

A Deep Learning Solution for Integrated Traffic Control Through Automatic License Plate Recognition

Riccardo Balia[1], Silvio Barra[2], Salvatore Carta[1], Gianni Fenu[1], Alessandro Sebastian Podda[1(✉)], and Nicola Sansoni[1]

[1] University of Cagliari, DMI, 09124 Cagliari, Italy
{r.balia,salvatore,fenu,sebastianpodda,n.sansoni}@unica.it
[2] University of Naples "Federico II", DIETI, 80125 Napoli, Italy
silvio.barra@unina.it

Abstract. Nowadays, Smart Cities applications are becoming steadily popular, thanks to their main objective of improving people daily habits. The services provided by the aforementioned applications may be either addressed to the entire digital population or narrowed towards a specific kind of audience, like drivers and pedestrians. In this sense, the proposed paper describes a Deep Learning solution designed to manage traffic control tasks in Smart Cities. It involves a network of *smart lampposts*, in charge of directly monitoring the traffic by means of a bullet camera, and equipped with an advanced System-on-Module where the data are efficiently processed. In particular, our solution provides both: i) a risk estimation module, and ii) a license plate recognition module. The first module analyses the scene by means of a Faster R-CNN, trained over an ad-hoc set of synthetically videos, to estimate the risk of potential traffic anomalies. Concurrently, the license plate recognition module, by leveraging on YOLO and Tesseract, is active for retrieving the plate number of the vehicles involved. Preliminary experimental findings, from a prototype of the solution applied in a real-world scenario, are provided.

Keywords: Deep Learning · Smart Cities · Anomalies detection · License plate recognition

1 Introduction

Smart Cities services and applications are becoming ubiquitous in the modern life, positively influencing the daily habits their citizens. Currently, most of these applications fall in the areas of physical and logical security [12], targeted both at pedestrian [15] and drivers [3], including privacy aspects [12].

These brand new services are commonly provided by means of *Internet of Things* (IoT) sensors, spread in specific districts and roads or, in an increasing number of cases, all over the city. Such sensors are instrumental in sensing the environment, to infer new information and to provide people-oriented solutions

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12951, pp. 211–226, 2021.
https://doi.org/10.1007/978-3-030-86970-0_16

[38]. Nevertheless, their applications nowadays pave the entire range of the road safety and facility topics, not comprehensively:

- Smart Parking [10,11];
- Intelligent Transportation Systems [33];
- Smart Traffic lights [4];
- Automated Vehicles and Pedestrian Safety [17].

In this context, several scientific and industrial approaches have been proposed, relying on the mere vehicle/object tracking [9,24], as well as aimed at addressing related issues like camera calibration [30], scene segmentation [16], just to mention a few. More recently, with the advent of *learning approaches* [37], these approaches have been boosted towards a more generalist scenario, involving heterogeneous data sources and fusions [23], scene understanding proposals [6,8], video captioning techniques [7,18].

In view of the foregoing, this paper aims to propose a novel solution, based on Deep Learning and Computer Vision methods, to automatically monitor traffic flows for recognizing possible abnormal traffic situations and identifying the license plates number of the vehicles involved. The main benefit provided is to drastically optimise the recording and storing of such events for subsequent audits, as well as lay the foundations for the integration with automatic sanctioning or traffic analysis systems. Indeed, current solutions in literature or in industry are either not focused on scalability, or they require expensive hardware and software. In this sense, the proposed system is cost-effective and leverages on open-source techniques only.

Specifically, the original contributions of this work are:

- the description of a novel distributed architecture for the automatic monitoring of traffic flows in Smart Cities;
- the definition of a taxonomy of traffic anomalies, so that they can be recognised by an intelligent automatic system, and the specification of a monitoring module based on Faster R-CNN, trained on synthetic and real videos, to estimate the probability that such anomalous events have occurred;
- the design of an original and efficient automatic number plate recognition approach, based on YOLO and Tesseract OCR;
- the development of a prototype implementing the above approaches, that supports the real-time execution on NVIDIA-based System-on-Modules to be embedded on Smart City lampposts;
- the validation of the proposed solution through preliminary experimental simulations, and by providing an in-depth analysis of the results obtained.

The remainder of this document is organised as follows. Section 2 briefly introduces relevant related work in the literature, whereas Sect. 3 describes in detail the proposed system and the approach on which it is based. Section 4 presents the evaluation results and discusses findings and limitations of our proposal. Finally, Sect. 5 ends the paper with conclusions and future directions where we are headed.

