



# Smart Tramway Systems for Smart Cities: A Deep Learning Application in ADAS Systems

Marco Guerrieri<sup>1</sup> · Giuseppe Parla<sup>1</sup>

Received: 16 December 2021 / Revised: 28 July 2022 / Accepted: 2 August 2022 / Published online: 27 August 2022  
© The Author(s) 2022

## Abstract

Artificial intelligence and deep learning-based techniques undoubtedly are the future of Advanced Driver-Assistance Systems (ADAS) technologies. In this article is presented a technique for detecting, recognizing and tracking pedestrians, vehicles and cyclists along a tramway infrastructure in a complex urban environment by Computer Vision, Deep Learning approaches and YOLOv3 algorithm. Experiments have been conducted in the tramway Line 2 “Borgonuovo –Notarbartolo” (Palermo, Italy) in correspondence of the tramway segments crossing a roundabout having an external diameter of 24 m. A survey vehicle equipped with a video camera was used in the study. The results of the research show that the proposed method is able to search and detect the position and the speed of road users near and over the rails in front of the tram in a very precise way as demonstrate by the estimated values of the Accuracy, Loss and Precision obtained during the neural networks training process. The implementation of this advanced detection method in ADAS systems may increase the safety of novel autonomous trams and autonomous rapid trams (ARTs).

**Keywords** Tramway · ADAS · Deep learning · YOLOv3 · Road safety

## 1 Introduction

Various types of Advanced Driver Assistance Systems (ADAS) have been developed for safe driving of different types of private and public vehicles. In 2014 the Society of Automotive Engineers International (SAE) classified 6 levels of driving automation, from the most basic systems (level 0) to 100% autonomous driving one (level 6) [1]. Several research on the ADAS demonstrate that passive forward collision warning (FCW) and intelligent speed assistance (ISA) systems have been efficacious in reducing the number of pedestrian/vehicle collisions [1–3]. ADAS can be defined as a vehicle-based intelligent safety system which may well increase road safety in terms of crash avoidance, crash severity mitigation and user’s protection during post incident phases. The most well-known ADAS systems include several technologies such as:

- Collision Warning (i.e.: Blind Spot Warning; Forward Collision Warning; Lane Departure Warning; Parking Collision Warning; Rear Cross Traffic Warning);
- Collision Intervention (i.e.: Automatic Emergency Braking; Automatic Emergency Steering; Reverse Automatic Emergency Braking (AEB));
- Driving Control Assistance (i.e.: Adaptive Cruise Control (ACC); Lane Keeping Assistance; Active Driving Assistance);
- Parking Assistance (i.e.: Active Parking Assistance; Remote Parking Assistance; Trailer Assistance);
- other driver assistance systems (Automatic High Beams; Driver Monitoring; Head-Up Display; Night Vision).

Currently ADAS systems are largely used in cars and trucks. Autonomous buses (or self-driving, driverless buses and automated shuttles) are in the experimental stage in numerous cities worldwide thanks to their high automation levels [2–4]. Experiments have been conducted in controlled zones such as the university campus, parking areas and small villages [4]. Autonomous vehicles can be classified as follows: (1) private autonomous cars, (2) shared autonomous cars/taxis, (3) autonomous buses and trucks; (4) autonomous Trams (5); autonomous-rail rapid tram (ART) [4–7].

---

✉ Marco Guerrieri  
marco.guerrieri@unitn.it  
Giuseppe Parla  
peppeparla@yahoo.it

<sup>1</sup> DICAM, University of Trento, Trento, Italy

The International Association of Public Transport (UITP) defines ATO, ATC, ATP as Grades of Automation (GoA). Each GoA is identified by the operational responsibilities of the train basic functions either to an automatic system or to man. In particular, the UITP identify the following 5 GoA levels [8]:

According to [9, 10] the grade of automation (GoA) of metro lines can be categorized into four categories (Fig. 1):

- GoA1 (Operation with a driver). The driver of the Tram is actively involved in all the driving activity and the vehicles is equipped with an ATP (Automatic Train Protection system);
- GoA2 (Semi-automatic Train Operation (STO)). The driver is involved in driving only if a failure occurs, and is responsible for opening and closing the doors. The Tram is equipped with ATP and ATO (Automatic Train Operation system);
- GoA3 (Driverless Train Operation (DTO)). The Tram does not require a driver. An attendant is answerable for the opening and closing of the doors and intervenes in the event of failures. Tram is equipped with ATP and ATO systems.
- GoA4 (Unattended Train Operation (UTO)). The Tram does not require a driver and attendant; it is equipped with ATP and ATO systems.

Similar classification can be used in case of tramway systems. Consequently, trams operation can be assumed to be automatic if the vehicles are driverless (GoA4 and GoA3).

Driverless systems applied to public transport mode may improve capacity, efficiency and safety and decrease the operational costs (lower operation personnel cost) and the road congestion. In particular, the autonomous buses and autonomous-rail rapid trams could be characterized by demand-driven schedules and therefore could potentially dynamically regulate their trajectory, capacity and stopping in accordance with users demand. Several researches have shown positive attitudes of passengers towards autonomous buses [4, 11–13]. Some Adaptive Autonomous vehicles as ART use ACC in order to increase the transportation system capacity. In operating conditions ACC allows small distance between each vehicle pair of a fleet.

In general, the main purpose of autonomous systems is to safely drive a vehicle without human supervision. For long time, researchers have proposed several artificial vision algorithms for implementing automatic systems into private or public vehicles but many problems including changes in obstacles, lighting, background and speed turns this technical approach into a difficult task with no simple solution [14].

The main problem in the ADAS systems is the obstacle detection.

Grade of automation	Type of tram operation	Setting Tram in motion	Stopping Tram	Door Closure	Operation in event of disruption
GoA 1	ATP with driver	Driver	Driver	Driver	Driver
GoA 2	ATP and ATO with driver (STO)	Automatic	Automatic	Driver	Driver
GoA 3	Driverless (DTO)	Automatic	Automatic	Tram attendant	Tram attendant
GoA 4	UTO	Automatic	Automatic	Automatic	Automatic

Fig. 1 Classification of grade of automation (adapted from [9, 10])

Nowadays several obstacle detection technologies are used in this field, based on sensors that allow obtaining several information including the presence of pedestrians, cyclists and vehicles along the way. The most common Detection systems can be classified as follows [15]: 1) Active detection that uses radar systems; 2) Passive detection founded on visual detection technology (the vehicle is equipped with a CCD camera in order to obtain a large amount of information directly from the environment).

In smart cities of the near future, there will be many types of autonomous vehicles, including trams. ADAS systems operating on trams will improve the safety of public transport systems and thus the safety of the entire mobility system of the smart city.

Unlike the ADAS operating on cars, in tramway systems the ADAS should identify and classify only the anterior obstacles, this is because the trajectories of tram vehicles are imposed by the rails. In addition, for each time instant, ADAS should calculate the Emergency braking curve EBD that describes braking distance in function of the instantaneous speed of trams [16]. Over the years a large number of accidents resulting in many fatalities and a high social economic cost have occurred due to human factors-related problems [17].

After many decades of stagnation, traditional and innovative tramway systems (i.e. catenary-free systems) are currently experiencing a period of strengthening with a notable expansion in new railway lines [18, 19]. In fact, since 2000, about 130 new tramway systems have been constructed in the world and many other existing systems have undergone deep renovations [1]. Nowadays, modern trams enjoy great popularity because they are characterized by low environmental impact, adequate construction and maintenance costs, and high transport capacity. At the same time, the number of people using cycling and other types of micro-mobility systems is increasing in cities, which causes a greater number of dangerous situations [20]. In urban contexts walking is a fundamental mode of transportation. Nevertheless, more than 7000 pedestrians were killed each year in the European Union, equal to 27% of all traffic victims [21]. According to [18, 22] some of the most important hazards in tram traffic are: 1) Collision between trams, 2) Tram collision with other types of vehicles, cyclists or pedestrians, 3) Running into an obstacle on the tracks. In traditional tramway systems the train driver must scan the approaching physical environment for the presence of hazards, warnings and imposed speed limits [23]. In order to reduce such hazards are required several actions both from the transportation system (operators, train drivers, infrastructure management, etc.) and the road users.

