Multimed Tools Appl (2018) 77:14673–14703
DOI 10.1007/s11042-017-5054-6

Driver information system: a combination of augmented reality, deep learning and vehicular Ad-hoc networks

Lotfi Abdi[1] **· Aref Meddeb**[2]

Received: 20 December 2016 / Revised: 12 May 2017 / Accepted: 24 July 2017 /
Published online: 3 August 2017

Abstract Improving traffic safety is one of the important goals of Intelligent Transportation Systems (ITS). In vehicle-based safety systems, it is more desirable to prevent an accident than to reduce severity of injuries. Critical traffic problems such as accidents and traffic congestion require the development of new transportation systems. Research in perceptual and human factors assessment is needed for relevant and correct display of this information for maximal road traffic safety as well as optimal driver comfort. One of the solutions to prevent accidents is to provide information on the surrounding environment of the driver. Augmented Reality Head-Up Display (AR-HUD) can facilitate a new form of dialogue between the vehicle and the driver; and enhance ITS by superimposing surrounding traffic information on the users view and keep drivers view on roads. In this paper, we propose a fast deep-learning-based object detection approaches for identifying and recognizing road obstacles types, as well as interpreting and predicting complex traffic situations. A single convolutional neural network predicts region of interest and class probabilities directly from full images in one evaluation. We also investigated potential costs and benefits of using dynamic conformal AR cues in improving driving safety. A new AR-HUD approach to create real-time interactive traffic animations was introduced in terms of types of obstacle, rules for placement and visibility, and projection of these on an in-vehicle display. The novelty of our approach is that both global and local context information are integrated into a unified framework to distinguish the ambiguous detection outcomes, enhance ITS by superimposing surrounding traffic information on the users view and keep drivers view on roads.

✉ Lotfi Abdi
lotfiabdi@hotmail.com

Aref Meddeb
aref.meddeb@infcom.rnu.tn

1 Networked Objects Control and Communication Systems Laboratory, National Engineering School of Tunis, University of Tunis El Manar, Tunis, Tunisia

2 Networked Objects Control and Communication Systems Laboratory, National Engineering School of Sousse, University of Sousse, Sousse, Tunisia

Keywords Intelligent transportation systems · Augmented reality · Head-up display · Deep learning · Cooperative safety systems

1 Introduction

In recent years, the use of the automobile as the primary mode of transportation has increased significantly and driving has become an important activity of daily life. As such, road traffic generates serious problems in terms of congestion, safety, and environmental impact. Today, due to the rapidly-growing number of vehicles, road accidents constitute the eighth leading cause of death. With the technological advances in the areas of mobile computing, wireless communications, and remote sensing, Intelligent Transport Systems (ITS) have recently emerged as a promising technology that will enable the deployment of diverse safety, traffic efficiency, and infotainment applications.

Driving is a complex, continuous, and multitask process that involves driver's cognition, perception, and motor movements. The development of digital image sensors, communication technologies, and computer vision techniques offer a great deal of advantages and enable a variety of compelling ITS applications and components such as Driver Assistance Systems (DAS), Traffic Signs Recognition (TSR), Traffic and Activity Monitoring, Network Traffic Behavior Analysis, Traffic Management, etc. Vision-based driver-assistance systems for the overtaking maneuver require low data transmission delay to achieve reliable detection of vehicles coming from the opposite direction. The way road traffic signs and vehicle information is displayed impacts strongly driver's attention with increased mental workload leading to safety concerns. Today, the automotive industry is focused on Human Systems Interactions (HSI) related to safety and convenience while driving. Next-generation DAS will increase road safety by helping drivers become better aware of the road and its potential hazards. Research in perceptual and human factors assessment is needed for relevant and correct display of this information for maximal road traffic safety as well as optimal driver comfort.

In the context of automotive, HSI constitutes a big challenge taking into account road safety issues and complexity of the driving task under high time constraint. There is a strong correlation between a vehicles accident involvement and its drivers hazard perceptionc as most of severe road accidents occur because the driver becomes aware of a hazardous situation too late to react properly. In order to improve driving safety and minimize driving workload, the information provided should be represented such that it is easily understood and imposes less cognitive load onto the driver. In this context, future cooperative perception technologies provide the possibility of assisting drivers by so called advisory warnings in potentially dangerous driving situations. In vehicle-based safety systems, it is more desirable to prevent an accident (active safety) rather than reduce the severity of injuries (passive safety). However, active safety systems pose more difficult and challenging problems than passive ones. Critical traffic problems such as accidents and traffic congestion require the development of new transportation systems aimed at addressing critical issues like passenger safety and traffic congestion. This can be achieved by integrating information and communication technologies into transportation infrastructure and vehicles.

The rapid growth of technology has propelled novel ways in which drivers senses may be augmented. One such technology is Augmented Reality (AR) applied to a vehicular context, opening new opportunities for merging digital content with real world environment. To support this task, existing on-board systems display mainly visual messages, forcing the drivers to move their eyes away from the road. Today, vehicle manufacturers have been pointing to AR as a next-generation visualization technology for in-vehicle driving displays. AR uses embedded vision technology to enhance the driver's view of real-world situations with computer-generated graphics. Further, AR-based Head-Up Displays (AR-HUD) are emerging as a next-generation in-vehicle display technology, potentially reducing drivers' cognitive workload. AR-HUD based vehicular safety information system can enhance ITS by superimposing surrounding traffic information onto the users view and keep drivers attention on the road.

On the other hand, advanced computer vision algorithms are a driving force in the next-generation Advanced Driver Assistance Systems (ADAS). A key component is vision-based machine intelligence that can provide information to the control system or the driver to maneuver a vehicle properly based on the surrounding road conditions. Vision is the most important sense used for driving and therefore computer vision algorithms are the most critical for ADAS. Computer vision tries to acquire, process, analyze, and understand visual data captured by all kinds of sensors from the real world. Though many aspects of scene understanding in computer vision have significantly advanced in the past decades, recognition and parsing of complex objects in real scenes is still largely unsolved. It is a discipline at the crossroad between computer science and artificial intelligence, attracting a large number of researchers. Deep convolutional networks have become the most popular architecture for large-scale image recognition and segmantation tasks. ADAS and driverless vehicles will heavily rely on deep learning-based machine vision for identifying and recognizing pedestrians and vehicle types, as well as interpreting and predicting complex traffic situations.

The problem of detecting and localization has been dealt with in various areas of computer vision. Detection of multiple objects in the presence of occlusions has been a notorious problem in computer vision. Vision algorithms for driver assistance systems usually need to comply with strong real-time constraints. Recent advancements in artificial neural networks and so-called deep learning are accelerating the reality of self-driving vehicles faster than was originally expected as hardware and software vendors are taking the lead in pushing the technology to enable autonomous vehicles forward. Artificial Intelligence (AI), and in particular Deep Learning based on neural network computing can recognize objects even faster and with more accuracy than humans. The concept of deep learning, also known as machine learning, has been in place since the early 80's but only recently has the technology advanced to a point where it has become feasible. Its architecture are designed to support the fusion of many different camera into a coherent data set that can be analyzed to obtain positions placement of obstacles.

The idea of deep learning is to attempt to artificially emulate the functionality of the human brain via hardware and software. Deep learning, in which computers learn a desired behavior using artificial intelligence and neural network concepts, could be a viable solution for ADAS. It fuses them to accurately detect objects, identify them, determine where the car is relative to the world around it, calculate its optimal path for safe travel, and be able to build that 3D environment model of everything going on around the vehicle. Deep learning-based convolutional neural networks have recently emerged as the leading approach for achieving state-of-the-art object detection accuracy for a wide range of object classes. A significant amount of work has set the state-of-the art in object detection by using deep learning descriptors generated with Convolutional Neural Network (CNN). In object detection, methods such as R-CNN have reached excellent results by integrating CNNs with region proposal generation algorithms, such as selective search [34]. CNNs also demonstrated excellent performance on a number of visual recognition tasks that include classification of entire images [12], predicting presence/absence of objects in cluttered scenes or localizing objects by ROIs. The idea is to remove object proposal generation and regress a grid based object representation for an image.

Overtaking maneuver on roads without a clear view can cause serious accident. We aim to improve the accuracy of AR traffic information system in order to assist the driver in various driving situations, increase the driving comfort, and reduce traffic accidents. Hence, systems and methods for adaptive streaming with augmented video stream in highway scenariohave been proposed as an overtaking maneuver assistance system. Therefore, the main goals of our proposed adaptive framework are to perform video streaming over highway VANET in a reliable and efficient way without reducing quality of video and incurring a high load into the network.

The services and applications that provide by video streaming over VANETs have recently become very attractive. They can be utilized to guarantee road safety, create new forms of inter-vehicle communications, and avoid potential accidents. Cooperative tracking methods can provide DAS with more information about vehicles. The objectives of road safety applications are to decrease the number of accidents and to help increase safety. For instance, these applications warn driver when another driver makes immediate break. As part of this, the dynamic information displayed on the windscreen makes it possible to better understand how driving aids work and, in autonomous mode, to quickly send a signal to the driver in order to take back control of the vehicle. The targeted application of V2V video streaming delivery is overtaking maneuver assistance systems using a see-through video.

The performance of video streaming suffers from the delay and packet loss incurred by the long time disconnection. In order to achieve a high-quality and real-time video streaming on VANETs, an error recovery technique should be applied to the video stream. Although many solutions have been proposed to handle the high mobility problem, few of them addressed the problem in the context of video transmission. The design of a feasible solution for the successful and timely deliver of video frames over VANETs has to be in accordance with all features of video streaming.

In the future, ADAS and driver-less vehicles may more and more deeply rely on deep learning-based machine vision to recognize objects and predict potential dangers. Such systems may as well help interpret and predict complex traffic situations. Moreover, in a world that is becoming increasingly complex, communication and navigation systems exact a high cognitive price for the assistance they provide to the driver. Cooperative systems, which allow vehicles to communicate with each other to achieve a common goal, are widely recognized. Today, car manufacturers are looking to enhance safety and personalize and augment the driving experience by building AR technology into newer car models.

In order to prevent traffic accidents by distracted driving, the proposed system presents a cooperative overtaking assistance systems based on real-time video streaming, where a video stream captured with a camera installed at the windshield of a vehicle is augmented, compressed, broadcasted, and displayed in surrounding vehicles. The video is augmented with computer-generated highlighting of pedestrians, traffic signs, and vehicles. The vehicle in front sends a video stream captured by a camera placed on its windshield to the vehicle that wants to initiate the overtaking maneuver. The receiving vehicle then uses sensors that measure the distance through computer vision to determine the geometry of a translucent image that is projected over its windshield by means of a holographic projector.

In this context, this paper aims to deliver innovative intelligent driver information systems based on AR and deep learning for identifying and recognizing road obstacles as well as interpreting and predicting complex traffic situations. We aim to provide a prototype implementation of a visual AR system that can significantly improve driving experiences. By analyzing information from both looking in out of the vehicle, such systems can actively prevent vehicular accidents and improve driver safety and experience. The aim is to combine sensors and algorithms to understand the vehicular environment so that the driver can receive assistance or be warned of potential hazards. We propose an assistance system that explores this concept for delivering information about the topology of the network on the windshield in order to avoid road accidents in low-visibility situations. This cooperative system is able to increase the visibility of the driver to avoid overlooking obstacles, thus making critical maneuvers faster and safer.

