and shadows, vehicle type and camera locations.

Figure 22 shows the confusion matrices of the predictions from each vehicle used to compile the test batch of images. Note that the confusion matrix for the Aalto research vehicle is based on two camera modules: the camera unit used in the current research and test images from an old camera unit which was used in [55]. The accuracies of detections were 92% and 88% for the SEAT and Aalto research vehicles respectively. Therefore, the results prove that there is no significant drop in the accuracies of the prediction model of both the vehicles, which indicates the generalisability of the algorithm for passenger vehicles of other types. There were also notable sources of specific errors in the results. First, it was important to ensure that the images were taken when there was sufficient daylight to ensure that the captured images were not too dark to hinder the visibility of any object in the car. Figure 23 presents an instance in which items (bottle on the floor and wallet on the seat) were left undetected due to a lack of visibility. Essentially, these were instances of false negatives due to the lack of contrast between a particular object and its neighbouring background pixels. The confusion matrix in Fig. 16 proves that the percentage of false negatives was 6.00% with a total of 22 instances in which objects were left undetected. However, there were only 6 instances of false positives (ratio of 1.55%) observed to primarily exist due



Fig. 21 Examples of successful detections

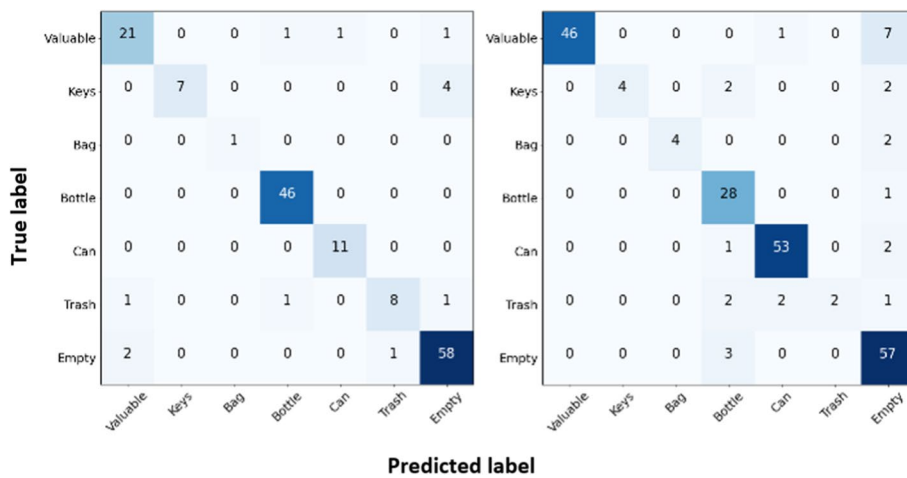


Fig. 22 Confusion matrices for test images obtained from the SEAT concept car (left) and Aalto research vehicle (right)

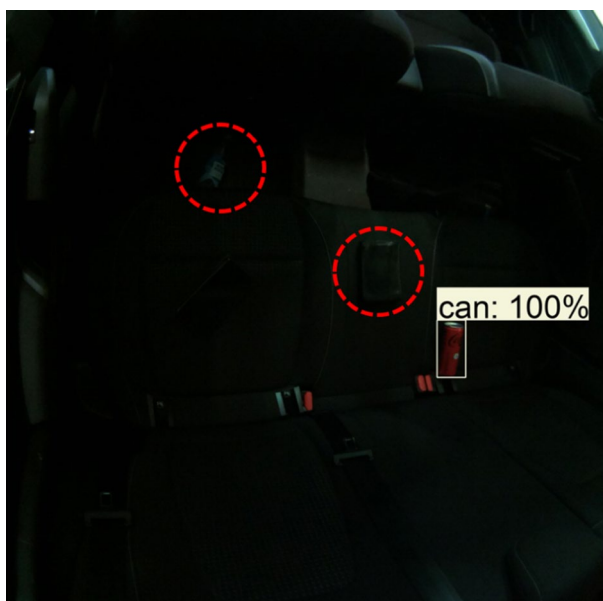


Fig. 23 Predicted test image lacking visibility. Red dotted circles correspond to the undetected items

Table 1 Breakdown of prediction accuracies for each applicable item class

	Val	Keys	Bag	Bottle	Can	Trash
Ratio (capture)	0.20	0.06	0.04	0.34	0.33	0.03
Ratio (total)	0.22	0.04	0.13	0.21	0.20	0.21
Acc (%)	85.90	55.00	71.43	98.67	95.52	55.56

to shadows. Figure 24 (left) presents an instance in which the shape of the shadow is misinterpreted by the prediction model as being a valuable.