2 Related Works

The challenges tackled by this work have been topics of research from the 90s [29], falling within the broad scope of the anomaly detection, which took its first steps in the network security area [13,34] and then expanded to the wider field of surveillance [2]. Indeed, nowadays, the problems related to automatizing video surveillance and traffic monitoring tasks in Smart Cities [1], in particular those aimed at improving the safety of pedestrian and vehicles, are still open and remain vastly studied [21]. While earlier works mainly employed techniques based on pattern recognition and image manipulation [40] to extract useful data gathered from the sensors, today the developments in Internet of Things [19], Edge Computing [22], Machine/Deep Learning [31] and Computer Vision [26] are the prominent focus of the research.

In this context, Appathurai et al. [5] exploited an artificial neural network, with weights optimally selected through an oppositional gravitational search optimization, to implement a moving vehicle detection system. Through their approach, it is hence possible to gather data about vehicle tracking, vehicle normal speeds, movement examinations and vehicle classification. Makhmutova et al. [28] highlighted the usefulness of improving existing closed-circuit television systems with the application of intelligent information processing and data analysis. They employed computer vision methods to perform object detection and tracking and neural networks for real-time classification and detection of anomalies. They also tested the proposed system on real video streams provided by the city of Kazan.

Similarly, in [2], the authors focused on vehicle identification and counting, with the aim of identifying traffic congestions. They used blob analysis, background subtraction, a dynamic autoregressive moving average model, a boundary block detection algorithm and vehicle tracking to perform automatic vehicle counting. Chakraborty et al. [14] proposed an artificial vision-based smart city solution to the problems concerning traffic and transportation. They also implemented a human activity detection system to improve safety and security of Smart Cities.

This work aims to take a step forward in this direction, by exploiting different Deep Learning techniques in a coherent fashion, and by proposing a structure for the development and deployment of a comprehensive system that handles the problem of traffic safety in a Smart City environment.

3 The Proposed Solution

The proposed solution is designed for automatic traffic monitoring in Smart Cities, through the application of Deep Learning and Computer Vision techniques. The primary purpose of this solution is to carry out video surveillance of urban traffic, by implementing the distributed system shown in Fig. 1.

The solution consists of a network of *smart* lampposts, installed on the public street and equipped with a bullet camera and a general purpose *System-on-Module* (SoM), specifically a *NVIDIA Jetson AGX Xavier* board (performing a

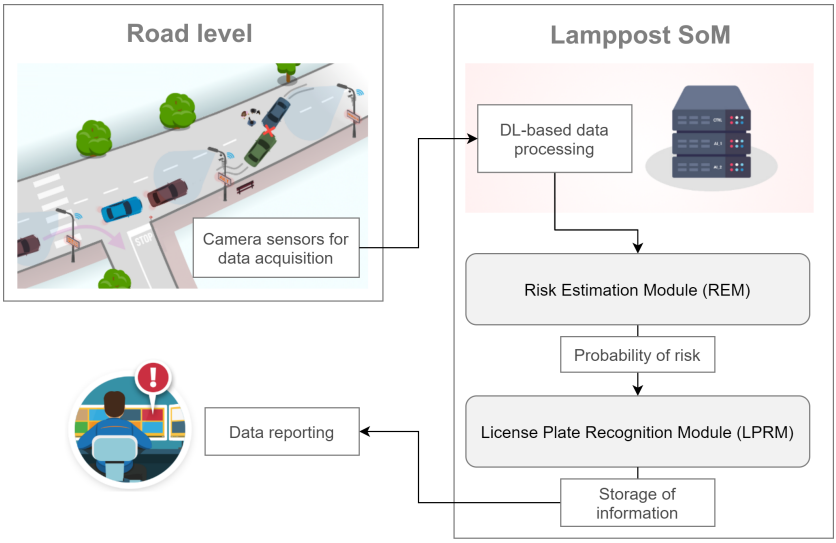


Fig. 1. The proposed solution.

Volta GPU with Tensor Core, an ARMv8 CPU and 32 GB of RAM), particularly suitable for running Neural Networks (NN) and Deep Learning (DL) algorithms on such embedded configuration.