The remainder of the paper is organized as follows. Section 2 presents related work while Section 3 introduces the system architecture of the proposed cooperative system. Section 4 provides experimental results performed with vehicles on urban roads and Section 5 concludes this paper.

2 Related work

Vehicular safety has been actively explored in recent years. Even before the appearance of motorized vehicles, many devices were developed and placed not only in vehicles, but also in the road environment as a means to regulate traffic, to provide better awareness, and to increase road safety for all road users. Among the factors that may contribute to traffic accidents, human error is one of the most important factors, such as driver's inattention and wrong decisions. To reduce driving accidents, it is important for drivers to obtain driving information easily and efficiently.

Automobile assistance systems can be split into in-Vehicle information systems and ADAS. The major concern of such systems is to minimize driver's distraction. Information and Communication Technologies (ICT) are key tools to provide access to knowledge and services, and also facilitate several kinds of systems interaction. ICT are essential for interaction between vehicles and also between vehicles and their environment [7]. To give new answers to the increasing mobility demand and other open issues, several solutions have been adopted. Nowadays, the research field of intelligent vehicles systems benefits from the emerging inter-vehicle communication to perform cooperation. There are yet many new technologies in development for vehicles that are going to enrich even more the options of services to be deployed over VANETs.

Most research on intelligent vehicle systems were devoted to single vehicle system. On the other hand, the rapidly developing vehicular communication technology [28] stimulates research interests on vehicle cooperation. For example, in cooperative platooning systems [5], cooperative Adaptive Cruise Control systems [39], and cooperative collision warning systems [24], each vehicle shares its own state with other relevant vehicles via inter-vehicle communication. The main advantage of cooperative perception can be summarized as an increase in situational awareness, without substantial additional costs. For example, the system in [2] assists the driver in avoiding collision through a Smart Dashboard system at a very low cost based on video-based analysis. Other proposals address overtaking in the intelligent vehicle suc as [63] and [42]; or different detection approaches based on technologies such as optical flow [3]. Most of those proposals focus on the blind-spot issue [40, 44] and lane-departure warning [53], alerting the driver of a potential hazard and thus, preventing accidents.

Recent studies have explored the relationship between the allocation of visual attention and driving behavior [64]; but the relationships between the allocation of visual attention, AR and driving performance are yet to be explored [32]. It is intuitive then that the next step is to provide on-board communication capabilities to merge sensed information (e.g. speed, vehicle malfunction, objects proximity), automated mechanisms (e.g. breaking, cruise control, parking) and infotainment equipment (e.g. on-board screens, video and audio systems, video streaming) available at several vehicles with many valuable services (e.g. collision avoidance, accident alerts, notification to first responders) [50, 66]. Such a system may assist the driver by monitoring the driver or vehicle behaviors to predict/detect driving situations and alert the driver to take corrective action. Collision avoidance systems warn drivers of potential collision threats that may be in the Line-Of-Sight (LOS) of the driver or out of the LOS of the driver (e.g., determined from wireless communications).

Via inter-vehicle communications, drivers can be informed of crucial traffic information such as treacherous road conditions and accident sites by communicating with each other and/or with the roadside infrastructure [29]. With better knowledge of traffic conditions, it is plausible that the problem of accidents can be alleviated [8]. In [9], the authors illustrate an automatic speed controllable vehicle exploiting V2I communication system. A vehicle collects information and transmits it to a Road Side Equipment (RSE). This latter receives information from each vehicle within a management section and transmits information to other vehicles to adjust their speed. The vehicle receiving information required to adjust the speed may also display a warning notification for the driver. Collision avoidance systems may generate visual and/or auditory alerts to warn the driver of the potential collision threats [59]. The fusion of these devices, information, and services should be maintained through the cooperation of all the entities involved. In [45], the authors introduced a see-through system based on VANETs for assisting overtaking maneuvers.

Video quality at the receivers is affected by distortion due to packet loss and delay. Streaming delay of video content in a real-time cannot exceed a few seconds. Due to the high density and high mobility of vehicles, it is difficult to design an efficient broadcast protocol for VANETs in urban areas where packet loss is also a major issue. In order to improve the delivery ratio, redundancy is used as an error correction mechanism to make inter-vehicle data transmissions more reliable. The use of redundancy is ideal for VANETs in handling packet loss as they do not require any interaction between source and receivers nodes.

Several recent studies focus on the forwarding error correction mechanism, which creates redundant information for important information to ensure the transmissions. However, because of the limited resources of wireless networks and the fragile connections in a VANET, forwarding error correction presents some drawbacks such as burst of messages caused by the redundant information and burst of message loss [65]. To reduce the contention and redundant messages, a multi-hop relay node selection scheme is proposed in [30]. The authors propose a rebroadcast mechanism for vehicle node selection over VANET.

Some research activities were also devoted to evaluate methods for directing driver attention with the use of AR cues [31], [62]. The primary goal was to determine the costs and benefits of dynamic conformal AR cues to alert experienced drivers to potential roadway hazards. The enclosed aspect of a car, allied with the configuration of the controls and displays directed towards the driver offer significant advantages for AR systems when considering the amount of immersion it can provide to the user [55]. Recent studies have investigated how AR displays impact elderly drivers' performance [31, 52] and [14]. They found that AR cues in general help elderly detect hazardous target object of low visibility. In [38], authors present the effectiveness of various tactile warning signals under relatively low perceptual load and high presentation rate.

In [56], the authors proposed a lane change aid system wherein the driver of a motor vehicle traveling along a highway is warned if unsafe lane change or merge maneuver is attempted, regardless of information available through the vehicle's rear-view mirror system. It requires that virtual objects are placed in the correct 3D position, orientation, and comply with the human eyes scale factor. In [35], the authors presented a method of cooperative perception for AR application. The idea is to transform the occluded part of the first vehicle to a perception of the rear vehicle based on the 3D perspective in order to have a common reference for vehicle perception.

Object recognition in computer vision comes in many flavors, two of the most popular being object detection and semantic segmentation. Previous work has proposed explicit modeling of object part appearances and locations for more accurate recognition and localization. Local methods like sliding windows need to calculate a complex prediction function for each pixel of the input image, which can be expensive to compute and makes it difficult to take the context into account [41]. State-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations. Prominent methods like SPPnet [22], Fast R-CNN [16] and [48] have reduced the running time in networks used in deep-learning-based object detection system to about 5-17 fps. Deep CNNs have demonstrated excellent performance in image classification [25], but here is still room for improvement in object-detection tasks with many categories, in particular for cluttered scenes and occlusion.

Previously proposed methods are not equipped to accurately recognize the location of a detected obstacle. In this context, our research is focused on combining computer vision techniques to develop real-time algorithms able to assist driving activities. Computer vision systems may help check other sensors, ensuring that the total reliability of vehicle safety applications is maintained. We want to augment todays ADAS with deep learning systems that will learn the behavior of drivers over time [13]. By adding the computer vision and AR features to the human machine interface, an automaker takes the user experience to an entirely new level. Vital alerts and road information can be displayed on a windscreen or a head-mounted wearable device. This is one of the key enabling technologies not only for user experience, but for the future autonomous driving.

3 Cooperative driving systems

The augmented perception and situation awareness capability can contribute to better driving in terms of decision making and planning. In order to propose a visualization metaphor adapted to the driving situation, it is necessary to know what is the current situation and what the driver looks at. The proposed cooperative perception technology needs the recognition technology which generates driving-related information based on information collected from various sensors, the decision technology which determines what information to guide for various driving situation based on the recognition information and the technology which represents the decision information as the AR [33].

Our goal was to identify ways in which the overall driving experience can be improved in terms of safety, navigation, and traffic control. Smart vehicles will be able to give route directions, warn drivers of impending collisions, keep drivers alert, increase the driving comfort and reduce traffic accidents. We proposed a cooperative information system by combining the new paradigm of dedicated wireless communications, AR-HUD technologies and deep learning–based machine vision for identifying and recognizing road obstacles types, as well as interpreting and predicting complex traffic situations. We provide a prototype implementation of a visual AR system that can significantly improve the driving experience as shown in Fig. 1.

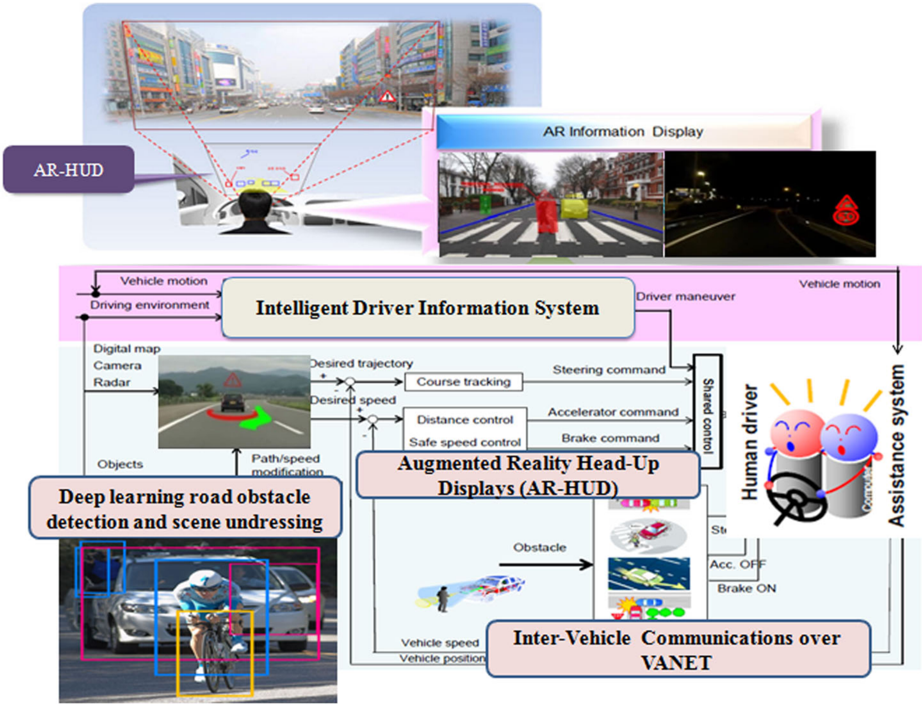

Fig. 1 Cooperative Driving Information System

3.1 Deep convolutional neural network for road obstacle detection

The vision algorithms for driver assistance systems usually need to fulfill strong real-time constraints. Hence, we draw a particular focus on real-time capability of the algorithms evaluated here. Object detection systems based on the deep convolutional neural network (CNN) have recently made groundbreaking advances on several object detection benchmarks. The recognition of road obctacles correctly at the right time for that particular place is very important for any vehicle driver to ensure a safe journey for themselves and their passengers. However, the way road traffic, signs or vehicle information is displayed impacts strongly driver's attention with increased mental workload and safety concerns. In the current developed system, each module recognizes forward vehicle, pedestrian and traffic signs respectively and also determines what information to guide for the driving situations and displays respectively. This paper investigates how to extract objects-of-interest without relying on handcraft features and sliding windows approaches, as shown in Fig. 2.