Based on the results, the binary classification accuracy (detected vs undetected) was 93%. Consequently, if the manner of removing items from the vehicle is unique across all items, it is already useful information to the maintenance personnel that the algorithm detected the presence of an item (regardless of the classification).

The remaining errors in the confusion matrix were misclassifications in which the prediction model classified the detected objects incorrectly shown in Fig. 24 (trash classified as a bottle). The total number of misclassified instances was 12 which corresponds to a percentage of 3.11%. Table 1 presents the breakdown of prediction accuracies for each class for which the prediction model was trained. The row wise data of the table correspond the ratio of the self-gathered (captured) training images, ratio of the total training images and the final test accuracy respectively. The table suggests that the accuracy values are low for the classes which have the lowest ratio of instances in the dataset. From Fig. 8 it was evident that the class with the lowest quantity of images was “keys” with a mere total of 45 images due to practical difficulties in obtaining unique images of key sets. Thus, this category also has the lowest accuracy with respect to the detections. On the other hand, the best image classes with respect to quantity and variety (valuables, bottles and cans) had the highest relative accuracies.

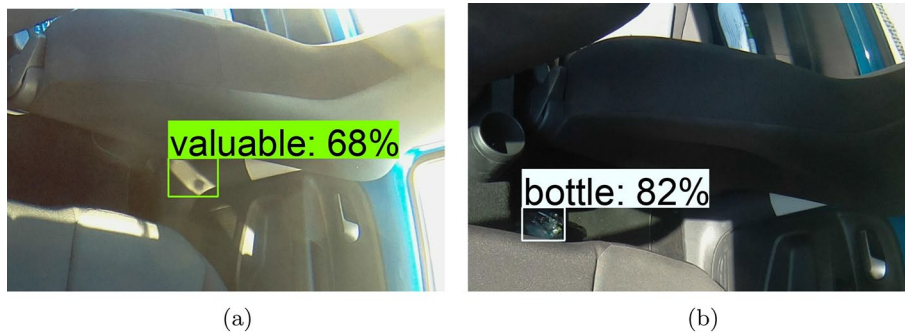


Fig. 24 Predicted test images of a false positive as a result of (a) a shadow and a misclassification in which (b) a miscellaneous trash item was classified as a bottle

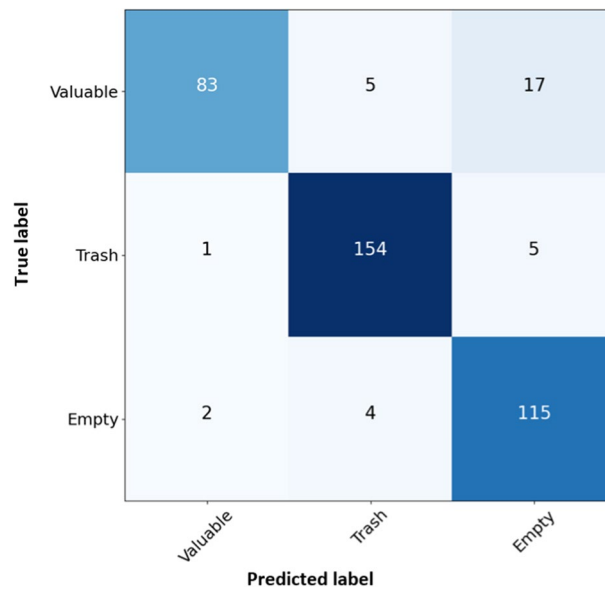


Fig. 25 Confusion matrix with respect to the three fundamental classes

Additionally, the use of images from open-source platforms, such as Kaggle [13], proved successful in balancing the quantities among most classes for better training. The “keys” class was an exception since no images were available for this class from public datasets. It was especially useful to employ this technique, of importing images to expand the training dataset, for the “trash” class since it was not practical to repeatedly litter the inside of the vehicle.

Furthermore, it is notable that most of the self-gathered images of trash items are cans and bottles. Although these were trained as separate classes, they essentially belong to the miscellaneous “trash” category. Therefore, misclassifications among these classes are technically harmless with respect to the final use-case scenarios. The confusion matrix in Fig. 25 presents the results of merging these three classes into a single category. Based on this, the accuracy has now elevated to 91%.