Through this board, the lamppost is able to run the two modules based on Deep Learning and Computer Vision techniques, which perform automatic traffic monitoring functions: (i) a risk estimation module (REM), that automatically marks the segments of the entire observed video stream which are characterised by abnormal traffic conditions (i.e., anomalies are present); and, (ii) a license plate recognition module (LPRM), responsible for registering the license plate numbers of the vehicles present in each segment. In this way, the system is able to store only the information relevant for the traffic analysis, guaranteeing better storage space occupation and a longer time of data archiving. Moreover, such a specification is also effective where the public authority intends to permanently certify digests of these data in a distributed context, for instance by exploiting metadata storage on public blockchains [25,27,39]. In the remainder of this section, the approach adopted and the detailed functioning of these two modules are outlined.

3.1 Risk Estimation Module (REM)

The first of the two modules that make up the proposed solution, i.e. the one dedicated to the estimate the risk of abnormal traffic conditions, is described below.

Categorisation of the Anomalies. In order to build an effective risk detection module, part of the preliminary work on the development of this solution was

devoted to categorise the different typologies of anomalies that can occur in the urban environment and that can be (potentially) captured through the exploitation of Deep Learning techniques. Specifically, the taxonomy of anomalies employed in this study, illustrated in what follows, specifies 8 different macro-categories, each of them consisting of a set of subcategories, with the aim of highlighting similarities and differences among them. This categorisation is presented in Table 1.

It is important to emphasise that such a categorisation is not intended to cover all possible road anomalies. In the specific case of this study, it is in fact the result of an early analysis within the scope of the project of which this work is part. However, the rationale adopted was to mainly include, as previously pointed out, anomalies that were potentially dangerous to the movement of vehicles and other road users, as well as violations that could cause traffic disruption (such as parking in prohibited areas).

In addition, the anomalies taken into account are those suitable for automatic recognition by artificial intelligence techniques based on computer vision, such as those based on Deep Learning and exploited in this study. In this context, the macro-categories defined in Table 1 need to be further discriminated between *static* and *dynamic* ones, with regard to how time factors into the detection process. In detail, static anomalies are those than can be identified by looking at a static image, as they are mainly characterized mainly by the position and appearance of the involved entities (i.e., vehicles, pedestrians, *etc.*) in the scene; in contrast, dynamic ones necessarily requires to observe the behaviour and movements of such entities in a protracted interval, thus implying the time component. In particular, the macro-categories can be subdivided as follows:

- **Static**: traffic, entity on the road, collision, prohibited stop;
- **Dynamic**: failure to give way, illegal lane change, invasion, reckless driving.

Also note that the risk estimation module of the proposed solution is currently focused only on identifying static anomalies, through frame-by-frame analysis (starting from the video acquired by the lamppost camera) by exploiting a Convolutional Neural Network (CNN). On the other hand, the part concerning the analysis of dynamic anomalies is still under development and is not the subject of this paper.

Faster R-CNN and Ad-hoc Dataset Generation. A supervised learning approach was adopted to estimate the risk associated with the currently observed traffic situation. To do so, an artificial intelligence algorithm, specifically a neural network, was trained by exploiting a dataset of static anomalies (identified through the categorisation described in the previous paragraph) and in order to enable the recognition of moderate/high risk traffic situations.

As previously stated, the goal of this approach is not that of estimating the traffic risk with high accuracy, but rather to improve the performance of the traffic control system by marking as *relevant* those clips – acquired from the continuous loop recording – whose associated risk is estimated as higher than

Table 1. Taxonomy of anomalies.

Macro-category	Subcategory	Description
Traffic	Congestion	A large number of vehicles in the monitored area is proceeding at a very low speed
Entity on the road	Animal	An animal invaded the roadway
	Person	A person invaded the roadway (in the absence of pedestrian crossings)
	Object	An object in the roadway constitutes an obstacle to vehicles
Failure to give way	Red light violation	A vehicle engaged the crossroads with a red traffic light
	Intersection	A vehicle engaged the intersection while another vehicle with the right of way is approaching
	Roundabout	A vehicle engaged a roundabout while another vehicle with the right of way is approaching
Illegal lane change	U-turn	A vehicle has performed a U-turn
	Wrong way	A vehicle was driving on the wrong side of the road
	Continuous line	A vehicle crossed the continuous line of the road
Collision	Head-on	Some vehicles have been involved in a frontal impact
	Sideswipe	Some vehicles collided sideways with each other
	Rear-end	A vehicle impacted the rear of another vehicle
	Obstacle	A vehicle impacted an obstacle, a wall or another obstacle in the environment
	Pedestrian investment	A vehicle impacted with a pedestrian
Invasion	Bike lane	A vehicle invaded a bike lane
	Sidewalk	A vehicle invaded a sidewalk or a pedestrian area
	Off-road	A vehicle was driving off-road but neither on a bike lane or on a pedestrian area
Reckless driving	Hard braking	A vehicle halted in a sudden way
	Zigzagging	A vehicle is zigzagging or changing lanes abruptly
	High speed	A vehicle is driving at very high speed, above the permitted limits
Prohibited stop	No parking zone	A vehicle is parked on-road, in a no parking zone
	Illegal parking	A vehicle is parked off-road, where not allowed