3.1.1 Deep CNNs for region proposal

Object detection is the process to identify and localize objects of a specific category in an image. The accuracy and short processing time are extremely important for TSR. Our Deep CNN is a fully-convolutional network that simultaneously predicts object localization in images and object-ness scores at each position. Our network uses global image features to predict detections, which drastically reduces its errors from background detections. It also

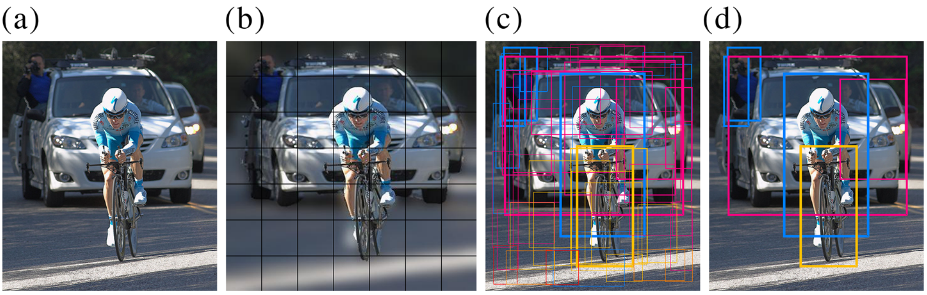

Fig. 2 Object Detection using Deep Neural Networks: **a** Test image, **b** Divides the image into a regular grid, **c** Object categories present in the image, **d** Object class detection

predicts all ROIs for an image simultaneously, it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network. For each image, algorithms produce ROIs indicating the position and scale of all road obstacle along with a confidence score corresponding to each ROIs. We associate a set of default ROI with each feature map cell, for multiple feature maps at the top of the network. The default ROI tile the feature map in a convolutional manner, so that the position of each ROI relative to its corresponding cell is fixed. At each feature map cell, we predict the offsets relative to the default ROI shapes in the cell, as well as the per-class scores that indicate the presence of a class instance in each of those ROI [36].

Algorithm 1 Deep CNNs for Region Proposal

```
Input: Image I
Output: Final list of detected object bounding boxes
Initialization: Grid[][]=Divide-Regions(I) ;
for i = 1; i < W; i++ do
    for j = 1; j < H; j++ do
        if Grid[i][j] != 0 then
            C = Pr(Obj) * IOU;
            P = Pr(Class_k|Obj);
        else
            C = 0;
            P = 0;
        end
    end
end
return Vect = [X, Y, W, H, P];
```

We aim to detect all objects in an image by dividing into grid cells (7 $\times$ 7 pixels) and regressing the probability of ROI falling into cells and also regressing the ROIs of each object. Specifically, we introduce binary variables for each cell in the ROIs indicating whether or not the pixels in the cell are from the object centered in the ROIs. The first step, referred to as the *grid cell analysis*, divides the image into a grid of regular cells each of which predicting a distribution over class labels, as well as a ROIs for the object whose center falls into it. The regions of objects are detected as transformed objects, which are different from the previously registered background. At each feature grid cell, we predict 5 coordinates as (x, y, w, h) and confidence P. Here x, y are spatial coordinates of the ROI's center coordinates, and w, h its width and height as shown in Fig. 3.

- Confidence Scores as: $C \leftarrow Pr(Obj) * IOU$
- Confidence Prediction is obtained as: $P \leftarrow Pr(Class_i|Obj)$
- Class-specific confidence score as: $Pr(Class_i|Obj) * Pr(Obj) * IOU = Pr(Class_i) * IOU$

In order to ensure a correct detection process, we need to estimate a set of ROI for each image. The aim is to output a single image-level score for each object classes.

3.1.2 Network architecture

A deep CNN that simultaneously predicts ROIs and object-ness scores at each position. We train our network on full images and directly optimize detection performance. The full input image is fed into a deep neural network consisting of several convolutional layers, recurrent layers, and fully connected layers. The network first passes the input image into several convolutional and max-pooling layers to extract feature maps. Afterwards, for each object proposal a pooling layer extracts a fixed-length feature vector from each ROIs using a CNN. In other words, the max-pooling layer operation is applied for every location of the object in the image at the position with the maximum object-ness score, as illustrated in Fig. 4.

The ability to generate coherent sets is particularly important in our case because our framework needs to remember previously generated predictions and avoid multiple predictions of the same object. Pixel similarity between two objects is computed by comparing the presence of pixels relative to each object's ROIs. At the same time, we minimize the confidence of the remaining predictions, which are deemed not to localize true objects.

The goal of object detection is to locate and identify instances of an object category within an image. First, we define object detection as a regression problem to the coordinates of several ROIs. In addition, for each predicted box, the net outputs a confidence score of how likely this ROIs contains an object. The main difference with the localization task is to predict a background class (non-object class) when no object is present. To address this

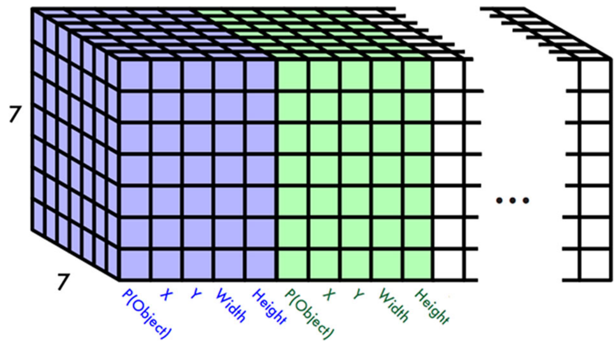

Fig. 3 Example of predicts object localization in images and object-ness scores at each position

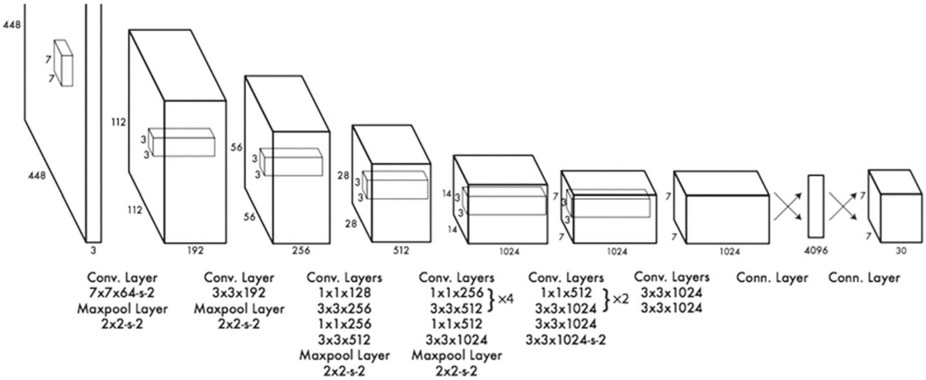

Fig. 4 Architecture of the deep convolutional neural network for object detection

problem , each class is treated as a separate binary classification problem. The loss function is therefore a sum of K binary logistic regression losses, one for each of the K classes $k \in \{1, , K\}$.

$$f(g_k(x, y_k) = \frac{1}{Z(x)}(\exp[E(y, x)]) \tag{1}$$

Where $g_k(x)$ is the output of the network for the image test x and $y_k \in \{1, 1\}$ is the image label indicating the presence of class k in the image test x. Each class score $g_k(x)$ can be interpreted as a posterior probability indicating the presence of class k in image x with transformation. The ROI pooling layer uses max-pooling to convert the features inside any valid ROI into a small feature map with a fixed spatial extent of $H \times W$ (e.g., 7×7), where H and W are layer hyper-parameters that are independent of any particular ROI [16]. As such, they can be classified with a subsequent classifier to achieve object detection. We demonstrate that a single fully convolutional neural network, if designed and optimized carefully, can detect objects under different scales with heavy occlusion extremely accurately and efficiently.

3.2 Inter-vehicle video communications

In order to prevent traffic accidents by distracted driving, the proposed system presents a cooperative overtaking assistance systems based on real-time video streaming, where a video stream captured with a camera installed at the windshield of a vehicle is augmented, compressed, broadcasted, and displayed in surrounding vehicles. The video is augmented with computer-generated highlighting of pedestrians, traffic signs, and vehicles. The vehicle in front sends a video stream captured by a camera placed on its windshield to the vehicle that wants to initiate the overtaking maneuver. The receiving vehicle then uses sensors that measure the distance through computer vision to determine the geometry of a translucent image that is projected over its windshield by means of a holographic projector.

The services and applications that provide by video streaming over VANETs have recently become very attractive. They can be utilized to guarantee road safety, create new forms of inter-vehicle communications, and avoid potential accidents. Cooperative tracking methods can provide DAS with more information about vehicles. The objectives of road safety applications are to decrease the number of accidents and to help increase safety. For instance, these applications warn driver when another driver makes immediate break. As part of this, the dynamic information displayed on the windscreen makes it possible to better understand how driving aids work and, in autonomous mode, to quickly send a signal to the driver in order to take back control of the vehicle. The targeted application of V2V video streaming delivery is overtaking maneuver assistance systems using a see-through video.

To overcome problem of loss and delay of emergency packets, the node selection must be achieved in a limited time. The selected vehicle nodes should be as central as possible in order to broadcast the content to a maximum number of neighbors, without the need of further retransmissions [6]. The idea is that an actor is central if it can quickly interact with all others, and it is based on the geodesic distances among nodes. The proposed mechanism selects a minimum sub-set of rebroadcaster vehicles in order to achieve high video quality and reduce interferences. The vehicles are ranked based on their capacity to reach other vehicles and their strategic location in the network using a new centrality metric inspired from the social network analysis, referred to as dissemination capacity [50]. Figure 5 illustrates an example of rebroadcast mechanism for vehicle node selection over VANET.

In a traditional broadcasting mechanism system, each node in the network receiving the content will rebroadcast it. Intensive rebroadcasts increase packet loss and the video quality is degraded. Our proposed broadcasting mechanism relies on an algorithm where each node selects a subset of nodes to forward message. It combines the advantage of neighbor coverage knowledge and delayed rebroadcast mechanism so as to decrease the number of retransmissions. This idea prompted us to enhance our video streaming system over VANET by new re-broadcaster selection mechanism, which selects a minimum subset of neighboring vehicles to rebroadcast the content. In the broadcasting process, the current sender selects the best node that will rebroadcast the message to other vehicles as fast as possible using an efficient forward node selection mechanism.