In addition, two sources of error were observed from the results of the prediction model which were dependent on the placement of the items. First, false negatives were

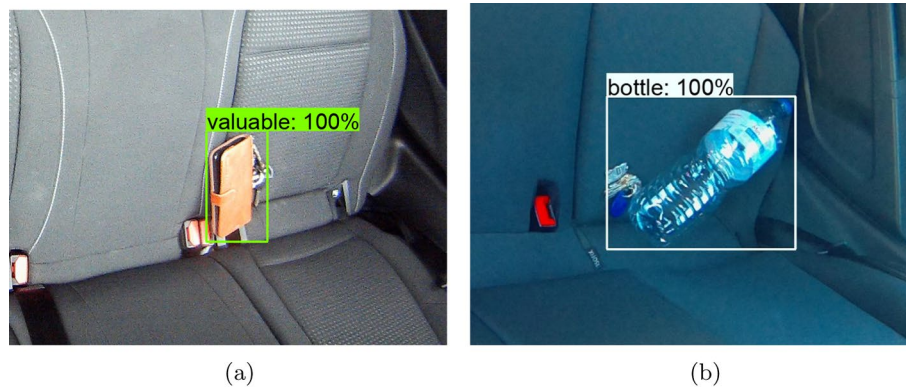


Fig. 26 False negatives of two predicted images in which key sets are left undetected due to overlapping



Fig. 27 Two instances of the bounding box merging for items placed in close proximity to each other. The image in **a** depicts items from the same trained class and the image in **b** depicts items from two different trained classes

evident in a few images due to overlapping of items in which objects were left unclassified as depicted in Fig. 26.

Yet, the model does generate a bounding box over the dominant item in terms of the area of visibility, which is a positive trait since, when alerted, the maintenance personnel can eliminate all the items.

Additionally, there were a couple of instances when the bounding box merged when two similar items were placed close together as shown in Fig. 27. However, this was not identified as an error since the desired outcome of detection was achieved.

Moreover, there were instances of false negatives when there was a large quantity of different items scrambled across the seating area as depicted in Fig. 28. This was identified as the biggest source of error in the prediction model. However, images constituting of items from the trash class produced successful results despite the objects overlapping and being placed in high quantities as shown in Fig. 29.

Moving on to indoor air quality monitoring, the pilots conducted with the IAQ unit enabled us to define an IAQ index calculated with respect to the CO₂ and VOC levels within the vehicle. Prior to passenger occupation, the IAQ index was calculated with the raw concentration of VOCs to ensure that its initial conditions are low. Once



Fig. 28 Two examples of predicted images with a high quantity of items scattered across the seating area. Red dotted circles correspond to instances of undetected items (**a** bag and **b** keys)

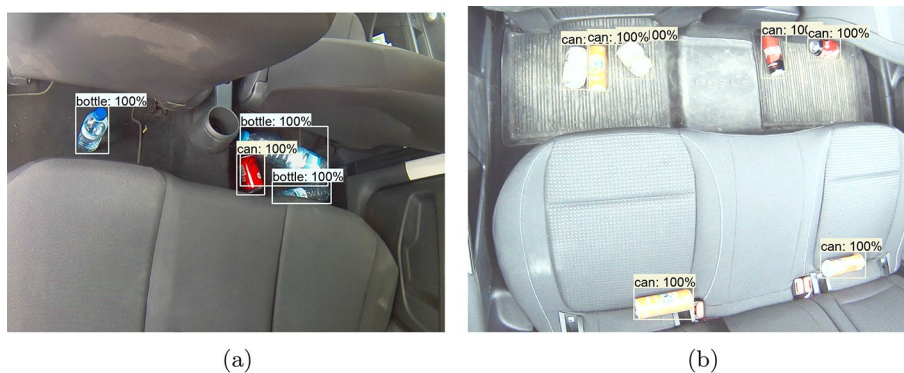


Fig. 29 Two examples of scattering a high quantity of trash items across the seating area

reference conditions were stable, relative changes in the concentration of VOCs were considered, since it was useful to capture changes in the environment caused by the activity of passengers. Hence, the IAQ index considered the raw concentration of CO₂ and VOC levels, as well as the relative change in VOCs in comparison to clean conditions of air which was the baseline or reference level.