normal. These clips are therefore stored for a longer period of time and remain available for possible subsequent human inspection; moreover, the number plate recognition algorithm is executed on them (in a subsequent non-real time phase) in order to identify the vehicles involved.

Given the performance improvement purpose of the risk estimation module, the approach is agnostic with respect to the type of neural network used to identify static anomalies. However, in the implementation of the experimental prototype of this work, the selected neural network was a *Faster R-CNN*, built on top of *ResNet* [20]. In our preliminary tests, in fact, the Faster R-CNN showed a discrete ability to identify static anomalies, thanks to its peculiarity of being able to analyse individual regions of each frame acquired. However, instead of the Faster R-CNN, it is also possible to leverage on a lighter *single shot detector* (SSD), e.g. *YOLO* [36], properly trained to detect the targeted anomalies, by sacrificing the accuracy of the risk estimation in favour of a higher supported frame rate. The architecture of the Faster R-CNN adopted in our module is shown in Fig. 2.

In order to train the network to recognize the static anomalies defined by the taxonomy in Table 1, we decided to create an *ad-hoc* training dataset, composed by about 800 synthetic videos, manually realized in order to cover each of the target categories with an adequate number of samples (in widely varying scenarios, shots, lighting and traffic conditions).

These synthetic videos have been generated through a Rockstar Games' editor, that exploits the traffic simulator engine of the video game *Grand Theft Auto V*, produced by the same company. The main reason behind this choice falls in the absence of literature datasets that include all the kind of anomalies that do not belong to the category of *collisions*, as well as the impracticality of creating a balanced and comprehensive dataset by capturing real video, in a reasonable amount of time. However, the simulations mimic the real-world scenario at a good level (i.e., traffic dynamics and abnormal events are realistic, while the consequences, such as impact damage, are less so, although they are not considered in this analysis), and even rare events can thus be added to the training set very efficiently. Sample footage depicting normal behaviour, i.e. without anomalies, is also generated and used so as to minimize the amount of false positives. A part of the dataset has also been used as a test set (out-of-sample), together with a small set of real videos, for the validation of the obtained model.

Finally, *Python3* and the `detecto` library are used to train and the deploy the Faster R-CNN, while the tool *LabelImg* to manually annotate images.

3.2 License Plate Recognition Module (LPRM)

The license plate recognition module (LPRM) acts downstream of the risk estimation module (REM), recognising and storing, fully automatically, the number plates of vehicles involved in a situation of potential anomaly or risk to traffic.

This module operates through a pipeline consisting of three fundamental steps: i) the plate detection; ii) the character segmentation (which includes rotation, threshold and blobs rejection); and iii) the character recognition. Such a

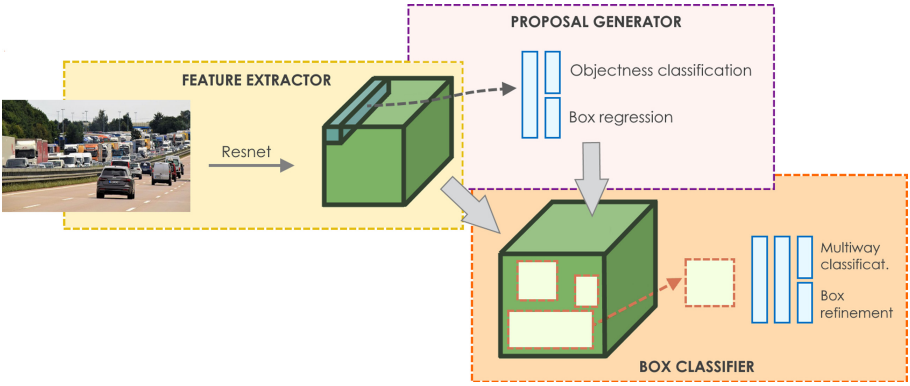


Fig. 2. Architecture of the Faster R-CNN used in the Risk Estimation Module.