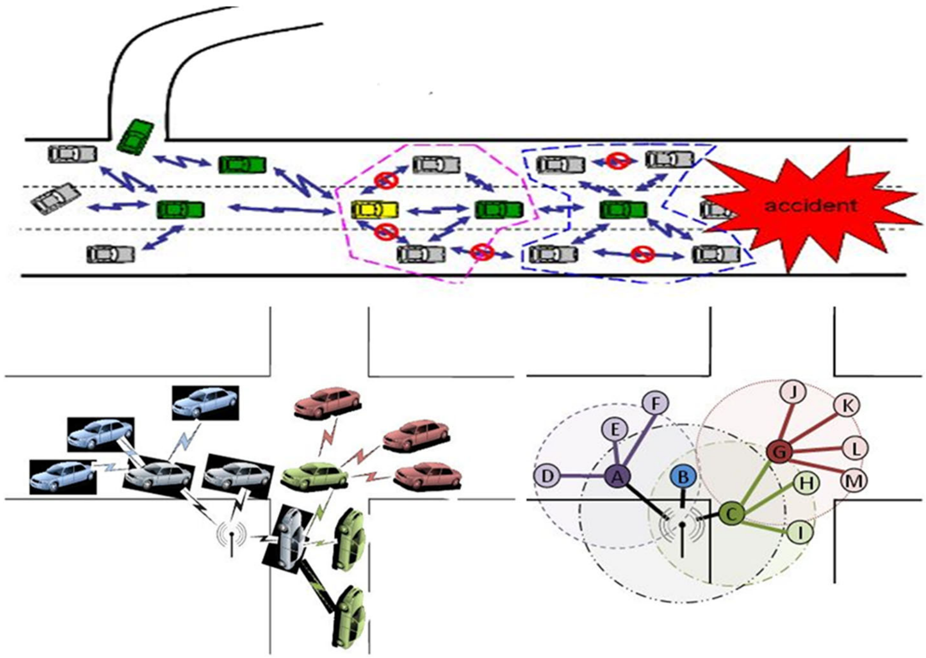

Fig. 5 Rebroadcast mechanism for vehicle node selection over VANET

The combination of visual AR and sensor-based systems with connected vehicles can be more effective in reducing road accidents and provide better traffic flow. Cooperative driving systems are a possible solution to improve traffic efficiency and road safety. A vision-bstructing vehicle such as a bus or a truck is equipped with a dashboard camera and a V2V communication device. As a vehicle equipped with a virtual windshield approaches the vehicle in front, it establishes a real-time video stream from the windshield camera to the virtual windshield. Driver using AR applications can read raffic information in a non-distracting wayt and keep their eye on roads without distracting attentions. Based on the vision-obstructing vehicle's dimensions and relative position to the vehicle, the video stream is overlaid onto the vehicle using the AR capabilities of the virtual windshield and computer vision to seamlessly overlap the preceding vehicle.

In the next section, we aim to provide a prototype implementation of a visual AR system that can significantly improve the driving experience. We employ this approach to improve the accuracy of AR traffic information system to assist the driver in various driving situations, increase the driving comfort, and reduce traffic accidents. AR applications can enhance ITS by superimposing surrounding traffic information on the users view and keep drivers and pedestrians view on roads.

3.3 Projection-based augmented reality head-up display

To make the information correctly match the real environment according to the variations in the drivers' view, the pose estimation of target ROIs is an essential approach in the AR. Although the problem of camera calibration has been extensively studied, our scope of application imposes some restrictions to the generic problem. More precisely, we are interested in a low cost resolution for on-board vehicular cameras. In order to render a virtual object into the real world, a virtual camera has to be placed in the same position and orientation as the real camera. To estimate the camera pose, we perform a camera calibration off-line and store the camera's intrinsic parameters. The intrinsic parameters are those specific to the camera, such as the focal length, principal point, and lens distortion. The extrinsic parameters refer to the 3D position and orientation of the camera.

In order to estimate the camera's extrinsic parameters for a given frame, some correspondences between 2D points from the image and 3D points from the model are needed. The camera pose can be found from projecting the 3D coordinates of the features into the 2D image coordinate, along with minimizing the difference between their corresponding 2D features. The localization of the object in the real-time image uses the homography matrix to find the corresponding comers of the object in the reference and real-time images. Basically, the system takes four known points from the image of a scene and sets four separate tracking windows around the points. The first three input parameters, namely *Rotation*, *Scale* and *Distortion*, are extracted directly from the homography. The relationship between camera model and tracking target is shown in Fig. 6.

The key to realize a AR 3D-registration is to obtain a camera projection matrix, which represents the relationship between the 2D points from the image and the 3D points from the model. The geometric relationship between 3D world lines and their projections on the camera image are built to estimate the relative 6-DOF camera pose consists of rotation parameters and translation parameters [21]. From the planar homography, we can easily compute the camera position and rotation, which provides the motion estimates. The used mathematical model is the projection transformation, which is expressed by (2) where λ is the homogeneous scale factors unknown a priori, where P is a 3×4 projection matrix, $x = (x, y)$ are the homogeneous coordinates of the image features, $X = (X, Y, Z)$ are

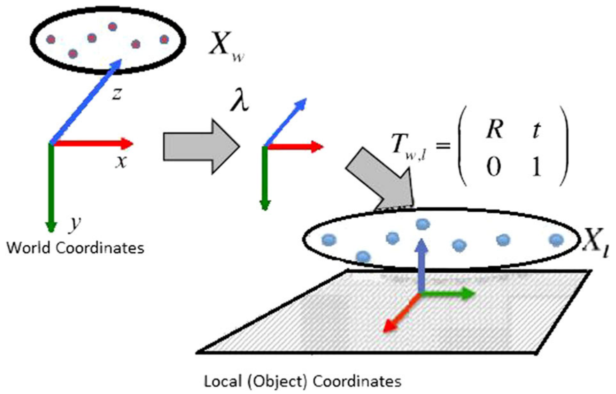

Fig. 6 Coordinate systems involved in camera calibration

the homogeneous coordinates of the feature points in the world coordinates, $K \in R^{3\times3}$ is the matrix with the camera intrinsic parameters, also known as camera matrix, the joint rotation-translation matrix $[R|t]$ is the matrix of extrinsic parameters, $R = [r_x r_y r_z]$ is the 3×3 rotation matrix, and $T = [t]$ is the translation of the camera.

$$x = \lambda P X = K[R|t]X \tag{2}$$

The projection matrix P is the key to creating a realistic augmented scene using the intrinsic parameters of the camera, the dimensions of the video frame, and the distances of the near and far clipping planes from the projection center. In our method, we assume that the intrinsic parameters are known in advance and do not change, and this is reasonable in most cases.

$$\begin{aligned} P &= \overbrace{K}^{\text{Intrinsic matrix}} \quad * \quad \overbrace{[R|t]}^{\text{Extrinsic matrix}} \\ &= \overbrace{\underbrace{\begin{pmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D translation}} * \underbrace{\begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D scaling}} * \underbrace{\begin{pmatrix} 1 & s/f & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D shear}}}^{\text{Intrinsic matrix}} * \overbrace{\underbrace{(I|t)}_{\text{3D translation}} * \underbrace{\left(\begin{array}{c|c} R & 0 \\ 0 & 1 \end{array}\right)}_{\text{3D rotation}}}^{\text{Extrinsic matrix}} \end{aligned} \tag{3}$$

Once K is known, the extrinsic parameters for each image is readily computed. From equation 2, we have:

$$\begin{array}{l} r1 = \lambda + K^{-1}h_1 \\ r2 = \lambda + K^{-1}h_2 \\ r3 = r1 * r2 \\ t \;\; = \lambda + K^{-1}h_3 \end{array} \; where \left\{ \begin{array}{l} h_1 = [h_{11}\; h_{21}\; h_{31}]^T \\ h_2 = [h_{12}\; h_{22}\; h_{32}]^T \\ h_3 = [h_{13}\; h_{23}\; h_{33}]^T \\ r_1 = [r_{11}\; r_{21}\; r_{31}]^T \\ r_2 = [r_{12}\; r_{22}\; r_{32}]^T \\ r_3 = [r_{13}\; r_{23}\; r_{33}]^T \end{array} \right. \begin{array}{l} Where\ H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \\ and\ \lambda = \frac{1}{\|K^{-1} + h_1\|} \end{array} \Bigg\} \tag{4}$$

The Projection-based AR corresponds to the use of projection technology to augment and enhance 3D objects and spaces in the real world by projecting images onto their visible surfaces. Once there are enough successful matches, a RANSAC method is applied to calculate the homography matrix between the image of the frame and the image of the object. Then we are able to estimate the 3D pose and draw a virtual 3D object on the top of the real object. The camera calibration allows combining virtual and real world objects in a single display. In the case of images or videos, the relative position of an element on a screen can be calculated from the camera parameters and relative position information of the camera with respect to the element.

To correctly model the perspective projection of the camera, we must mimic the intrinsic camera parameters in the virtual environment. When we have the camera calibrated in a frame, we can synchronize the real camera with a virtual camera and project the virtual objects onto the real image using OpenGL. Technically, this can be described with a projection matrix that maps 3-D points onto a 2-D plane. After the world has been aligned with the camera using the view transformation, the conversion from an intrinsic matrix to the model view and projection matrices requires a conversion from the world coordinates to the normalized view volume coordinates used by OpenGL. The perspective projection matrix is expressed by (5), where width, height, far, near, represents the positions of the clipping planes.

$$\begin{bmatrix} x_{clip} \\ y_{clip} \\ z_{clip} \\ w_{clip} \end{bmatrix} = \begin{bmatrix} \frac{2*c_x}{width} & 0 & 1-\frac{2*x_0}{width} & 0 \\ 0 & \frac{2*c_y}{height} & -1+\frac{2*y_0}{height} & 0 \\ 0 & 0 & \frac{near+far}{near-far} & -2*\frac{near*far}{near-far} \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_{camera} \\ Y_{camera} \\ Z_{camera} \\ 1 \end{bmatrix} \tag{5}$$

As it is indicated, most of current marker-less tracking approaches require a 3D model of the environment for matching 2D features to those lying on the model. In addition to the complexity of building a model, such a strategy will result in performance problems when the model is very complex or the environment is dynamic. In contrast, our approach does not need to perform 3D engineering of the environment. Also, we use a simple virtual 3D model, with a known size, to define a reference coordinate system. For the robust tracking of the camera pose, we have developed a new marker-less approach that combines information from both real and virtual worlds.

The AR-HUD also reflects navigation information in the real exterior view by overlaying the field of vision with supplementary digital information, to improve driver safety, enhance the driver's experience and to protect the drivers from various road hazards. The system draws data from the vehicle's camera and combines it with the vehicle's dynamics data, which is used to create a model of the external view of the car as seen from the driver's perspective [61]. To that end, the use of vehicular communications and AR to enable safer and innovative intelligent driver information systems, each ITS entity periodically broadcasts safety messages to notify its neighborhood about its context and location information.

AR systems have the potential to communicate with other vehicles to improve tracking performance and make better prediction or detection of critical events. The AR information is here to increase visual saliency and attract driver's attention to the proximity of the next and coming vehicle. When driving in reduced visibility conditions, we tend to leave less inter-distance with the next vehicle and the preceding vehicle serves as a landmark for trajectory control [19]. We integrate AR-HUD based Forward Crash Warning, Lane Change Warning, Do Not Pass Warning, Left Turn Assist, Do Not Pass Warning, Intersection Movement Assist and vehicles tracking for driver assistance, this novel approach adds valuable safety functionality and provides a contextually relevant of the on-road environment for driver. Here, the driver sees which vehicle in front is detected by their vehicle system. The information relating to each Scenarios is reflected precisely into the road layout, supplementing reality with 3D information to guide the driver through the maneuver.