The introduction of Self-driving technologies could increase the quality of service, the system reliability and the safety of these transportation systems [24] both in

traditional and smart cities. A Self-driving vehicle necessitates of a multimodal suite of sensors (e.g. dual antenna GNSS-aided Inertial Navigation System (INS), Lidar, Radar and Cameras) to perform the localization, signal handling, and obstacle handling tasks [25]. In this direction, Bombardier, Alstom, CRRC, Siemens and other Companies have already carried out practical experimentations [24, 25]. Nevertheless, the complex traffic environment and the correlated hazards make the realization of unmanned vehicles problematic in tramway systems. Many researches and experiments are required above all in the field of object recognition and tracking [26]. Presently, in-vehicle detection technology is widely applied to the smart car field and it is divided into active detection and passive detection [6].

This research describes a technique for pedestrians, vehicles and cyclists recognition along a tramway infrastructure by Computer Vision and deep learning approaches. The proposed technique guarantees high accuracy and precision in object detection and therefore could be used in advanced driver assistance systems or in autonomous trams and autonomous rapid trams (ARTs) in order to achieve a high level of safety. The paper is structured as follows. Section 2 presents an overview of relevant literature of object detection and recognition systems based on deep learning techniques and YOLOv3 system. Section 3 briefly explains the proposed technique applied to a case study in urban context. Section 4 deals with the neural network training; Sect. 5 presents the model for the estimation of the distance between pedestrians and rails; instead the main results are presented in Sect. 6. Finally, conclusions are proposed in Sect. 7.

## 2 Methodology for the Detection of Road Users

The detection of pedestrians, cyclists and other road users play a key role in automatic driving emerging technologies. In fact, errors in the detection of road users could threaten users' lives [27]. Therefore, the performance of road users detection algorithms is of great importance in the field of autonomous vehicles.

This research presents a methodology based on computer vision, deep learning and Yolov3 algorithms applied to the recognition along a tramway infrastructure of pedestrians, vehicles and cyclists.

In accordance with the scientific literature, object detection algorithms are mainly subdivided into two classes: the one-stage method [8, 28, 29] and the two-stage method [30, 31]. One-stage method classifies the object and renders a regression of the object location from the raw image. Instead, in the two-stage method the detection process requires the following phases: 1) extraction the region of interest (ROI) from an image

where the objects of interest may be, and 2) correction and recognition of the candidate region [32]. YOLO is a typical one-stage detection algorithm, which was originally proposed by Redmon [8, 28] to achieve end-to-end target detection based on a single CNN model. Nowadays Object detection systems like YOLO (You Only Look Once), SSD (single-shot detector), and Faster R-CNN, not only classify images but also can locate and detect each object in images that contain multiple objects [33].

The structure of YOLO is composed of two different parts [8]: the first one is the feature extraction network that allows the general features of the researched object [34]. The second part of YOLO is a post-processing network which aims to evaluate the coordinates and the categories of the object under analysis. YOLOv3 is one of the latest versions of YOLO networks. It does not require a region proposal network (RPN) and directly performs regression to detect targets in the image [35].

The YOLOv3 algorithm uses a new network for extracting the object features. YOLOv3 includes 53 convolutional layers (Darknet-53) and 23 residual layers as displayed in Figs 2. YOLOv3 is characterized by a remarkable advancement in real-time object detection, particularly in the detection of smaller objects. In addition, YOLOv3 uses a multi-label classification. Therefore, in this research YOLOv3 is chosen as the detection system for vehicles, pedestrians and other road users along the analysed tramway line. As well explained in [35],  $1 \times 1$ ,  $3 \times 3/2$ , and  $3 \times 3$  convolution kernels of three sizes are applied in the convolutional layers to sequentially extract image features, ensuring

the model has remarkable classification and detection performances.

The remaining layers guarantee the convergence of the detection model [35]. For detecting the parts of the object of interest at the same time, YOLOv3 fuses three feature maps of different scales ( $52 \times 52$ ,  $26 \times 26$ , and  $13 \times 13$ ) by three times of downsampling. In conclusion, the YOLO algorithms family (i.e. YOLOv1, YOLOv2, YOLOv3) is a series of end-to-end deep learning models planned and designed for fast real-time object detection [33].

In YOLOv3 algorithm, the Loss function is composed by the following parts: classification loss, localization loss (errors between the predicted boundary box and the ground truth) and confidence loss (the objectness of the box) [37] as follows:

classification loss: if an object is detected, the classification loss at each cell is the squared error of the class conditional probabilities for each class [37]:

$$\sum_{i=0}^{S^2} I_{ij}^{obj} (p_i(c) - \hat{p}_i(c))^2 \quad (1)$$

where

$I_{ij}^{obj} = 1$ : if an object appears in the cell  $i$ , otherwise is 0;

$\hat{p}_i(c)$ : denotes the conditional class probability for class  $c$  in the cell  $i$ .

Localization loss: evaluates the errors in the predicted boundary box locations and sizes. It is only counted the box responsible for detecting the object [37]:

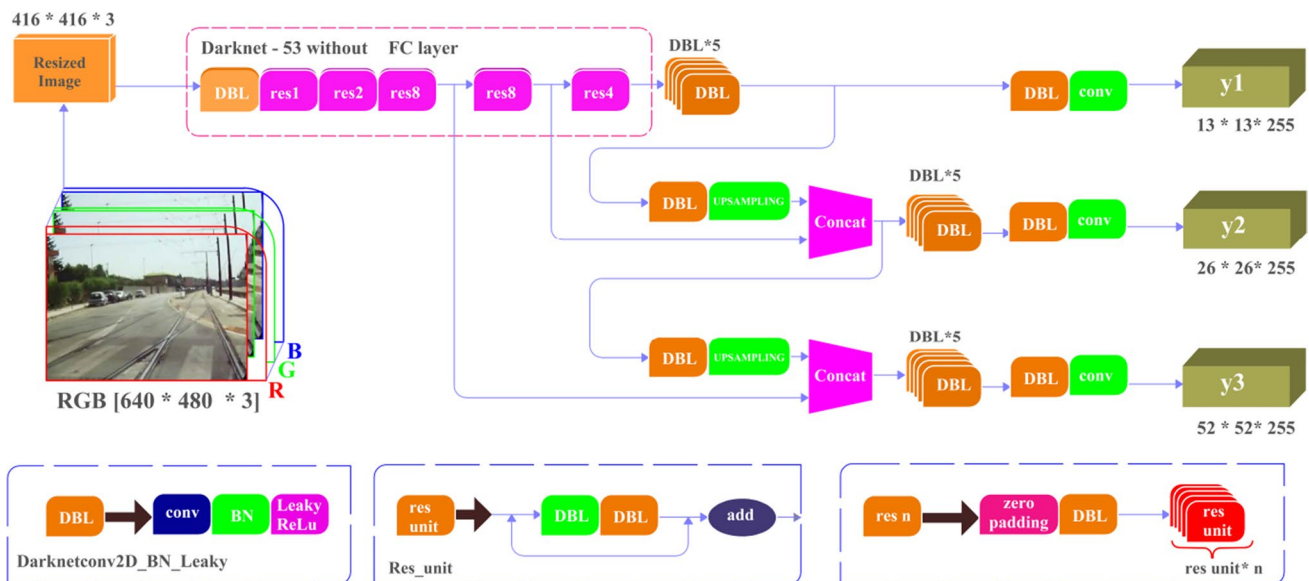


Fig. 2 YOLOv3 Network structure (adapted from [36])

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[ (x - \hat{x}_i)^2 + (y - \hat{y}_i)^2 \right] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[ \left( \sqrt{\overline{w}_i} - \sqrt{\widehat{w}_i} \right)^2 + \left( \sqrt{\overline{h}_i} - \sqrt{\widehat{h}_i} \right)^2 \right] \tag{2}$$

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e_{t_w} \\ b_h = p_h e_{t_h} \end{cases} \tag{6}$$

where

$I_{ij}^{\text{obj}} = 1$ : if in the  $j$ th boundary box in cell  $i$  is responsible for detecting the object, otherwise 0;

$\lambda_{\text{coord}}$ : increase the weight for the loss in the boundary box coordinates.