Figure 30 provides a summary of the data outputs and relevant supervisory information considered for each IAQ index derived from the results of our research. The threshold values were derived from the empirical results obtained in the current research, together with the information provided in the IAQ guidelines of indoor environments previously studied by organizations such as the German Environmental Agency [56] and the US Environmental Protection Agency [38].

Note that the bars in the figure have been associated with the data labels as their lengths are not drawn to scale in the chart. More importantly, the legend in the figure corresponds to the IAQ index based on the quantitative results of the three indicators. Table 2 presents an elaborated series of actions and messages that should be sent to relevant operators or actuators (such as the HVAC unit) in the vehicle cabin.

Based on the analysis of the results obtained, we were able to determine the aspects of improving and advancing the developed system. The most significant next step was

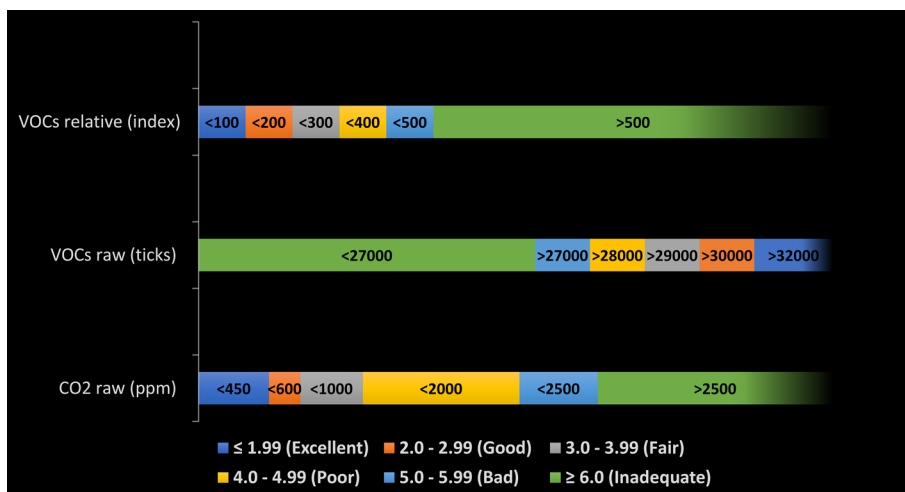


Fig. 30 Derived IAQ index for the air quality assessment in the vehicle cabin

Table 2 IAQ index evaluation

IAQ index	Air quality evaluation	Actions required
≤ 1.99	No appreciable changes	None
2.0–2.99	Negligible changes	None
3.0–3.99	Noticeable changes	Minor increase in ventilation
4.0–4.99	Slight deterioration in air	Increase in ventilation
5.0–5.99	Significant deterioration in air	Significant increase in ventilation. Search for external sources recommended
≥ 6.0	Severe deterioration in air	Mandatory refresh of air. Mandatory search for external sources

identified as the explicit fusion of the vision-based prediction model and the IAQ unit. The outputs of the two currently independent systems will be combined to yield an overall cleanliness level on a discrete predetermined scale. Figure 31 presents a potential instance of this expected outcome.

Moreover, we intend to expand the dataset of images for improving the prediction model with respect to identifying more types of leftover items. Essentially, the limitation of the dataset quantity in the current research is due to the requirement of manually placing items in the vehicle and capturing images for training the model. The process of manual item placement is tedious, repetitive and consumes a lot of time. To overcome this, we intend to install the camera module in a real-world shared vehicle and acquire images when it is in operation. The captured images could then be used to expand the training dataset and improve the prediction model. Having said this, we can confidently deduce that the magnitude of the current dataset was sufficient within the scope of this research as we were able to prove the viability of the application as originally intended.

Conclusion

The paper presented an architecture for determining the cleanliness status in shared passenger vehicles. An in-car camera unit was implemented to obtain images corresponding to the rear seating area of the vehicle. An indoor odour monitoring unit was developed