4 Experimental results

Nowadays, vehicular communications are considered as the main research area for developing technologies on road safety, accident alarming, anti-collision warning. Because of the contention nature of the IEEE 802.11p protocol, packet collisions and losses are common phenomena and significantly impact the service quality. Video streaming in VANETSs imposes stringent requirements in terms of delay, scalability, delivery ratio, reliability, and mobility in order to provide a satisfying level of service at the user's end. In this paper, we propose a novel framework for the transmission of video streams over VANET where we apply various adaptive techniques to transmit video streams while keeping video quality in unreliable and highly dynamic vehicular networks.

We conducted a number of simulation experiments with different scenarios using the Evalvid Tool [11], and real maps from open street map [18], which were imported into Simulation of Urban MObility (SUMO) [4], allowing us to generate the desired vehicle flows, and NS-3 [1] combined with traffic simulator SUMO and evalvid to simulate real-time traffic.

The proposed system have been installed to a test vehicle with a vehicle AR information system prototype and carried out in the real road environments. Data fusion from ADAS technologies can help to achieve this objective by combining information from multiple sensors, thus improving fault tolerance at the same time. Our implementation is based on the open source Caffe deep learning library [27] using OpenGL and OpenCV Library. To evaluate the performance of the proposed algorithm, we propose a single-stage training algorithm that jointly learns to classify object proposals and refine their spatial locations. Our implementation is based on the open source Caffe deep learning library [27] using OpenGL and OpenCV Library.

4.1 Object detection

In this section we will demonstrate a set of experiments to examine our controllers performance under different conditions created by varying simulation parameters. By combining object detection sensing with AR-HUD, we show that it is possible to design novel cooperative ADAS. In the current developed system, each module recognizes forward vehicle, pedestrian and traffic signs and also determines what information to guide for the driving situations and displays respectively.

4.1.1 Road obstacle detection

Autonomous driving requires detailed map and road knowledge. To be capable of fully autonomous driving, a vehicle must understand the environment, know the roads, accurately predict road changes, obstacles and pedestrians. Object detection systems based on the deep CNN have recently made ground-breaking advances on several object detection benchmarks. Similarly to the training pipeline of R-CNN [17], we fine tuned the deep CNN models pre-trained on ImageNet database using images from both train and validation sets of VOC 2007, Voc 2012. We primarily evaluate detection mean Average Precision (mAP), as this is the common metric for object detection.

We applied our proposed methods to standard visual object detection tasks on PASCAL VOC 2007. It consists mainly of complex scene images in which ROIs of 20 diverse object classes were labelled. The detailed comparison of the proposed framework with current leading approaches for object detection is presented in Table 1.

Table 1 Detection performance of our modules on VOC 2007 test set

	Bicycle	Bus	Car	Cow	Dog	Horse	Motorbike	Person	mAP
R-CNN [17]	72.8	66.3	74.2	63.5	61.2	69.1	68.6	58.7	58.5
fast-RCNN [16]	78.1	81.6	78.6	78.8	84.7	82.0	76.6	69.9	70.0
SPPnet [22]	72.3	74.4	73.0	73.6	70.3	74.6	74.3	54.2	63.1
RPN [48]	79.0	83.1	84.7	81.9	84.8	84.65	77.5	76.7	73.2
NoC [49]	77.7	78.0	75.5	77.2	81.1	75.9	75.1	61.6	68.8
Our Deep CNN	**80.4**	83.6	84.1	67.8	73.2	78.4	85.0	74.6	74.48

For a given test image, the ROI with detection scores are first predicted by object detectors. We present detection results for the PASCAL VOC 2007 dataset and compare our mAP to other top detection methods. The most performing detectors at the moment are: fast-RCNN [16] has 70 % mAP, the SPPnet [22] has 63.1 % mAP, the RPN [48] has mAP of 73.2 % and the NoC method [49] has 68.8 %. The first two methods represent two different lines of approaches for object detection. Our deep CNN detection system achieves mAP of 74.48 % on VOC 2007 detection challenges, thus surpassing the previous state-of-art by a significant margin. Furthermore, our approach improves by 1.28% the best competing approach, and obtains the best results for 8 out of 20 categories.

While our system is indeed competitive, there exist methods which have substantially larger computational cost, but that can achieve better detection performance, notably on VOC2007 localization. The main difference between our model and other state-of-the-art models is real-time capability of the algorithms. We demonstrated significantly improved performance over the state-of-the-art at different levels, our detection system is extremely fast at test time since it only requires a single network evaluation has a frame rate of 35 fps (including all steps) on a GPU, unlike classifier-based methods. In order to provide quantitative evaluation of the localization power of our detection system, Fig. 7 shows representative examples of successful detection using our method.

Some qualitative results are shown in Fig. 7 containing images with single object as well as images with multiple interacting objects with rigid transportation tools, articulated animals, and indoor objects. A Single CNN predicts region of interest and class probabilities directly from full images in one evaluation. Based on these results, it is fair to say that the proposed method can satisfactorily handle background clutters, objects with low contrast with the background, and multiple objects, as far as the detection is accurate enough. It increases localization accuracy and robustness to false positives over traditional non-maximum suppression.

4.1.2 Traffic sign detection

In order to evaluate the effectiveness of our method and compare it with state-of-the-art methods, our implementation is based on the open source Caffe deep learning library [27]. We conduct comprehensive experiments to demonstrate the performance of the proposed method and also present comparison with previous methods. We implement the suggested method in C++ and test the real-time performance on the German Traffic Sign Recognition Benchmark (GTSRB) dataset [57]. For training and testing GTSRB dataset contains 51839

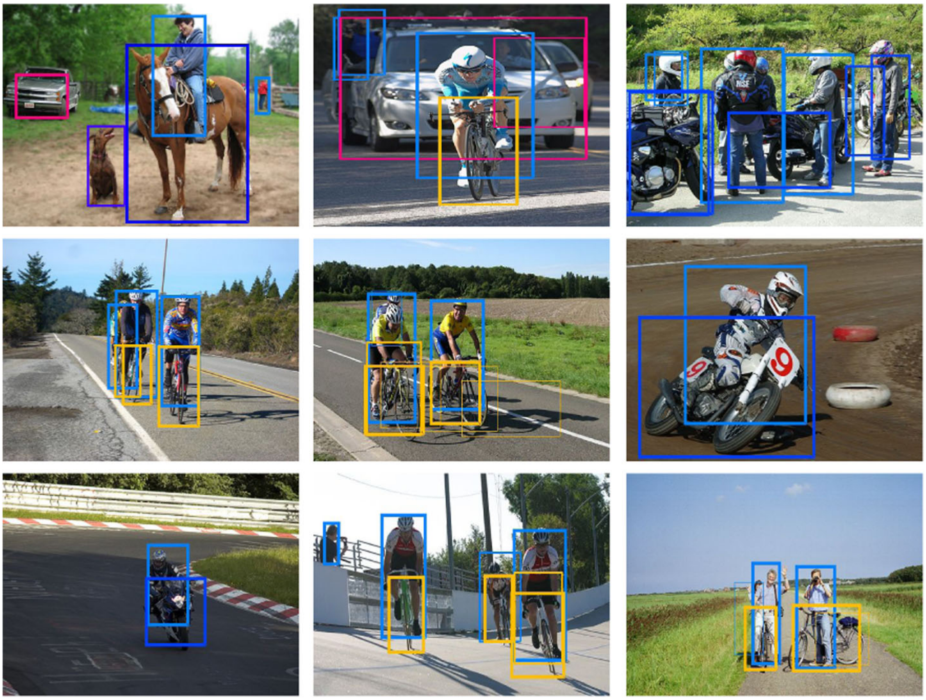

Fig. 7 Object detection in adverse conditions

images in 43 classes and 6,000 non traffic signs. We have selected 39,209 images for training and the rest for testing. The sizes of traffic sign examples are in range from 15x15 to 250x250 pixels.

We report these results in Table 2, where the results of the winning system from the IJCNN challenge are provided as references. We have compared the suggested method with other state-of-the-art algorithms such as the method [10, 54, 58, 67], and [20] . The performance is analyzed in terms of detection and recognition accuracy, and is presented in Table 2.

According to the results for the GTSRB data set, shown in Table 2, this work achieves a 99.31% recognition accuracy, which is a comparable performance of 0.24% less then the work by [10], and a performance of 0.17% higher than the work by [58] and 1.51% than the

Table 2 Performance comparison with other TSR methods

Method	[10]	[58]	[54]	[67]	[20]	Our
Speed limit	99.47	97.63%	98.61	95.95%	98.82%	99.13 %
Danger signs	99.07%	98.67%	98.03	92.08%	96.85%	98.97 %
Unique signs	99.22%	100.00%	98.63%	98.73 %	100.00%	99.51 %
Mandatory signs	99.89%	99.72%	97.18%	99.27 %	96.86%	99.45 %
Derestriction signs	99.72%	98.89%	94.44%	87.50 %	97.93 %	99.32 %
Other prohibitory signs	99.93%	99.93%	99.87%	99.13 %	98.27 %	99.47 %

Fig. 8 Detection of traffic signs in adverse conditions

work by [54]. The accuracy of recognizing unique signs reach 99.51%, which is comparable with the best achieved one. The danger signs which have triangular shape have given the worst results compared with other traffic sign categories.

In order to evaluate the system robustness, we have tested the accuracy of our algorithm when tracking the ROIs in the captured frames in various lighting and weather conditions, as shown in Fig. 8.

We evaluate the mAP and timing results for methods on PASCAL VOC 2007 test dataset, as shown in Table 3.

The main difference between our model and other state-of-the-art models is real-time capability of the algorithms. We demonstrated significantly improved performance over the state-of-the-art at different levels, our detection system is extremely fast at test time since it only requires a single network evaluation has a frame rate of 35 fps (including all steps) on a GPU, unlike classifier-based methods. A Single CNN predicts region of interest and class probabilities directly from full images in one evaluation. Based on these results, it is fair to say that the proposed method can satisfactorily handle background clutters, objects with

Table 3 Real-time object detection on pascal voc 2007

method	mAP (%)	rate (fps)
Fast R-CNN [48]	73.2	7
RPN (ResNet) [23]	76.4	2
SSD500 [37]	75.1	23
Yolo [47]	63.4	45
Our Deep CNN	74.48	35

low contrast with the background, and multiple objects, as far as the detection is accurate enough. It increases localization accuracy and robustness to false positives over traditional non-maximum suppression.

4.2 Inter-vehicle video communications

To simulate VANETs, we need two types of simulators: one for networking and the another for mobility. In our framework, we used NS-3 to simulate inter-vehicle communication in order to test our adaptive video streaming solution for the highway scenario. For mobility, we have used the SUMO package to generate a corresponding map scenario that includes roads, intersections, vehicles. EvalVid is responsible for generating the video stream for the transmission and evaluation of the quality of the transmitted video stream.