Confidence loss: if an object is detected in the box, the confidence loss (measuring the objectness of the box) is [37]:

$$\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (C - \widehat{C}_i)^2 \tag{3}$$

where

$\widehat{C}_i$ : stands for the box confidence score of the box  $j$ , in the cell  $i$ .

$I_{ij}^{\text{obj}} = 1$ : if the  $j$ th boundary box in the cell  $i$  is responsible for detecting the object, otherwise 0.

If an object is not detected in the box, the confidence loss is [37]:

$$\lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} (C - \widehat{C}_i)^2 \tag{4}$$

where  $\lambda_{\text{noobj}}$ : is the complement of  $I_{ij}^{\text{obj}}$ ;  $\widehat{C}_i$ : is the box confidence score of the box  $j$  in the cell  $i$ ;  $\lambda_{\text{noobj}}$ : weights down the loss when detecting background.

The final loss adds localization, confidence and classification losses together [37], as follows:

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[ (x - \hat{x}_i)^2 + (y - \hat{y}_i)^2 \right] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[ \left( \sqrt{\overline{w}_i} - \sqrt{\widehat{w}_i} \right)^2 + \left( \sqrt{\overline{h}_i} - \sqrt{\widehat{h}_i} \right)^2 \right] + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (C - \widehat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} (C - \widehat{C}_i)^2 + \sum_{i=0}^{S^2} I_{ij}^{\text{obj}} (p_i(c) - \widehat{p}_i(c))^2 \tag{5}$$

In YOLO v3 the method for predicting the bounding box is given by Eqs. (6):

In which  $t_x$ ,  $t_y$ ,  $t_w$ , and  $t_h$  are the predicted outputs of the model, which denote the relative position coordinates of the center of the bounding box and the relative width and height of the bounding box.  $c_x$  and  $c_y$  denote the net, instead  $p_w$  and  $p_h$  are the width and height of the predicted front bounding box. Finally  $b_x$ ,  $b_y$ ,  $b_w$ , and  $b_h$  are the true coordinates of the center of the bounding box, the true width and height of the bounding box obtained after prediction (Fig. 3).

The performance of a certain object detector can be measured mainly by the following metrics:

Frame per second (FPS) to measure detection speed (number of images processed every second);

Precision-recall curve (PR curve) in which Precision and Recall are calculated as follows:

$$\text{Re call} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Pre cision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \cdot 100 \tag{9}$$

The symbols TP, FN and FP stand for True Positive, False Negative and False Positive respectively.

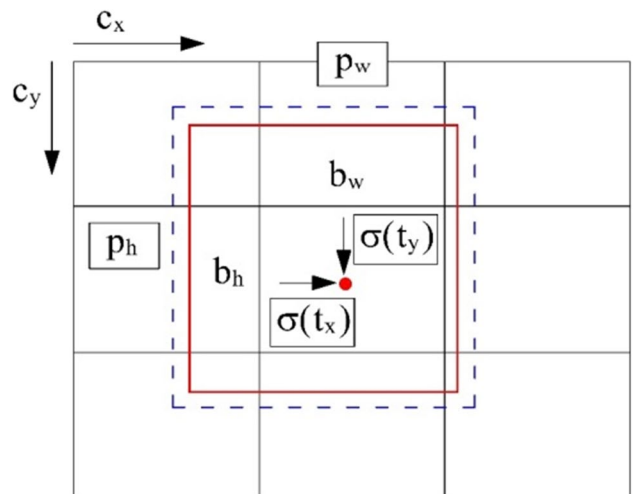


Fig. 3 Bounding Box with dimensions priors and location prediction (source: [8, 28])



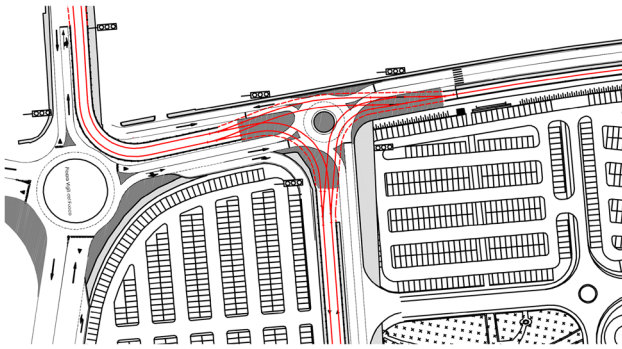


Fig. 4 Roundabout analyzed (in red the tramway tracks)

### 3 The Case Study

The city of Palermo (Italy) covers a territory of 158.88 km<sup>2</sup> with a population of 666,992 inhabitants. Apart from the capital, the metropolitan area includes 26 towns, with a total population of around 1 million inhabitants. The following tramway lines have been running since 2015:

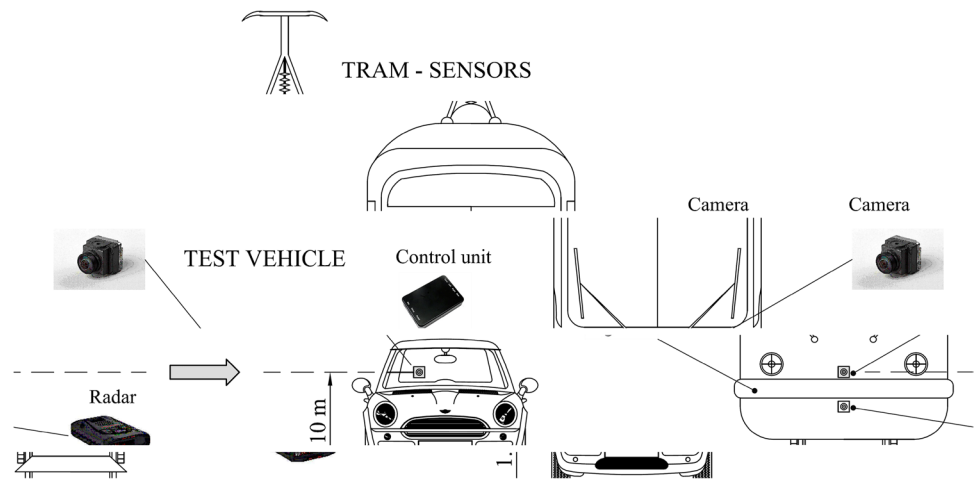
- Line 1 “Roccella”: about 5.5 km long, with double tracks;
- Line 2 “Borgonuovo –Notarbartolo”: 4.8 km, with double tracks;

- Line 3 “C.E.P. – Notarbartolo”: 5 km, with double tracks;
- Line 4 “Notarbartolo-Calatafimi-Notarbartolo”: 8 km with single track.

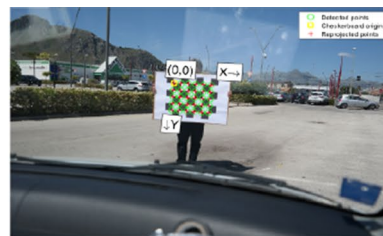
The municipality expects to extend the tramway system with seven new lines called “A, B, C, D, E, F, G”, with a total of more than 68 km.