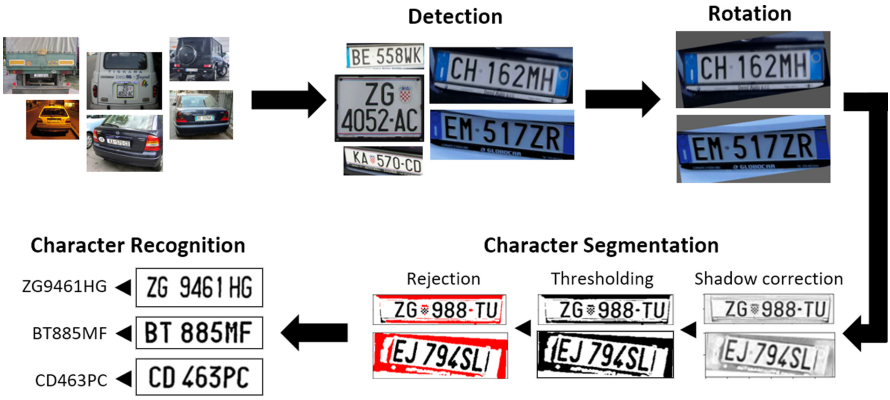


Fig. 3. The pipeline of the License Plate Recognition Module (LPRM).

pipeline is summarised in Fig. 3. The three steps are then described in the following of this section.

Plate Detection Phase. The plate detection phase is responsible for the identification and extraction of the plates within the video frames received by the REM. This step is performed by means of *ImageAI*[1], a Python library that acts as a wrapper for several deep learning algorithms and easily allows to train a custom model to detect objects from images or videos. In the proposed approach, the YOLOv3 algorithm has been used, trained on a dataset of license plates (described below) as to obtain a custom model.

Then, the detection algorithm takes as input an image and the generated model. Since YOLO can detect multiple objects of the same class within an

[1] https://github.com/OlafenwaMoses/ImageAI/releases/tag/essential-v4.

image, the plate with the highest probability is selected (with respect to the probability assigned by YOLO).

Figure 4 shows the output of the detection phase.



Fig. 4. Input (on the left) and output (on the right) of the detection phase.

Character Segmentation. Given the crop of the license plates, this step of LPRM pipeline aims to return an image containing only the characters of the license plate. This stage is performed through handcrafted pre-processing methods and it is divided, in turn, into four sub-phases: a) image rotation; b) image segmentation; c) anomalous blobs rejection; d) final refinements for the character recognition optimization.

The image rotation algorithm exploited by the LPRM represents an improved version of the one presented in [35]. First, an automatic thresholding is applied to generate a binary image, then edges are detected with the Canny algorithm. Finally, the module leverages on the Hough transform for the detection of the lines and their inclination, so as to straighten the image. Some improvements are applied to the Hough parameter space, i.e. a smoothing filter is used to remove outliers and the maxima pattern generated by the Hough transformation, making the angle detection more accurate. Furthermore, the rotation generates some empty space to entirely contain the final image and avoid cropping. An example of this rotation algorithm is depicted in Fig. 5.

After the image is rotated, the segmentation is performed. To do so, firstly the image is converted to grayscale and the height fixed to 100 pixels; then shadows are removed, by subtracting from the original image a copy of itself in which dilatation and smoothing filters are applied.

Subsequently, an adaptive thresholding based on the *Otsu's method* [32] is applied to extract the darker details. However, since several dark regions may have been lost due the removal of shadows, some isolated regions – whose shape may be similar to a character – can circumvent the blobs rejection. To overcome this issue, a map of darker areas is derived through a linear threshold and, secondly, dilation and closing filters, to make sure that the detected dark regions do not overlap with the characters. The final segmentation is thus obtained by subtracting the darker areas map from the processed license plate image.