Different scenarios were assessed in urban environments where we use OpenStreetMap to generate a realistic map of 2000 x 2000 m was used. The number of vehicles varies from 50 to 150 within 600s of simulation time. This environment was simulated with up to 150 vehicles at speeds ranging from 20 and 150 km/h using 802.11b to communicate with each other with a communication range of 300 meters. The standard video sequences are encoded into 300 frames at a rate of 30 frame/s and with an intra-period of 5 GOP (Group Of Picture) IBPBPB. These frames are divided into a payload of 1000 bytes that could fit 353 different packets streamed using RTP encapsulated in UDP for delivering real-time video streams.

Performances are analyzed under different conditions of distance of data transmission and vehicles' arrival rates. The performance of routing protocols depends on the various parameters such as speed, pause time, node density, and traffic scenarios. Our evaluation is based on Peak Signal-to-Noise Ratio (PSNR), Packet Delivery Ratio (PDR), Average End-to-End Delay, and imperceptibility. The vehicle density is also one key factor that may affect the performance of a routing protocol. Performance evaluations for different vehicle densities are presented in [60].

4.2.1 Peak signal-to-noise ratio

Peak Signal-to-Noise Ratio (PSNR) is an important criterion to measure the error between the reconstructed and the original frame. The PSNR has become the most widespread objective metric used to assess the application-level QoS of video transmissions [26]. In each test, by changing the speed of vehicles, the quality of the transmitted video under different routing protocols is measured.

The decoded video quality is measured in terms of PSNR of the luminance component. As the PSNR is calculated per frame, we report the mean of the PSNR of a video. The average PSNR of the decoded video without network transmission is about 39 dB, which is the upper bound of the simulation result. Decoded video quality at the receiver is therefore affected by two factors: encoder compression performance and distortion due to the packet loss or late arrivals. We recall that a PSNR value between 30 and 40 dB means that the video quality is good enough. Figure 9 shows the simulation results for the PSNR of the reconstructed video at the receivers' vehicles for different vehicles densities scenarios as the total number of vehicles increases from 50 to 150.

Figure 9 shows that the average PSNR for different node densities are all greater than 30 dB. This indicates that the quality of the received video is good. Thus, our scheme improves the performance in terms of PSNR as the number of vehicles increases on the road. We can see that the average PSNR value of the delivered video packets increases as the number

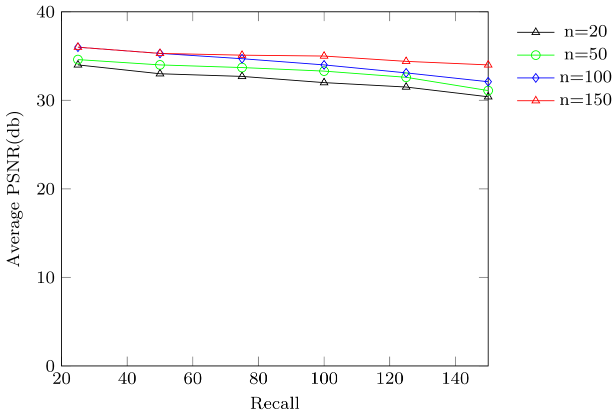

Fig. 9 Average PSNR (*db*) affected by vehicle density and different number of node (*vehicles*)

of nodes (vehicles) on the road increases. The reason for that is that when the total number of vehicles increases, the probability of connectivity between vehicles increases, which consequently results in reducing the distortion caused by packet loss.

4.2.2 Packet delivery ratio

Packet Delivery Ratio (PDR) is defined as the ratio of the number of data packets successfully received by the destination node to the total number of data packets generated by the source node as per (6). In Fig. 10, we give the PDR vs. the data-sending rate and under different speeds for various number of nodes.

$$PDR = \frac{\sum \text{Number of Packet Receive}}{\sum \text{Number of Packet Send}} \tag{6}$$

In Fig. 10, it can be observed that the PDR slowly increases as the numbers of nodes (vehicle) increases. Specifically, we can see that as the number of nodes increases from 20 to 100, the PDR increases from 95% to 97%. The PDR then sharply increases from 97%

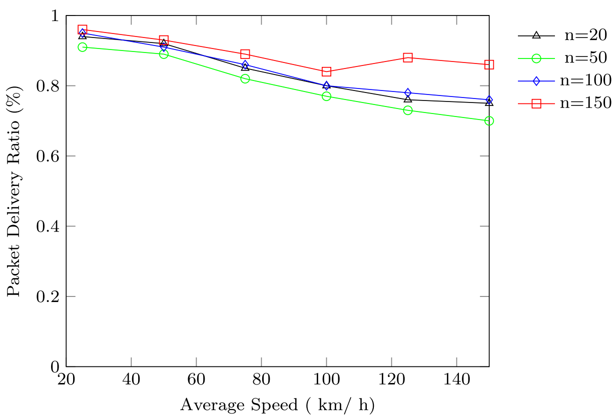

Fig. 10 Packet delivery ratio (%) vs. Number of Vehicles(Node) & Average Speed (km/h)

to 99% when number of nodes increases from 100 to 150. In fact, it is quite obvious that packet delivery is more likely to be successful as the network becomes highly connected. In a dense network, more forwarding nodes have the chance to be placed in the routing path, and consequently, the PDR of the routing protocols reaches high values [15] as well.

On the other hand, we can see from Fig. 10 that the average PDR decreases noticeably when the average velocity exceeds 100 km/h. As it shown, higher network density contributes to a better PDR. This result is also quite predictable since when the vehicles move at high speeds, the probability of transmission errors increases as well. This result stems from the fact that as the network topology becomes more dynamic, links/routes become more prone to signal loss and link disconnections.

4.2.3 End-to-end delay (E2E)

E2E Delay represents the average time for data packets to travel from the source node to the destination node. It also includes the delay caused by route discovery process, the queueing delay, contention time, and signal propagation. Only the data packets that are successfully delivered to destination are accounted for. To measure the E2E delay, we compute the difference of packet sending and receiving times. The average delay is obtained by adding the stored delay of each received packet and then dividing by the total number of received packets as shown in (7) [46].

$$\text{E2E} = \frac{\sum_i Arrival\_Time_i - Sending\_Time_i}{TotalNumberofPackets} \tag{7}$$

In Fig. 11, we give the average E2E delay (ms) vs. the average speed (km/h) and the number of vehicles (nodes). It can be observed that the average end-to-end delay decreases by increasing the number of nodes, because the forwarding node can find more qualified nodes in its neighborhood. While the density increases above 100 vehicles/km2, the network is more and more connected, and as a consequence, the forwarding mechanism is less used, thus the decrease of the average E2E delay. More precisely, since our packet transmission scheme is achieved with the help of leader nodes, the density of the vehicles does not affect the delay, except the case when the network is partitioned.

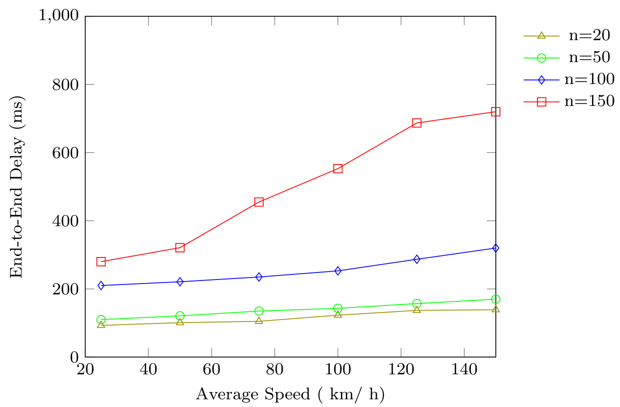

Fig. 11 End-to-End Delay (ms) vs. Number of Vehicles(Node) & Average Speed (km/h)

4.3 Wayfinding and navigation aids

To provide driving-safety information using the proposed AR-HUD, drivers receive all important information before their eyes in an easily comprehensible way. Our system can successfully detect objects that are in front of the vehicle and they reflect information such as your speed, driving warnings or navigational directions on the inside of the windscreen to make it easily viewable without taking your eyes off the road. With the augmentations embedded in the real exterior view, drivers intuitively recognize the significance of what they see in front of them.

The usability of navigation systems also improves with a better visual feedback regarding the road ahead, especially with our AR-HUD, where navigation information is seamlessly integrated in to the environment in front of your vehicle. Sensor-driven world-fixed graphics can also be used to cue drivers' attention to relevant hazards quickly and accurately, especially for near-invisible objects or low-visibility by super imposing virtual representations of pedestrians as shown in Fig. 12.

This study suggests that pedestrian collision warning improves driver performance as compared to the baseline, regardless of cue presentation method. We investigated the effect of visual warning presentation methods on drivers performance and braking behavior in pedestrian collision avoidance. AR can evolve to the next level by highlighting the exact locations of Vehicle, people and sign in the direct field of view of the driver, during both night and daytime. For example in Fig. 12 shows, if any person is detected it will show an alert and some visual representation to the driver to avoid the accident that is explained in the following figure, the AR generates an alert symbol and pedestrians position in the windshield. The sensor is used to continuously monitor and analyze road conditions, identifying situations that may be dangerous to the driver. If the system senses that the vehicle will collide with one of those objects, it will create an audio, visual, and haptic warning, designed to give the driver enough time to react and avoid or mitigate a collision.

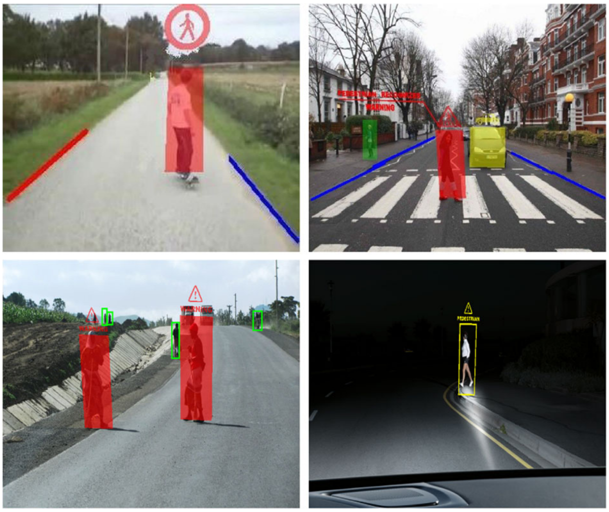

Fig. 12 Pre-Collision System with Pedestrian Detection and Collision Warning

Another important issue is the study of conventional traffic signs, in terms of rules for placement and visibility, types of traffic signs and the migration of these to in-vehicle display. By using such technology, vehicles are able to reduce the lock to-lock steering wheel travel as a function of the vehicle speed. We start the evaluation of the AR tracking by superimposing 3D graphics on target images. By deploying an on-board camera-based driver alert system against approaching traffic signs such as stop, speed limit, unique, danger signs, etc. Here we propose new signs and develop 3D models of these signs as well as other road infrastructures. If the vehicle is approaching an intersection with a stop sign, which is positioned after a bend, this can be shown to the driver earlier. To provide driving safety information using the proposed AR-TSR, various sensors and devices were attached to the experimental test vehicle, as shown in Fig. 13.