The experiments of this research have been conducted into the tramway Line 2 “Borgonuovo –Notarbartolo” in the city of Palermo (Italy) and more precisely in the tree-arms roundabout of Fig. 4 crossed by different tramway sections. From the geometric point of view, the roundabout is characterized as follows: external diameter: 24 m; width of circulating carriageway: 6 m; entry lanes width: 3.50 m; exit lanes width: 3.50 m. In the first phase of the research, Deep learning and Yolov3 algorithms are used for detecting vehicles and pedestrians in the tramway track space (i.e. road users that cross or travel along the track). For this purpose several runs of a survey vehicle (Fig. 5) were made traveling in the roundabout, following the tramway track. The detection processes are investigated via a video camera installed in the survey vehicle (Fig. 6a). Traffic video recordings, with the resolution of 1280×720 pixels, were analyzed using a workstation with Intel(R) Core(TM) i7-4510 CPU @ 2.00 Hz 2.60 GHz – Memory RAM 20 Gb, Windows 10 Home. The

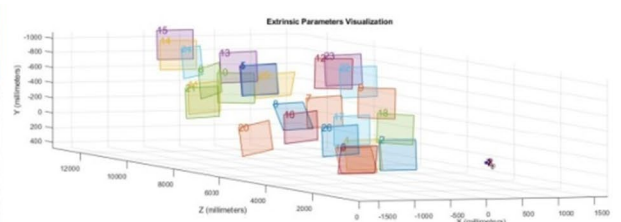
Fig. 5 Hypothetical Tram sensors and test vehicle used in the research



(a)



(b)



(c)

Fig. 6 Camera calibration procedure

first research phase was the Camera calibration. Camera calibration aims at establishing two sets of parameters: intrinsic and extrinsic. The intrinsic parameters are related to the focal length and optical center. The extrinsic parameters (pitch angle, yaw angle, and height above ground) determine the spatial offset of the camera. The Zhang algorithm [38] determines the extrinsic parameters of the system. For the calibration processes 64 different images of the chessboard of Fig. 6b have been considered using a set of different photos. Figure 6c shows the Extrinsic Parameters visualization obtained by the calibration processes. The calibrate model was then validated by several tests concerning the comparison between estimated and measured values of tramway gauge (see Sect. 5.1).

### 4 Training of the Neural Networks

As explained in previous Sect. 2 YOLOv3 includes 53 convolutional layers, hence the name Darknet-53. In Darknet-53 the weights of the custom detector are saved for every 100 until 1000 iterations, and it continues to be saved for every 10,000 iterations until it reaches the maximum batches [39]. The pre-existing road users dataset (including light and heavy vehicles, motorbikes, pedestrians and other road users) in Darknet-53 was split into 75% for training and 25% for testing.

In the course of the training process phase, data augmentation procedures (cropping, padding, flipping, etc.) were applied with the aim to prepare the large neural networks. Before, a bounding box labelling tool [40] has been applied to manually detect and recognize road users (vehicles and pedestrians) for the object to be detected [39] in the tramway track space. The outcomes of this phase and the class label consist of four points of the position coordinate. The label is converted into YOLO format and the tool changes the values to a format which the training algorithm YOLOv3 can employ. Figure 7 shows the training process consisting of 2000 iterations. The Accuracy (cf. Equation 9 and Fig. 7),

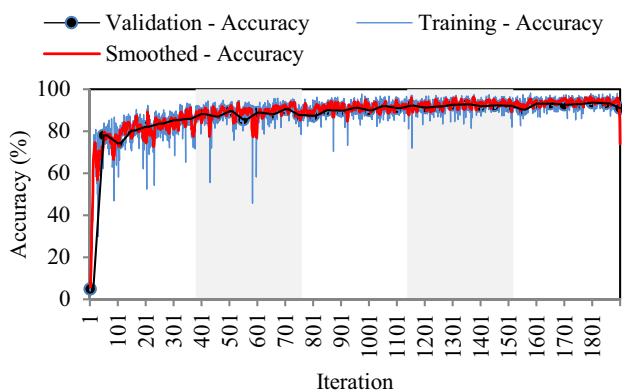


Fig. 7 Accuracy of the pre-trained networks

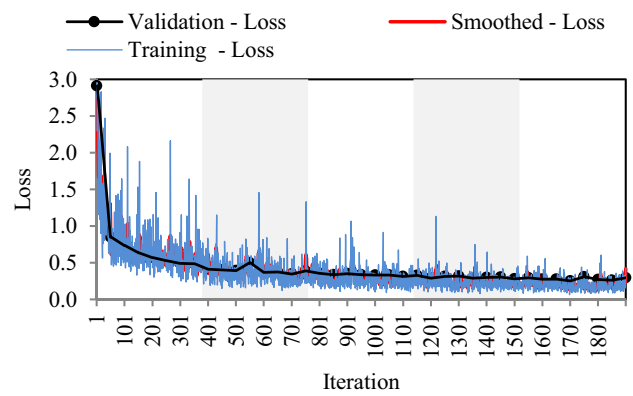


Fig. 8 Loss function of the pre-trained networks

the Loss (cf. Equation 5) and the Precision (Eq. 8 and Fig. 8) values demonstrate that the proposed training model detects the pedestrian and the vehicles with high accuracy.

## 5 Estimation of the Distance Between Pedestrians and Rails

### 5.1 Tramway Track Detection

The rails boundaries detection and tracking are based on the algorithm carried out by Nieto [41] for the detection of lane markings in roads and highways. In addition, a Top-hat filter was used to correct the non-uniform brightness of the original image [42].

Let  $I$  be the original image under analysis, the “Top-hat” algorithm can be written as follows:

$$I_t = (I - I \odot B) \tag{10}$$

Being  $\odot$  the binary morphology operator that denotes a morphological opening operation and  $B$  is the structured element, which is a matrix composed of binary values (0 or 1) and  $I_t$  is the filtered image. In this research the structured element is a circle with a radius of a 5 pixel. After having determined the filtered image, the algorithm determines its complement. The difference is used as the pixel value in the output image. The following types of elements are taken into account within any plane image: Pavement, tramway rails, Objects (i.e. vehicles, pedestrians and other road users). The object segmentation is based on the Bayesian decision theory according to the conditions given below. Let  $S = \{P, R, O, U\}$  be the set of classes that characterize, respectively, the pavement, the rails of the tramway track, the objects (i.e. vehicles, pedestrians and other road users) and the unidentified elements.

The main target of the classifier is to assign one of the classes {P, R, O, U} to each pixel of the image [42]. According with [41] we denote with  $X_i$  the event that a pixel whose coordinates are  $(x, y)$  and with an associated observation vector denoted as  $z_{xy}$  is classified as belonging to the class  $i \in S$ . By means of the Bayesian decision theory, the classification is obtained by selecting the class that maximizes the a posteriori conditional probability  $P(X_i|z_{xy})$  as follows:

$$P(X_i|z_{xy}) = \frac{p(z_{xy}|X_i)P(X_i)}{P(z_{xy})} \tag{11}$$

Being  $p(z_{xy}|X_i)$  the probability that a pixel belongs to class  $i$ ;  $P(X_i)$  the prior probability of each class and  $P(z_{xy}) = \sum_{i \in S} P(z_{xy}|X_i)P(X_i)$  a scale factor that ensures that the posteriors sum to unity. Then, it can be estimated the probability that a pixel belongs to a certain class.

Therefore, adopting the Likelihood models and an Expectation–Maximization (EM) algorithm, the segmented tramway rails (Boolean image) are obtained.