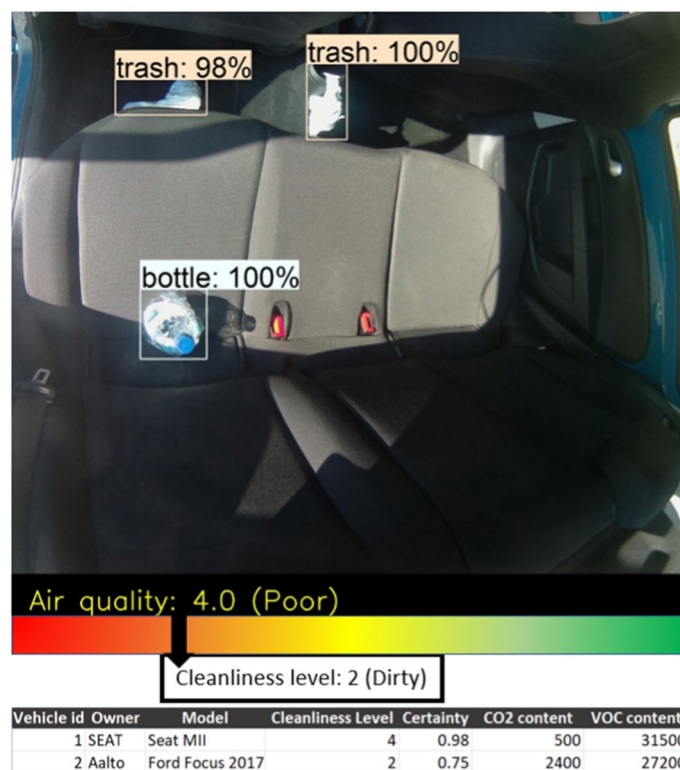


Fig. 31 A sample of the expected outcome of the future integration between the vision-based detection system and IAQ monitoring unit

to sense the concentrations of specific air pollutants within the vehicle. The sensor data from the vehicle were transmitted to a remote server, upon request, for processing.

The developed algorithm was built upon the architecture of the Efficientdet CNN-based object detection model. The trained prediction model in the current research yielded an overall accuracy of 89% across six distinct classes and a binary classification accuracy of 93%. From a practical standpoint, the overall accuracy of classifying between an empty image, an image containing trash and an image containing valuables is 91%. Therefore, the model proved to successfully work for the trained nature of images which included variation in ambient lighting and shadows. The existing sources of error with respect to the output of the prediction model were concretely analysed. Furthermore, the average execution time from capturing an image to displaying the results of detection on screen was experimentally obtained as 65 s. The target at the start of the research was 1 min. Thus, the obtained outcome was acceptable.

With respect to the indoor air quality-monitoring unit, an IAQ index was derived based on the results of four pilot categories. The index comprised 6 levels based on three indicators of air corresponding to the concentration of carbon dioxide and voltage organic compounds. Each index was associated with its definition of the interior environment and recommended actions (if any).

Although the system produced the desired research results, the final product of implementation is further expected to be near ideal in reliability due to the nature of the application. To yield a higher accuracy for more types of leftover items, the best form of

improving the prediction model was recognised as an expansion of the dataset for training by obtaining images from a shared vehicle in real-time operation (including the capture of images during nighttime or when the exterior is dark and requires illumination). Furthermore, the preprocessing stage of the detection unit will be integrated with a person detection algorithm which eliminates images containing one or more people in the vehicle, thereby preventing the transmission of such images to the server. Future work will further examine the possibility of improving the interior sensor units of the vehicle to be more modular with the potential the plug-and-play feature. This would enable the users to fit and remove the detection system into and from their vehicles as intended without disrupting the existing interior. Additionally, a separate article will be formulated and published to cover the in-depth architecture with respect to the specifics of the IAQ unit covering the applicable literature, lab experiments and odour evaluation. More importantly, the two systems of vision-based prediction and air quality monitoring will be explicitly integrated to generate an overall (and discrete) cleanliness level.

Abbreviations

CNN	Convolutional neural network
IAQ	Indoor air quality monitoring
VOCs	Volatile organic compounds
mAP	Mean average precision

Acknowledgements

We thank Jorge Navarro for providing remote support for data collection from Barcelona.

Author contributions

Conceptualization: NJ, EO, RO, KK and KT; methodology: NJ, EO and JP; software: NJ and EO; validation: NJ, and EO; formal analysis: NJ, KK, JV and KT; resources: NJ, EO and JP; data curation: NJ; writing-original draft preparation: NJ and EO; writing-review and editing: NJ, RO, KK, JV and KT; supervision: KK, JV and KT. All authors read and approved the final manuscript.