AR not only enhances the driving experience but can drastically change the way we behave while driving. AR-HUD of the future will enable a dialogue without words between the vehicle and driver. Safety of the driver is greatly improved with the navigation and directions projected onto the screen. This will increase safety on the roads and build trust in existing vehicle systems and new driving features such as automated driving. This is a revolution in itself, but also just the beginning of a new kind of interaction with drivers [51].

Experimental results show that the model can provide warning for the driver which can effectively improve vehicle safety. The AR-HUD supports the driver if the vehicle is in danger of unintentionally drifting out of a lane. Our lane departure warning system for example gives immediate feedback if you are going to leave the lane. The warning information is determined by analysing the state information of a front vehicle and the driving information of the vehicle gotten from each module of the system. This generation of virtual symbol supplements the exterior view of the traffic conditions in front of the vehicle with augmentations information for the driver.

4.4 Driver performance directing attention with AR cueing

Visual perception plays a large role in determining a drivers situation awareness of the environment. Relevant road traffic as well as useful navigation or path planning information

Fig. 13 Driving safety information using the proposed AR-TSR

may be displayed using partially or totally the windshield surface thanks to these emerging technologies. With improved feedback regarding the driving situation and the status of your car, it will become harder to miss important information. for example in Fig. 14 shows, a virtual symbol inserted precisely into the exterior view shows the driver the way on the curve in front of the vehicle. When distance controls (Forward Crash Warning) are enabled, a marking in the AR-HUD visualizes which vehicle in front is detected by the assistance system assures you that you are driving the correct distance from the car in front of you and gives you an early warning if you are not.

Explicit cues can help improve visual search performance but they can also cause perceptual issues such as attentional tunneling. In this section we evaluate the effects of AR cues designed to direct the attention of experienced drivers to roadside hazards. We evaluated driver performance using two measures: reaction time and response rate percentage for targets by cueing. AR can evolve to the next level by highlighting the exact locations of vehicle, people and sign in the direct field of view of the driver, during both night and daytime, we examine issues of driver distraction and how AR can influence it. Figure 15 shows the effect of AR cueing on response rates for target objects.

In this study, AR cueing increased response rate for detecting vehicles, pedestrians, and warning signs. This cue design provided information to the driver without obstructing the target. The results showed that no main effect of AR cueing was observed for objects of high visibility(vehicles). Vehicles were generally visible from a greater distance than pedestrian and warning sign targets because of their larger size and color contrast against the rural driving scene. A main effect of cueing was observed for pedestrian and traffic sign target objects. Participants responded to 24% more pedestrians and 6.1% more warning signs in cued conditions than in uncued conditions (in Fig. 15).

In addition, we hypothesized that alerts would help to direct attention to potential hazards. In summary, the most important results observed relating to the potential benefits of AR cueing included effects of cueing for detection of Vehicles, pedestrians and warning signs and an effect of cueing for response time (TCR) for warning signs. Figure 16 shows the effect of AR cueing on time to collision at response (TCR) for target objects.

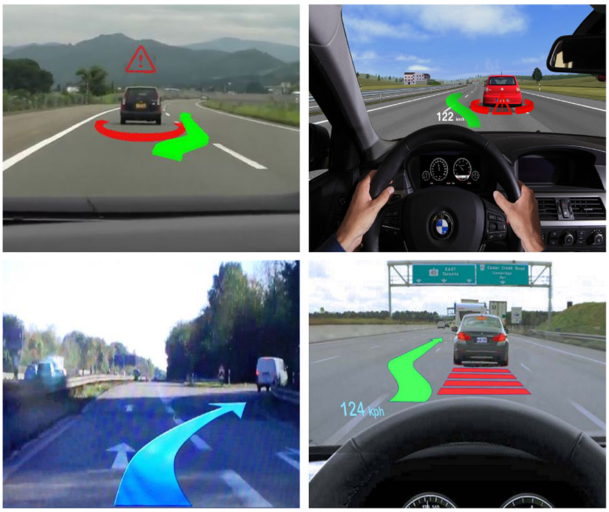

Fig. 14 Improving Road Safety with Information Visualization

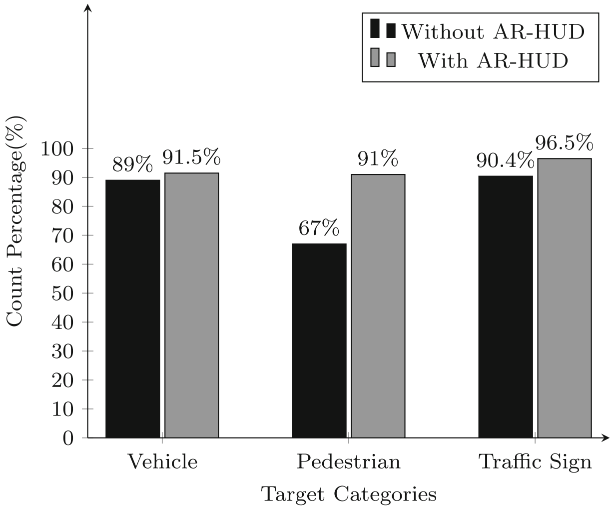

Fig. 15 Response rate (count) percentage for targets by cueing

The experiment result showed that early warnings helped distracted drivers react more quickly–and there by avoid more collisions–than no warnings. AR cues improved participant response time (TCR) to warning signs. Early warnings have been observed to help drivers react more quickly, particularly compared to when no warning is given in Fig. 16. Participants responded to these targets 0.5 seconds faster in cued conditions than in uncued conditions. Compared with the no-warning condition, an early RECAS warning reduced the number of collisions by 80.7% and the corresponding severity by 87.5%. The experiment result showed that warnings provide a potential safety benefit by reducing the time

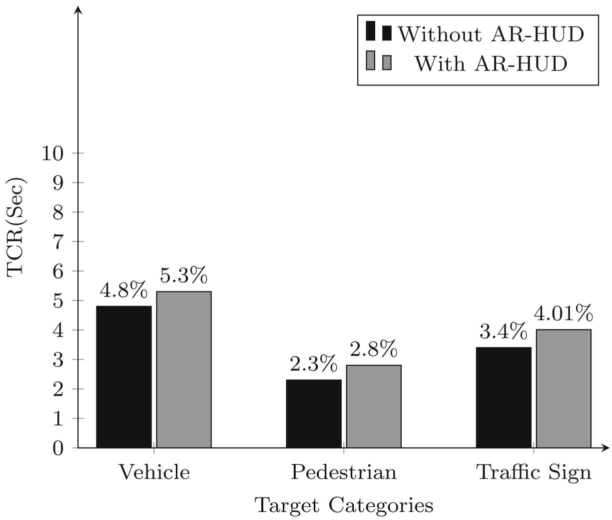

Fig. 16 Time to collision at response (TCR) for all target categories

required for drivers to release the accelerator. A user-centered process for creating and evaluating designs for AR displays in automobiles helps to explore what collaborative role AR should serve between the technologies of the automobile and the driver [43]. AR cueing is a promising technology to mitigate risk by directing driver attention to roadway hazards without interfering with other driving tasks such as maintaining safe headway.

– AR cueing aided the detection of targets of low visibility (e.g., pedestrians, traffic signs)
– Response times improved for targets of low visibility (e.g., traffic signs) with cueing
– AR cueing did not impair driver ability to maintain consistent distance behind a lead vehicle
– AR cueing did not impair discrimination of secondary objects

On the other hand, AR combined with other modalities such as wireless communications or sound, can serve as the intermediary for driver-car collaboration. AR can also enhance the situation awareness of other road vehicles and pedestrians surrounding the driver. In this study, no evidence was observed suggesting that AR cues interfered with driver perception of secondary objects, even for participants with cognitive impairment. To the extent that response likelihood, accuracy, and to hazards contribute to crash likelihood, then improvement on these measures represents a benefit and decreased performance represents a cost [52]. This is important as the goal of the AR cueing in this application was to aid the detection of critical objects such as vehicles, pedestrians, and warning signs without adversely affecting perception of other potential hazards.

5 Conclusion

Due to the growing number of vehicles on the roads worldwide, cooperative ITS are currently considered as a key emerging technology to improve road safety, traffic efficiency, and driving experience. The goal of this work is to design a next-generation driver information system that improves safety and sustainability of vehicular transportation. The development of cooperative vehicular driving systems undeniably requires a combination of dedicated wireless communications, deep learning and AR technologies as the building blocks of cooperative driving systems. In this paper, we have proposed an in-vehicle cooperative driver information system to improve driving safety by combining the new paradigm of dedicated wireless communications, AR technologies and deep learning based machine vision for identifying and recognizing road obstacles types, as well as interpreting and predicting complex traffic situations. Recent advancements in the field of artificial intelligence and in particular deep learning have been able to recognize objects even faster and with more accuracy than humans. We have proposed a novel method for localizing objects in an image, a single neural network predicts the region of interest and class probabilities directly from full image in one evaluation. We have demonstrated that a single fully convolutional neural network, if designed and optimized carefully, can detect objects under different scales with heavy occlusion extremely accurately and efficiently.

In addition, the implementation of communication and information technologies to improve transport systems will profoundly impact the transport sector ability to actively and positively contribute to achieving sustainable mobility. We have presented an adaptive video streaming solution for a highway VANET scenario, using an efficient broadcasting mechanism. The proposed mechanism optimizes and secures the video transmission against packet loss over VANETs by new re-broadcaster selection mechanism, which selects a minimum subset of neighbors vehicles to rebroadcast the content. The use of wireless technology based on the VANETs for information exchange can influence the drivers' behavior towards improving driving performance and reducing road accidents. Another important issue is the costs and benefits of dynamic conformal AR cues to alert drivers to potential roadway hazards. A new AR-HUD approach to create real-time interactive traffic animations is introduced, in terms of rules for placement and visibility, types of objects, and migration of these to an in-vehicle display. The AR-HUD supplements the exterior view of the traffic conditions in front of the vehicle with virtual information for the driver.