After the determination of the Boolean image of the tramway rails (Fig. 11d) the proposed procedure needs to match the estimated alignment of the rails with respect to the real alignment. In order to fit the model to the real geometric alignment of the track, the RANSAC “Random Sampling Consensus” model [43] was implemented. RANSAC is an iterative technique used to estimate the parameters of a mathematical model from a set of observed data that contains a large proportion of outliers in the input data. In short, RANSAC is a resampling technique that produces realistic solutions by means of the minimum number of observations which are required to estimate the underlying model parameters [44]. In RANSAC the number of iterations  $N$  is selected to ensure that the probability  $p$  (usually set to 0.99) that at least one of the sets of random samples does not include an outlier:

$$N = \frac{\log(1 - p)}{\log[1 - (1 - v^m)]} \tag{12}$$

In which  $v$  represents the probability of finding an inlier in any selected point,  $m$  is the minimum number of points required to estimate a model and they are selected independently. In case of a parabolic curve in the planimetric tramway alignment, the proposed algorithm estimates the coefficients  $[a, b, c, d]$  of the transition curve whose equation is  $y = a \cdot x^3 + b \cdot x^2 + c \cdot x + d$  from the candidate coordinates (input data). In addition this procedure allows validating the camera calibration phase. In fact, it can be estimated the Euler Angle of

the camera by comparing a sample of estimated values of gauge of the tramway track with respect to the real value (1435 mm). Through an iterative procedure, the Euler angles are modified until the difference between the estimated values and the real ones becomes less than 1 cm. Fig. 9 explains in a synthetic way the proposed process for camera calibration and validation.

### 5.2 Top view Image Transformation

In each instant of time the methodology allows determining the distance of road users from the rails and therefore to identify the critical conditions that require emergency braking of the Tram. The distance considered is that of the users' centroid projected on the ground plane (road pavement surface) with respect to the nearest rail [46]. The procedure for the distance estimation is specified below and it is founded on the inverse perspective mapping (IPM). Let  $\{F_w\} = \{X_w, Y_w, Z_w\}$  be the world frame centered at the camera optical center,  $\{F_c\} = \{X_c, Y_c, Z_c\}$  the camera frame and  $\{F_i\} = \{u, v\}$  an image frame (Fig. 10). It is assumed that the optical axis has no roll; in other words, the camera frame  $X_c$  axis stays in the world frame  $X_w Y_w$  plane. The height of the camera frame with respect to the ground plane is  $h$  (Fig. 10).

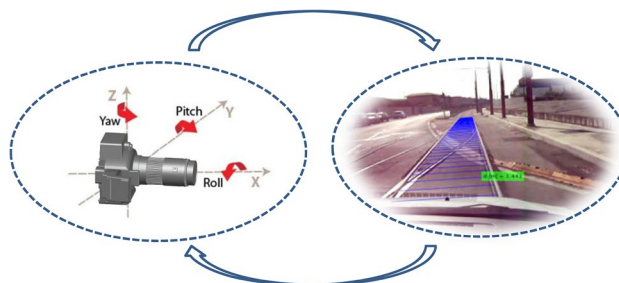


Fig. 9 Camera calibration and validation schematic procedure

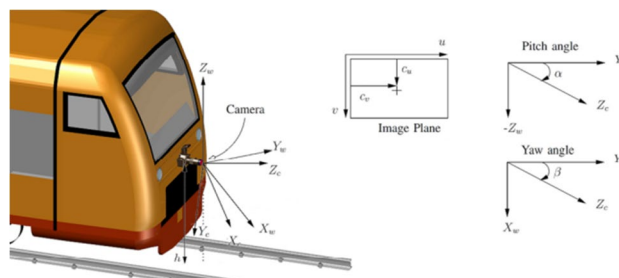


Fig. 10 IPM coordinates. Left: coordinate axes (world). Right: definition of pitch and yaw angles



Using the following homogeneous transformation can be found the projection point on the road plane of each point  ${}^iP = \{u, v, 1, 1\}$  in the image plane [46]:

$${}^i_gT = h \begin{bmatrix} -\frac{1}{f_u}c_2 & -\frac{1}{f_v}s_1s_2 & \frac{1}{f_u}c_u c_2 - \frac{1}{f_v}c_v s_1 s_2 - c_1 s_2 & 0 \\ \frac{1}{f_u}s_2 & \frac{1}{f_v}s_1 c_1 & -\frac{1}{f_u}c_u s_2 - \frac{1}{f_v}c_v s_1 c_2 - c_1 c_2 & 0 \\ 0 & \frac{1}{f_v}c_1 & -\frac{1}{f_v}c_v c_1 + s_1 & 0 \\ 0 & -\frac{1}{hf_v}c_1 & \frac{1}{hf_v}c_v c_1 - \frac{1}{h}s_1 & 0 \end{bmatrix} \quad (13)$$

In which:  $\{f_u, f_v\}$ : are the horizontal and vertical focal lengths;  $\{c_u, c_v\}$ : are the coordinates of the optical center;  $c_1 = \cos\alpha$ ,  $c_2 = \cos\beta$ ,  $s_1 = \sin\alpha$ , and  $s_2 = \sin\beta$ .

On the contrary from a point on road plane  ${}^gP = \{x_g, y_g, -h, 1\}$  we can obtain its subpixel coordinates on the image frame by the relationship  ${}^iP = {}^gT {}^gP$  using the inverse of the transform [46]:

$${}^gT = \begin{bmatrix} f_u c_2 + c_u c_1 s_2 & c_u c_1 c_2 - s_2 f_u & -c_u s_1 & 0 \\ s_2 (c_v c_1 - f_v s_1) & c_2 (c_v c_1 - f_v s_1) & -f_v c_1 - c_v s_1 & 0 \\ c_1 s_2 & c_1 c_2 & -s_1 & 0 \\ c_1 s_2 & c_1 c_2 & -s_1 & 0 \end{bmatrix} \quad (14)$$

With the previous relationships, using the function “EstimateMonoCameraParameters” in MATLAB environment the extrinsic parameters of the camera ( $EulerAngle = \langle \alpha | \beta | \gamma \rangle = \langle pitch | yaw | roll \rangle$ ) can be calculated.

It is worth pointing out that varying the angles “pitch”, “yaw” and “roll” in MATLAB the Bird’s eye view of the analyzed image changes accordingly. The Birds Eye View of the images is of fundamental interest for analyzing the distance between pedestrians and railway tracks. Figure 11 shows a schematic illustration of the top view image transformation [45].

### 5.3 Road Users Tracking

In the tracking phase of pedestrians, vehicles and cyclists, the Kalman filter was implemented with the purpose of tracking the projection on the pavement surface of the centroid of the objects of interest (i. e. pedestrians, vehicles and cyclists).

Kalman filter [47] is a recursive predictive filter able to evaluate the state of a dynamic system. The Linear Kalman Filter is applied for the evaluation of the coordinate of the object of interest. The dynamic equation is [48, 49]:

$$x_{n+1} = A_n x_n + B_n u_n \quad (15)$$

Considering the error covariance [48, 49]:

$$P_{n+1} = A_n P_n A_n^T + Q_n \quad (16)$$

in which  $x_n$  is the state value (object coordinates) at step  $n$ ,  $A_n$  is the state transition matrix,  $u_n$  is the measurement and the input and at step  $n$ .  $Q_n$  is the white noise covariance [49]. This step is called the “prediction step” since it estimates the  $n + 1$  state. Kalman gain value is given by the following relationship [49]:

$$K_n = P_n C^T (C P_n C^T + R_n)^{-1} \quad (17)$$

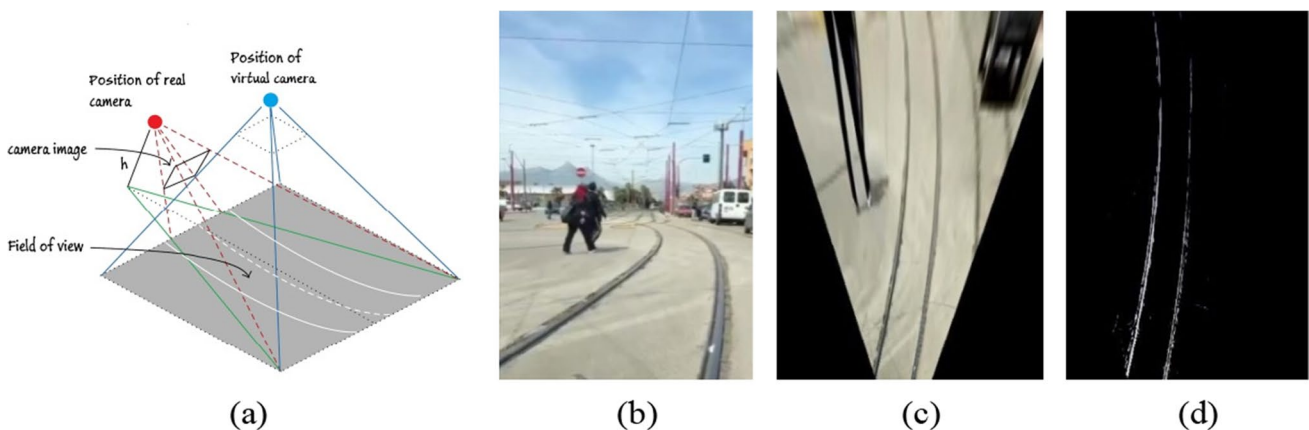
where  $C$  is the measurement matrix and  $R$  is measurement noise.