Funding

This work is partially supported by EIT Urban Mobility. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The current version of the custom dataset generated during the study is available in the Clean-Mobility github repository, <https://github.com/nilushacj/Clean-Mobility.git>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 May 2022 Accepted: 21 January 2023

Published online: 02 February 2023

References

1. Statista: Vehicles & Road Traffic. <https://www.statista.com/markets/419/topic/487/vehicles-road-traffic/>. 2020.
2. Future Mind: Car Sharing and Transportation Trends. <https://www.futuremind.com/blog/car-sharing-and-transportation-trends>. 2020.
3. McKinsey: How shared mobility will change the automotive industry. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-shared-mobility-will-change-the-automotive-industry>. 2017.
4. Shaheen S, Cohen A, Jaffee M. Innovative mobility: carsharing outlook-worldwide carsharing growth. <https://escholarship.org/uc/item/49j961wb>. 2018.

5. Fellows NT, Pitfield DE. An economic and operational evaluation of urban car-sharing. *Transport Res Part D Transport Environ.* 2000;5(1):1–10.
6. France 24: France's car-sharing system Autolib' hits the end of the road. <https://www.france24.com/en/20180621-france-paris-end-road-car-sharing-system-autolib>. 2018.
7. CleanAI: How dirty are shared cars. <https://medium.com/@CleanAI/how-dirty-are-shared-cars-24b8445ffb6>. 2018.
8. Curtale R, Liao F, van der Waerden P. User acceptance of electric car-sharing services: the case of The Netherlands. *Transport Res Part A Policy Pract.* 2021;149:266–82.
9. Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.2020*; pp. 10781–10790.
10. Partin DL, Sultan MF, Thrush CM, Prieto R, Wagner SJ. Monitoring driver physiological parameters for improved safety. *SAE transactions. SAE Technical Paper 13.2006*; 633–639.
11. Schewe F, Cheng H, Hafner A, Sester M, Vollrath M. Occupant monitoring in automated vehicles: Classification of situation awareness based on head movements while cornering. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.2019*; vol. 63, pp. 2078–2082.
12. Yang M, Thung G. Classification of trash for recyclability status. CS229 Project Report.2016; 2016.
13. Chandrasekhar XYZZSGTLWV, Hoi S. Deep learning for practical image recognition: Case study on kaggle competitions. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.2018*; pp. 923–931.
14. Tharani M, Amin AW, Rasool F, Maaz M, Taj M, Muhammad A. Trash detection on water channels. In: *Neural Information Processing.2021*; pp. 379–389.
15. Redmon J, Farhadi A. Yolov3: an incremental improvement. *arXiv preprint arXiv:1804.02767* 2018.
16. Liu J, Jiang Y. Design of intelligent trash can be based on machine vision. In: *2020 International Conference on Image, Video Processing and Artificial Intelligence.2020*; vol. 11584, pp. 245–250.
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014.
18. Adedeji O, Wang Z. Intelligent waste classification system using deep learning convolutional neural network. *Proc Manuf.* 2019;35:607–12.
19. Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2016*; pp. 770–778.
20. Zhihong C, Hebin Z, Yanbo W, Binyan L, Yu L. A vision-based robotic grasping system using deep learning for garbage sorting. In: *2017 36th Chinese Control Conference (CCC).2017*; pp. 11223–11226.
21. Kong T, Yao A, Chen Y, Sun F. Hypernet: Towards accurate region proposal generation and joint object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2016*; pp. 845–853.
22. Rad MS, Kaenel Av, Droux A, Tische F, Ouerhani N, Ekenel HK, Thiran J-P. A computer vision system to localize and classify wastes on the streets. In: *International Conference on Computer Vision Systems.2017*; pp. 195–204.
23. Alfarrarjeh A, Kim SH, Agrawal S, Ashok M, Kim SY, Shahabi C. Image classification to determine the level of street cleanliness: A case study. In: *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM).2018*; pp. 1–5.
24. Ghildiyal A, Sharma S, Kumar A, et al. Street cleanliness monitoring system using deep learning. In: *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).2021*; pp. 868–873.
25. Exsilio Blog: Accuracy, precision, recall & F1 score: Interpretation of performance measures. <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>. 2016.
26. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2016*; pp. 2818–2826.
27. Jayasinghe L, Wijerathne N, Yuen C. A deep learning approach for classification of cleanliness in restrooms. In: *2018 International Conference on Intelligent and Advanced System (ICIAS).2018*; pp. 1–6.
28. Ojala R, Kinnunen T, Aakko M, Mattila J, Kiviluoma P, Kuosmanen P. Monitoring cleanliness of public transportation with computer vision. In: *Baltic Mechatronics Symposium. 2020*.
29. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. Ssd: Single shot multibox detector. In: *European Conference on Computer Vision.2016*; pp. 21–37.
30. Cuhadar C, Lau GPS, Tsao HN. A computer vision sensor for efficient object detection under varying lighting conditions. *Adv Intel Syst.* 2021;3(9):2100055.
31. Fu Q, Hou Y-L, Hao X, Shen Y, Zhang L. On-road vehicle detection under varying lighting conditions. In: *2018 IEEE International Conference on Information and Automation (ICIA).2018*; pp. 1454–1458.
32. LLC U. Vehicle interior air quality: addressing chemical exposure in automobiles. UL LLC: Northbrook, IL, USA; 2015.
33. Faber J, Brodzik K. Air quality inside passenger cars. *AIMS Environ Sci.* 2017;4(1):112–33.
34. Zulauf N, Dröge J, Klingelhöfer D, Braun M, Oremek GM, Groneberg DA. Indoor air pollution in cars: an update on novel insights. *Int J Environ Res Public Health.* 2019;16(13):2441.
35. Barnes NM, Ng TW, Ma KK, Lai KM. In-cabin air quality during driving and engine idling in air-conditioned private vehicles in Hong Kong. *Int J Environ Res Public Health.* 2018;15(4):611.
36. Wolkoff P. Volatile organic compounds. *Indoor Air.* 1995;3:1–73.
37. Gladyszewska-Fiedoruk K. Concentrations of carbon dioxide in the cabin of a small passenger car. *Transport Res Part D Transport Environ.* 2011;16(4):327–31.
38. EPA United States Environmental Protection Agency: Volatile Organic Compounds' Impact on Indoor Air Quality. <https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality>. 2022; Accessed 03 Jan 2023.
39. Szczurek A, Maciejewska M. Classification of air quality inside car cabin using sensor system. In: *SENSORNETS.2015*; pp. 211–219.
40. Szczurek A, Maciejewska M. Categorisation for air quality assessment in car cabin. *Transport Res Part D Transport Environ.* 2016;48:161–70.