References

1. Accesed 22 may 2014. https://www.nsnam.org/
2. Akhlaq M (2010) A smart-dashboard: Augmenting safe & smooth driving
3. Batavia PH, Pomerleau D, Thorpe CE et al (1997) Overtaking vehicle detection using implicit optical flow. In: IEEE conference on intelligent transportation system, 1997. ITSC'97. IEEE, pp 729–734
4. Behrisch M, Bieker L, Erdmann J, Krajzewicz D (2011) Sumo–simulation of urban mobility. In: The 3rd international conference on advances in system simulation (SIMUL 2011). Barcelona, Spain
5. Bergenhem C, Coelingh E, Johansson R, Tehrani A (2014) V2v communication quality: measurements in a cooperative automotive platooning application. SAE Int J Passenger Cars Electron Electr Syst 7(2014-01-0302):462–470
6. Bradai A, Ahmed T, Benslimane A (2014) Vicov: efficient video streaming for cognitive radio vanet. Veh Commun 1(3):105–122
7. Caird J (2004) In-vehicle intelligent transportation systems. Transportation in an Aging Society, p 236
8. Cheng HT, Shan H, Zhuang W (2011) Infotainment and road safety service support in vehicular networking: from a communication perspective. Mech Syst Signal Process 25(6):2020–2038
9. Choi RA (2015) Automatic speed controllable vehicle and method for controlling speed thereof. US Patent 20,150,149,059
10. Cireşan D, Meier U, Masci J, Schmidhuber J (2012) Multi-column deep neural network for traffic sign classification. Neural Netw 32:333–338
11. De Felice M, Cerqueira E, Melo A, Gerla M, Cuomo F, Baiocchi A (2015) A distributed beaconless routing protocol for real-time video dissemination in multimedia vanets. Comput Commun 58:40–52
12. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2013) A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531
13. Fang Z, Fei F, Fang Y, Lee C, Xiong N, Shu L, Chen S (2016) Abnormal event detection in crowded scenes based on deep learning. Multi Tools Appl 75(22):14,617–14,639
14. Fu WT, Gasper J, Kim S (2013) Effects of an in-car augmented reality system on improving safety of younger and older drivers. In: 2013 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 59–66
15. Ghoreyshi SM, Shahrabi A, Boutaleb T (2016) A novel cooperative opportunistic routing scheme for underwater sensor networks. Sensors 16(3):297
16. Girshick R (2015) Fast r-cnn. arXiv preprint arXiv:1504.08083
17. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 580–587
18. Haklay M, Weber P (2008) Openstreetmap: User-generated street maps. IEEE Pers Commun 7(4):12–18
19. Halit L, Kemeny A, Mohellebi H, Garbaya S, Mérienne F, Michelin S, Albaut V Augmented reality head-up-display for advanced driver assistance system: A driving simulation study
20. Haloi M (2015) A novel plsa based traffic signs classification system. arXiv preprint arXiv:1503.06643
21. Han P, Zhao G (2016) Line-based initialization method for mobile augmented reality in aircraft assembly. The Visual Computer, pp 1–12
22. He K, Zhang X, Ren S, Sun J (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Computer Vision–ECCV 2014. Springer, pp 346–361
23. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
24. Huang CM, Lin SY (2014) Cooperative vehicle collision warning system using the vector-based approach with dedicated short range communication data transmission. IET Intell Transp Syst 8(2):124–134

25. Huang J, Li B, Zhu J, Chen J (2017) Age classification with deep learning face representation. Multimedia Tools and Applications, pp 1–17
26. Jama M, Risaq A, Issa S, Khalifa OO (2010) Performance evaluation of mpeg-4 video transmission over ieee802. 11e. Int J Comput Netw Secur 2(5):11–15
27. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia, pp 675–678. ACM
28. Khaleda Y, Tsukadaa M, Santab J, Choia J, Ernsta T (2009) A usage oriented analysis of vehicular networks: from technologies to applications. J Commun 4(5):357–368
29. Khandelwal SA, Abhale AB, Nagaraj U (2014) Accident prevention and air pollution control using vanet under cloud environment
30. Kim H (2010) An efficient alert broadcasting scheme considering various densities in vanet. In: Networked digital technologies
31. Kim S, Dey AK (2009) Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 133–142
32. Kim S, Dey AK (2016) Augmenting human senses to improve the user experience in cars: applying augmented reality and haptics approaches to reduce cognitive distances. Multi Tools Appl 75(16):9587–9607
33. Lee JW, Yoon CR, Kang J, Park BJ, Kim KH (2015) Development of lane-level guidance service in vehicle augmented reality system. In: 2015 17th international conference on advanced communication technology (ICACT). IEEE, pp 263–266
34. Lenc K, Vedaldi A (2015) R-cnn minus r. arXiv preprint arXiv:1506.06981
35. Li H, Nashashibi F (2011) Multi-vehicle cooperative perception and augmented reality for driver assistance: a possibility to 'see'through front vehicle. In: 2011 14th international IEEE conference on intelligent transportation systems (ITSC). IEEE, pp 242–247
36. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S (2015) Ssd: single shot multibox detector. arXiv preprint arXiv:1512.02325
37. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: single shot multibox detector. In: European conference on computer vision. Springer, pp 21–37
38. Meng F, Spence C (2015) Tactile warning signals for in-vehicle systems. Accid Anal Prev 75:333–346
39. Milanés V, Shladover SE, Spring J, Nowakowski C, Kawazoe H, Nakamura M (2014) Cooperative adaptive cruise control in real traffic situations. IEEE Trans Intell Transp Syst 15(1):296–305
40. Mota S, Ros E, Ortigosa EM, Pelayo FJ (2004) Bio-inspired motion detection for a blind spot overtaking monitor. Int J Robot Autom 19(4):190–196
41. Moysset B, Adam P, Wolf C, Louradour J (2015) Space displacement localization neural networks to locate origin points of handwritten text lines in historical documents. In: ICDAR 2015 workshop on historical document imaging and processing
42. Naranjo JE, Gonzalez C, Garcia R, De Pedro T (2008) Lane-change fuzzy control in autonomous vehicles for the overtaking maneuver. IEEE Trans Intell Transp Syst 9(3):438–450
43. Ng-Thow-Hing V, Bark K, Beckwith L, Tran C, Bhandari R, Sridhar S (2013) User-centered perspectives for automotive augmented reality. In: 2013 IEEE international symposium on mixed and augmented reality-arts, media, and humanities (ISMAR-AMH). IEEE, pp 13–22
44. Ngai DC, Yung NH (2007) Automated vehicle overtaking based on a multiple-goal reinforcement learning framework. In: Intelligent transportation systems conference, 2007. ITSC 2007. IEEE. IEEE, pp 818–823
45. Olaverri-Monreal C, Gomes P, Fernandes R, Vieira F, Ferreira M (2010) The see-through system: A vanet-enabled assistant for overtaking maneuvers. In: Intelligent vehicles symposium (IV), 2010 IEEE. IEEE, pp 123–128
46. Paul B, Ibrahim M, Bikas M, Naser A (2012) Experimental analysis of aodv & dsr over tcp & cbr connections with varying speed and node density in vanet. arXiv preprint arXiv:1204.1206
47. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
48. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497
49. Ren S, He K, Girshick R, Zhang X, Sun J (2015) Object detection networks on convolutional feature maps. arXiv preprint arXiv:1504.06066
50. Rezende C (2014) Video streaming in vehicular ad hoc networks: challenges, protocols and the use of redundancy. Ph.D. thesis, University of Ottawa

51. Richter P Ar-hud – technology for a dialog without words. http://www.continental-corporation.com/www/pressportal.comen/themepressreleases/3automotivegroup/interior/pressreleasespr20140721arhudtechniken.html
52. Schall MC, Rusch ML, Lee JD, Dawson JD, Thomas G, Aksan N, Rizzo M (2012) Augmented reality cues and elderly driver hazard perception. Human Factors: The Journal of the Human Factors and Ergonomics Society p pp. 0018720812462029
53. Schofield K (2013) Driver assistance system for a vehicle. US Patent App. 13/919,483
54. Sermanet P, LeCun Y (2011) Traffic sign recognition with multi-scale convolutional networks. In: The 2011 international joint conference on neural networks (IJCNN). IEEE, pp 2809–2813
55. Silvéria M. K. (2014) Virtual windshields: merging reality and digital content to improve the driving experience. arXiv preprint arXiv:1405.0910
56. Sivaraman S, Trivedi MM (2014) Dynamic probabilistic drivability maps for lane change and merge driver assistance. IEEE Trans Intell Transp Syst 15(5):2063–2073
57. Stallkamp J, Schlipsing M, Salmen J, Igel C (2011) The german traffic sign recognition benchmark: a multi-class classification competition. In: The 2011 international joint conference on neural networks (IJCNN). IEEE, pp 1453–1460
58. Stallkamp J, Schlipsing M, Salmen J, Igel C (2012) Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural Netw 32:323–332
59. Stebbins MA, Kiefer RJ (2014) Vehicle seat back haptic alert systems and methods. US Patent App. 14/453,391
60. Tian D, Zhou J, Wang Y, Zhang G, Xia H (2016) An adaptive vehicular epidemic routing method based on attractor selection model. Ad Hoc Netw 36:465–481
61. Timofte R Kul belgium traffic signs and classification benchmark datasets. http://btsd.ethz.ch/shareddata
62. Tonnis M, Sandor C, Lange C, Bubb H (2005) Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In: Proceedings of the 4th IEEE/ACM international symposium on mixed and augmented reality. IEEE Computer Society, pp 56–59
63. Wang F, Yang M, Yang R (2009) Conflict-probability-estimation-based overtaking for intelligent vehicles. IEEE Trans Intell Transp Syst 10(2):366–370
64. Werneke J, Vollrath M (2012) What does the driver look at? The influence of intersection characteristics on attention allocation and driving behavior. Accid Anal Prev 45:610–619
65. Xie H, Boukerche A, Loureiro A Mervs: A novel multi-channel error recovery video streaming scheme for vehicle ad-hoc networks
66. Yang J, Xu R, Cui J, Ding Z (2016) Robust visual tracking using adaptive local appearance model for smart transportation. Mult Tools Appl 75(24):17,487–17,500
67. Zaklouta F, Stanciulescu B, Hamdoun O (2011) Traffic sign classification using kd trees and random forests. In: The 2011 international joint conference on neural networks (IJCNN). IEEE, pp 2151–2155

Lotfi Abdi received the graduate degree in computer science from the Higher Institute of Applied Sciences and Technology of Sousse, University of Sousse, Tunisia, in 2009, the engineer degree in Computer Science and the master degree in Intelligent and Communicating Systems from the National Engineering School of Sousse, University of Sousse, Tunisia, in 2012 and 2013, respectively, and the Ph.D. degree in Communication Systems from the National Engineering School of Tunis, University of Tunis El Manar, Tunisia, in 2016. His research interests include Intelligent Transport Systems, Computer Vision, Signal, Image and Video Processing, Augmented Reality, Artificial Intelligence and Neural Networks.

Aref Meddeb obtained his Engineer's degree from the National School of Engineering of Tunis, Tunisia, in 1992, and both his M.S. and Ph.D. degrees, in Electrical Engineering and Computer Science, from Ecole Polytechnique of Montreal, Canada, in 1995 and 1998, respectively. He worked with Alcatel (Tunisia/France), INRS (Canada), Teleglobe (Canada), and Nortel (France). From 2002-2010, he was assistant professor and deputy director at the Higher Institute of Computer Science and Communication Technology, University of Sousse where he also headed the Telecommunications Department from 2003-2005. Since 2012, he joined the faculty staff of the National School of Engineering, University of Sousse where he is currently full professor and deputy director. His research interests include Internet of Things, Sensor Networks, and Intelligent Transport Systems. He mainly focuses on Security, Quality of Service, and Optimization.