Actual measurement value at the updated time and error covariance is [49]:

$$P_n = (I - K_n H) P_n \quad (18)$$

Being  $K_n$  the measurement value and  $H$  the mapping matrix from true state to observation.

The initial covariance matrix is a diagonal matrix having high values since the centroids of the Bounding Boxes



**Fig. 11** a) Schematic illustration of the top view image transformation (adapted from [45]), (b) original image of pedestrians near the tramway track, (c) road-plane transformation, (d) Boolean image of the tramway track.

that identify the pedestrians in many frames of the analyzed video sequences often are not always clearly distinguishable. The Kalman filter can be applied to different object motions (i.e. constant velocity and constant acceleration); the amount of deviation from the ideal motion model allows fitting the real motion conditions with respect to the real ones.

Figure 12 shows some examples of pedestrians and vehicles detection near the tramway track under study, instead Fig. 13 gives an example of the Kalman filter application for the evaluation of pedestrians' trajectories. It is worth pointing out that the Kalman filter is able to estimate the trajectory of a pedestrian even if it overlaps with another object in the scene.

## 6 Results

Figure 11c shows an example of IPM: it can be noted the position of a pedestrian with respect to the rails of the tramway track. Therefore, the functions described in the previous sections allow to estimate the distances between the pedestrians and the rails and to evaluate the risk conditions that require emergency braking maneuvers even in situations in which prospectively the trajectories of pedestrians are arranged between them in partial or total occlusion. As shown in Figs. 14, 15, for the pedestrians, the procedure is capable of estimating the speed component orthogonal to the rails when they approach or cross the tramway track (Fig. 14). Similar results can be obtained in case of vehicles, bicyclists, animals, etc.



Fig. 12 Example of pedestrians and vehicles detection near the tramway track



Fig. 13 Results of the Kalman filter application

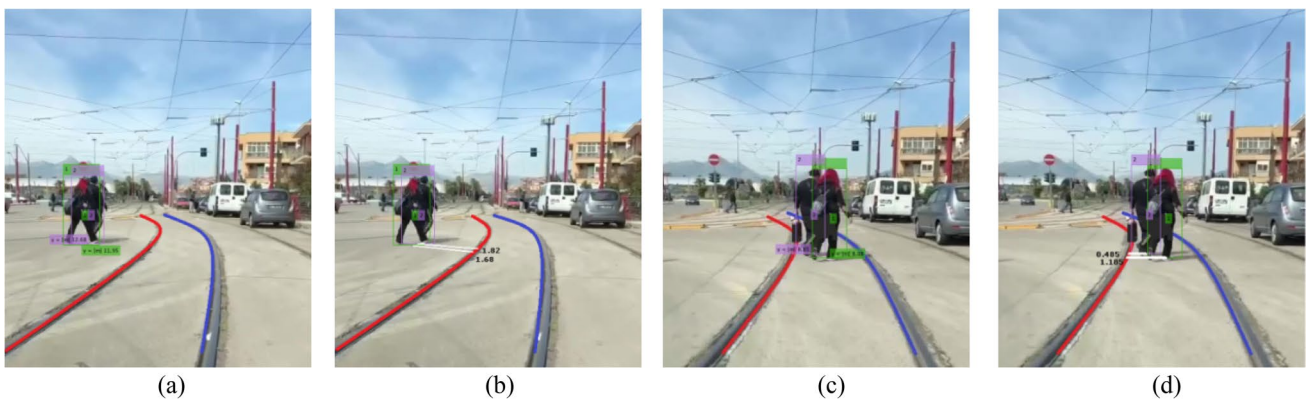
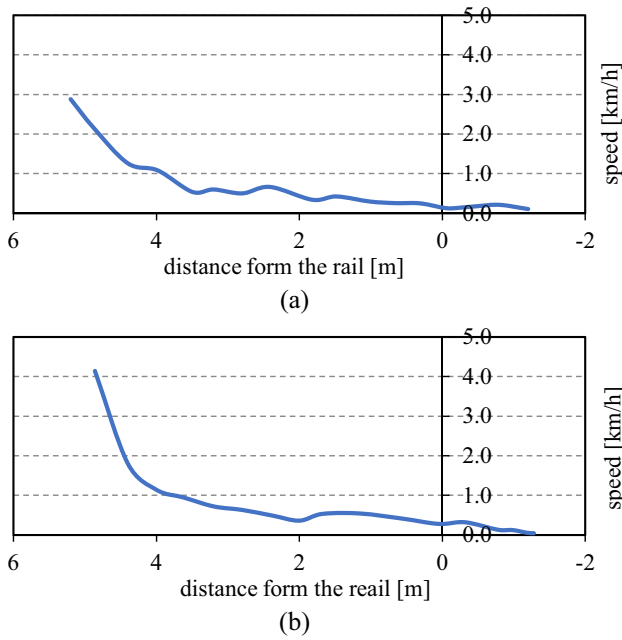


Fig. 14 Examples of detection, tracking and distances measurement of two pedestrians



**Fig. 15** Speed of two pedestrians (cf. Fig. 14): a) pedestrian n. 1; b) pedestrian n. 2

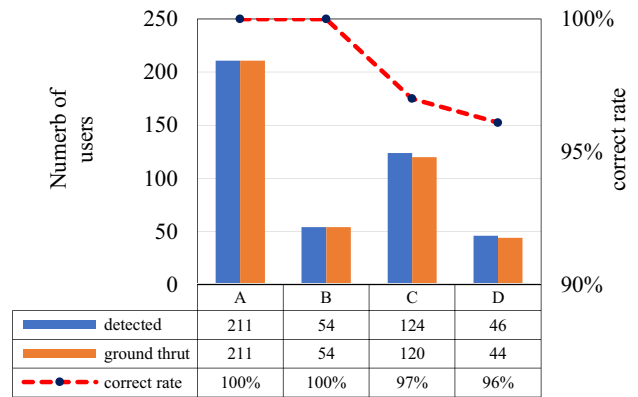
The results show the effectiveness of the proposed algorithm in detecting with high accuracy and precision the tramway track, the road users (pedestrians) and their trajectories and speeds and therefore this method could be used in advanced driver-assistance systems ADAS or in on autonomous trams and autonomous rapid trams (ARTs) in order to achieve a high level of safety.

The proposed method is able to detect and follow the position of each road user using the tracking algorithm.

However, some FP and FN were found during the detection phase of the algorithm [50, 51]. In general, detection errors are associated with a variety of causes, including the camera viewing angle, the test vehicle oscillations, other objects near the tramway line and many others [52]. Consequently, to better appreciate the reliability of the proposed technique, an error analysis was performed by comparing the number of road users detected and the real one. Figure 16 exemplifies the main results of the validation procedure: the correct detection rate ranges from 96.0% to 100% depending on the road user type.

The tracking phase permits the detection of the same vehicle or pedestrian in all the positions which may be shown in a sequence of consecutive video frames. The system successfully identified pedestrians and vehicles in the lateral position with respect to the rails.

Therefore, the speed component of the users orthogonal to the rails can be estimated and by specific safety protection algorithm the collision warning function can be activated.