41. Moreno T, Pacitto A, Fernández A, Amato F, Marco E, Grimalt JO, Buonanno G, Querol X. Vehicle interior air quality conditions when travelling by taxi. *Environ Res.* 2019;172:529–42.
42. Tille T. Automotive suitability of air quality gas sensors. *Sens Actuators B Chem.* 2012;170:40–4.
43. Yang J, Chen Y, Liu Y, Makke O, Yeung J, Gusikhin O, MacNeille P. The effectiveness of cloud-based smart in-vehicle air quality management. In: 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC).2016; pp. 325–329.
44. Miletiev R, Damyanov I, Iontchev E, Yordanov R. Smart in-vehicle environment monitoring system. In: 2020 XXIX International Scientific Conference Electronics (ET). 2020; pp. 1–4.
45. European Commission|CORDIS (Horizon 2020): High-performance filter to reduce in-car air pollution (2018). <https://cordis.europa.eu/article/id/244886-highperformance-filter-to-reduce-incar-air-pollution>. Accessed 03 Jan 2023.
46. ISO-AIRE): What is a HEPA filter and how does a HEPA filter work?; 2022.
47. Gulli A, Kapoor A, Pal S. Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API. Packt Publishing Ltd, 2019.
48. Bradski G, Kaehler A. *OpenCV. Dr Dobb's journal of software tools.* 2000;3:2.
49. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. 2014; pp. 740–755.
50. Nagpal A. L1 and L2 regularization methods. *Towards Data Sci.* 2017; 13.
51. Szandala T. Review and comparison of commonly used activation functions for deep neural networks. In: Bio-inspired Neurocomputing.2021; pp. 203–224.
52. Liu Y, Gao Y, Yin W. An improved analysis of stochastic gradient descent with momentum. *Adv Neural Informat Process Syst.* 2020;33:18261–71.
53. Sensirion, AG: Datasheet SGP40 Indoor Air Quality Sensor for VOC Measurements. Sensirion. AG version. 2022;1:2.
54. Wu P. Analysis of the wireguard protocol. Master's Thesis, Analysis of the WireGuard protocol, Eindhoven University of Technology. 2019.
55. Jayawickrama N, et al. Detecting trash and valuables with machine vision in passenger vehicles. 2020.
56. Umwelt Bundesamt: Volatile Organic Compounds (VOC); 2018.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