**Fig. 16** Number of detected road users, total number of real road users and correct detection rate (A: cars, B: heavy vehicles; C: pedestrians, D: cyclists)

The implementation of this advanced detection method based on deep learning together to a collision warning system in ADAS may increase the safety of novel autonomous trams and autonomous rapid trams (ARTs). For such novel types of transportation systems, a Safety Protection Framework could be structured like the one shown in Fig. 17. The implementation of this advanced detection method based on deep learning together to a collision warning system in ADAS.

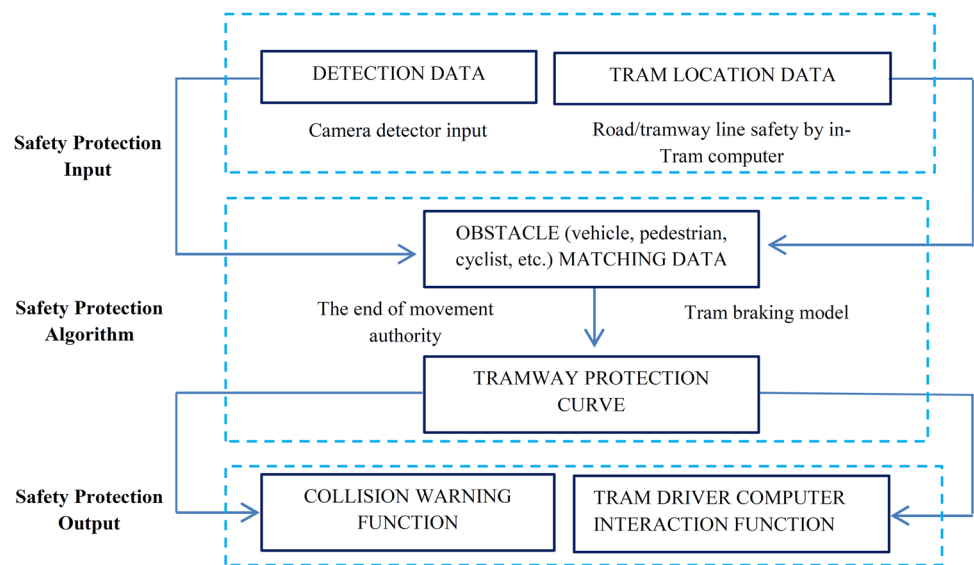
### 7 Conclusions

Artificial intelligence and deep learning-based techniques are the future of Advanced Driver-Assistance Systems (ADAS) technologies. In this article is presented a technique for real-time pedestrians, vehicles and cyclists detection, tracking and recognition along a tramway infrastructure in a complex urban environment by Computer Vision and deep learning approaches. In particular, the YOLOv3 algorithm, RANSAC model and Kalman filter were used.

Experimental activities were conducted in the tramway Line 2 “Borgonuovo –Notarbartolo” in the city of Palermo (Italy) on the segment crossing a Tree-arms roundabout. A survey vehicle equipped with a video camera was used in the research. Traffic video recordings, with the resolution of 1280 × 720 pixels, were analyzed using a workstation with Intel(R) Core(TM) i7-4510 CPU @ 2.00 Hz 2.60 GHz – Memory RAM 20 Gb, Windows 10 Home.

The proposed method is able to search and detect the position of private vehicles, pedestrians and cyclists near and over the rails in front of the tram in a very precise way despite the detection and the tracking of obstacles in complex urban environment have many complications due to the tram speed, the dissemination of forms (blurring) and the

**Fig. 17** Safety Protection Framework



changing of the background which can greatly complicates its accuracy.

In fact, the Accuracy, the Loss and the Precision values obtained with the YOLOv3, during the Neural networks training process prove that the proposed training model detects even the position of several obstacles (vehicles, pedestrians and cyclists) at the same time with high accuracy in each frame. In particular the pedestrian trajectories and the speeds component orthogonal to the rails can be estimated. This type of information is useful in the safety protection algorithm and in the collision warning function.

Although the proposed procedure needs to be validated with a greater number of tests, the first results demonstrate the effectiveness of the proposed algorithm with overall good outcomes (minimum correct detection rate: 96%).

Therefore, the proposed method guarantees high accuracy and precision in object detection and therefore could be used in advanced driver-assistance systems or in autonomous trams and autonomous rapid trams (ARTs) in order to achieve a high level of safety both in traditional and smart cities.

**Funding** Open access funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement. This research was funded by the Projects of National Interest—PRIN 2017 “Stone pavements. History, conservation, valorisation and design” (20174JW7ZL) financed by the Ministry of Education, University and Research (MIUR) of the Italian Government.

## Declarations

**Conflicts of Interest** We have no conflicts of interest to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Di Palma, C., Galdi, V., Calderaro, V., De Luca, F.: Driver Assistance System for Trams: Smart Tram in Smart Cities. IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe, IEEEIC / I and CPS Europe (2020). <https://doi.org/10.1109/IEEEIC/ICPSEurope49358.2020.9160780>
- Ainsalu, J., Arffman, V., Bellone, M., Ellner, M., Haapamäki, et al.: State of the art of automated buses. Sustainability. 10(9) (2018). <https://doi.org/10.3390/su10093118>
- Nordhoff, S., De Winter, J., Payre, W., Van Arem, B., Happee, R.: What impressions do users have after a ride in an automated shuttle? An interview study. Transport. Res. F: Traffic Psychol. Behav. **63**, 252–269 (2019). <https://doi.org/10.1016/j.trf.2019.04.009>
- Mouratidis, K., Cobeña Serrano, V.: Autonomous buses: Intentions to use, passenger experiences, and suggestions for improvement. Transport. Res. F: Traffic Psychol. Behav. **76**, 321–335 (2021). <https://doi.org/10.1016/j.trf.2020.12.007>
- Yuan, X., Zhang, Q., Zhang, S., Huang, R., Zhang, X., Yunqin, H.: Longitudinal control of autonomous-rail rapid tram in platooning using model predictive control. 2020 IEEE Vehicle Power and Propulsion Conference, VPPC 2020 – Proceedings (2020). <https://doi.org/10.1109/VPPC49601.2020.9330878>



6. Blades, L., Douglas, R., Early, J., Lo, C. Y., & Best, R.: Advanced Driver-Assistance Systems for City Bus Applications. SAE Technical Papers (2020). <https://doi.org/10.4271/2020-01-1208>
7. Bösch, P.M., Becker, F., Becker, H., Axhausen, K.W.: Cost-based analysis of autonomous mobility services. *Transp. Policy* **64**, 76–91 (2018)
8. Redmon, J., Divvala, S., Girshick, R., Farhadi A.: You only look once: unified, real-time object detection. in Proc. IEEE Conf. CVPR. 779–788. (2016)
9. UITP: <http://www.uitp.org/sites/default/files/Metro%20automati on%20-%20facts%20and%20figures.pdf> (2013b). Accessed 2 Jan 2022.
10. Pyrgidis C.N.: Railway Transportation systems. CRC Press (2016)
11. Nordhoff, S., De Winter, J., Kyriakidis, M., Van Arem, B., & Happee, R.: Acceptance of driverless vehicles: Results from a large cross-national questionnaire study. *J. Adv. Transp.* 5382192 (2018)
12. Rehr, K., Zankl, C.: Digibus\_: Results from the first self-driving shuttle trial on a public road in Austria. *Eur. Transp. Res. Rev.* **10**(2), 51 (2018). <https://doi.org/10.1186/s12544-018-0326-4>
13. Salonen, A.O.: Passenger's subjective traffic safety, in-vehicle security and emergency management in the driverless shuttle bus in Finland. *Transp. Policy* **61**, 106–110 (2018). <https://doi.org/10.1016/j.tranpol.2017.10.011>
14. Rodriguez, L.A.F., Uribe, J.A., Bonilla, J.F.V.: Obstacle detection over rails using hough transform. STSIVA 2012 - 17th Symposium of Image, Signal Processing, and Artificial Vision. art. no. 6340602, 317–322 (2012)
15. Wang, G., Zeng, X., Bian, D., Wang, W.: Research on modern tram auxiliary safety protection technology based on obstacles detection. *Smart Innov. Syst. Technol.* **62**, 37–50 (2017)
16. Kahlouche, A., Chaib, R.: Anoverview of Constantine's tram safety. *Transport Telecommun.* **18**(4), 324–331 (2017)
17. Dhillon, B.S.: Human Reliability and Error in Transportation Systems, Springer-Verlag, London (2007). <https://doi.org/10.1007/978-1-84628-812-8>
18. Szmagliński, J., Grulkowski, S., Birr, K.: Identification of safety hazards and their sources in tram transport. MATEC Web of Conferences. 231, art. no. 05008 (2018)
19. Guerrieri, M.: Catenary-Free Tramway Systems: Functional and Cost-Benefit Analysis for a Metropolitan Area. *Urban Rail Transit.* **5**(4), 289–309 (2019). <https://doi.org/10.1007/s40864-019-00118-y>
20. Gerber, J.L., Suppiger, T., Sauter, T.C., Traschitzger, M., Müller, M., Exadaktylos, A.K.: Tram, rail, bicycle: An unhappy triad? Rising incidence and resource consumption of tramline-associated bicycling accidents in Bern, Switzerland. *Accid. Anal. Prev.* 151 (2021)
21. Strandroth, J., Sternlund, S., Lie, A., Tingvall, C., Rizzi, M., et al.: Correlation between Euro NCAP pedestrian test results and injury severity in injury crashes with pedestrians and bicyclists in Sweden. *Stapp Car Crash J.* **58**, 213–231 (2014)
22. Guerrieri, M.: Tramways in Urban Areas: An Overview on Safety at Road Intersections. *Urban Rail Transit.* **4**(4), 223–233 (2018)
23. Rjabovs, A., Palacin, R.: Investigation into Effects of System Design on Metro Drivers' Safety-Related Performance: An Eye-Tracking Study. *Urban Rail Transit.* **5**(4), 267–277 (2019). <https://doi.org/10.1007/s40864-019-00115-1>
24. Li, K.: The challenges and potential of risk assessment for active safety of unmanned tram. *iCCAIS 2018 - 7th Int Conf Control Auto Info Sci* 22–27 (2018). <https://doi.org/10.1109/ICCAIS.2018.8570696>
25. Palmer, A.W., Sema, A., Martens, W., Rudolph, P., Waizenegger, W.: The Autonomous Siemens Tram. *IEEE 23rd Int Conf Int Trans Sys* (2020) <https://doi.org/10.1109/ITSC45102.2020.9294699>
26. Zeng, X., Xiong, Q., Wang, Y., He, J., Liang, Y., Bian, D.: Modern Tram Auxiliary Safety Protection Technology Based on Obstacles Detection. *J. Tongji Univ.* **47**(1), 64–70 (2019)
27. Wang, W.: Detection of panoramic vision pedestrian based on deep learning. *Image Vis. Comput.* **103**:103986 (2020). <https://doi.org/10.1016/j.imavis.2020.103986>
28. Redmon J., Farhadi A.: YOLO9000: better, faster, stronger. in Proc. IEEE Conf. CVPR. 6517–6525. (2017). <https://doi.org/10.1109/CVPR.2017.690>
29. Li, G., Li, S.E., Cheng, B., Green, P.: Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities. *Transp. Res. Pt. C-Emerg. Technol.* **74**, 113–125 (2017)
30. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. in Proc. IEEE Conf. CVPR., 580–587. (2014)
31. Girshick, R.: Fast r-cnn. in Proc. IEEE ICCV. 1440–1448. (2015)
32. Li, G., Yang, Y., Qu, X.: Deep learning approaches on pedestrian detection in hazy weather. *IEEE Trans. Ind. Electron.* **67**(10), 8880634 (2020) <https://doi.org/10.1109/TIE.2019.2945295>
33. Elgendy, M.: Deep Learning for Vision Systems. Manning Shelter Island (2020)
34. Zhao, S., You, F.: Vehicle detection based on improved yolov3 algorithm. *Proceedings - 2020 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS.* 76–79. (2020)
35. Ge, L., Dan, D., Li, H.: An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision. *Struct. Control Health Monit.* **27**(12) (2020). <https://doi.org/10.1002/stc.2636>
36. Guerrieri, M., Parla, G.: Deep learning and yolov3 systems for automatic traffic data measurement by moving car observer technique. *Infrastructures* **6**(9), 134 (2021). <https://doi.org/10.3390/infrastructures6090134>
37. Hui, J.: Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3. Available: <https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088>. Accessed 1 Aug 2021 (2018)
38. Zhang, Z.: A flexible new technique for camera calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence.* *IEEE: Piscataway.* **22**, 1330–1334 (2000)
39. Dewi, C., Chen, R.-C., Yu, H.: Weight analysis for various prohibitory sign detection and recognition using deep learning. *Multimed. Tools Applic.* **79**, 43–44 (2020)
40. Chen, Q., Liu, L., Han, R., et al.: Image identification method on high speed railway contact network based on YOLO v3 and SENet. In: *Chinese Control Conference, CCC.* 8772–8777. (2019)
41. Nieto, M., ArróspideLaborda, J., Salgado, L.: Road environment modeling using robust perspective analysis and recursive Bayesian segmentation. *Mach Vis Appl* **22**(6), 927–945 (2011). <https://doi.org/10.1007/s00138-010-0287-7>
42. Golzales, R.C., et al.: *Digital Image Processing Using MATLAB.* Prentice Hall, Upper Saddle River, New Jersey (2004)
43. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. ACM.* **24**(6), 381–395 (1981). <https://doi.org/10.1145/358669.358692.S2CID972888>
44. Kripa, S., Monisha, M.: Single-Viewpoint Panorama Construction With Wide-Baseline Images using SIFT And SURF Features. *IOSR J. Electron. Commun. Eng.* 63–71 (2017)
45. Dorj, B., Lee, D.J.: A precise lane detection algorithm based on top view image transformation and least-square approaches. *J. Sens.* 4058093 (2016) <https://doi.org/10.1155/2016/4058093>
46. Aly, M.: Real time detection of lane markers in urban streets. *IEEE Intelligent Vehicles Symposium, Proceedings.* 7–12. (2008). <https://doi.org/10.1109/IVS.2008.4621152>



47. Kalman, R.E.: A new approach to linear filtering and predictions problems. *Journal of Basic Engineering*. **82**(D), 35–45 (1960). <https://doi.org/10.1115/1.3662552>
48. Welch, G., Bishop, Gary: *An Introduction to the Kalman Filter*. Department of Computer Science University of North Carolina at Chapel Hill Chapel Hill, NC 27599–3175 (2006). Available: [https://www.cs.unc.edu/~welch/media/pdf/kalman\\_intro.pdf](https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf) Accessed 1 Aug 2021
49. Niu, M.: *Object Detection and Tracking for Autonomous Driving by MATLAB toolbox*. Ohio State University. Available: [https://etd.ohiolink.edu/apexprod/rws\\_etd/send\\_file/send?accession=osu1523978995395614&disposition=inline](https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1523978995395614&disposition=inline). Accessed 10 Aug 2021. (2018)
50. Guerrieri, M., Parla, G., Corriere, F.: A new methodology to estimate deformation of longitudinal safety barriers. *ARNP J. Eng. Appl. Sci.* **8**(9), 763–769 (2013)
51. Guerrieri, M., Corriere, F., Parla, G., Ticali, D.: Estimation of pollutant emissions from road traffic by image processing techniques: A case study in a suburban area. *ARNP J. Eng. Appl. Sci.* **8**(8), 668–676 (2013)
52. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* **39**(6), 1137–1149 (2017). <https://doi.org/10.1109/TPAMI.2016.2577031>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.