From the previous step, the LPRM obtains a binary image containing the license plate characters fully separated from all the other regions. Then it

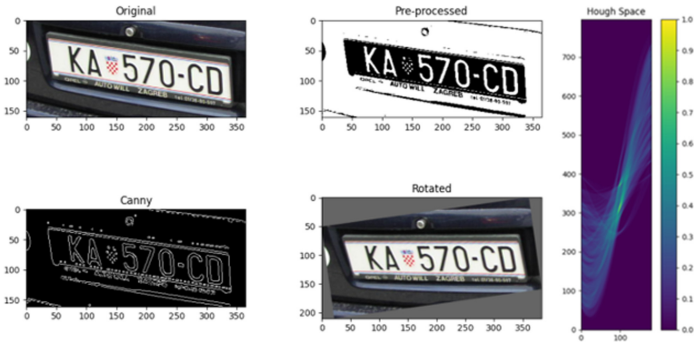


Fig. 5. An example of automatic image rotation.

needs to remove such anomalous regions using ad-hoc *blobs rejection* algorithms. Hence, three algorithms have been developed, based on *a priori* knowledge about the characteristic of the image; specifically, it is worth to note that the characters have a similar size and intensity to each other, and have a specific size within the image. Figure 6 graphically summarises the above steps.

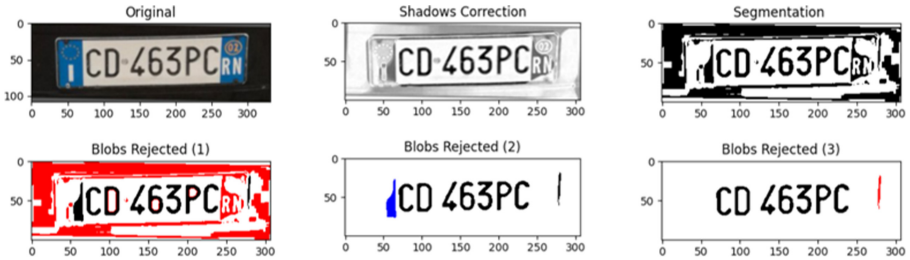


Fig. 6. The character segmentation sub-phases of the LPRM pipeline.

The last sub-phase of the character segmentation performs some refinements, aimed at optimizing the following recognition stage: all the white spaces are cut (substituted with a fixed padding), the image is resized, the height is fixed and the angle for a fine rotation adjustment is computed. Finally, a smoothing filter is applied.

Character Recognition. This is the final stage of the LPRM pipeline. It returns the extracted licence plate number by exploiting a majority voting approach. This stage leverages on an additional deep learning layer, based on the *Tesseract OCR*[2]. It is based on a Long-Short Term Memory (LSTM), lightweight and easy

[2] https://github.com/tesseract-ocr.

to use, which (compared to CNNs) required shorter training times while still ensuring high levels of accuracy.

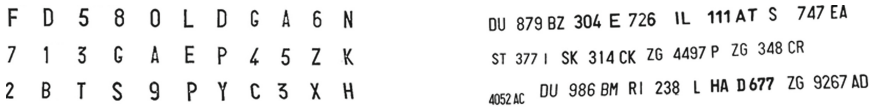


Fig. 7. Example of the images created to train Tesseract.

Over 60 images (Fig. 7) were tailored and used for training the LSTM, obtaining a custom model. In addition, the most accurate models in the Tesseract OCR repository were considered. However, despite of the optimizations, the reading of the images varied between configurations. To tackle this last issue, a majority voting system was built in order to combine the decisions of the individual models and thus to maximise the final accuracy.

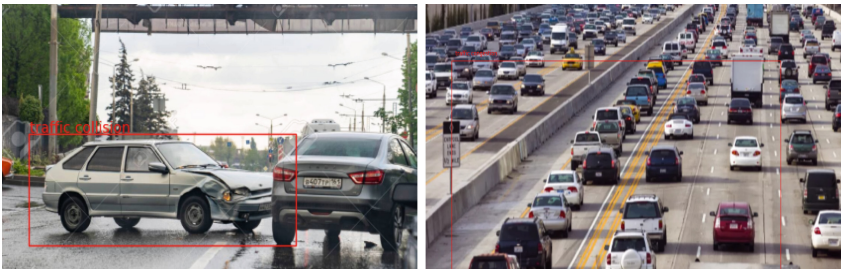
4 Results and Discussion

In this section, the preliminary experimental results for the proposed solution are shown and discussed, describing also the current limitations and possible future developments. The modules of the solution were developed and trained on a personal desktop configuration (Intel Core i7-8750H CPU @ 2.20 GHz with 32 GB of RAM and a GeForce GTX 1060 graphic card), while the smart lamppost operational execution has been simulated by providing, as input, a continuous stream of recorded videos to the Nvidia Jetson board.

In particular, a subset consisting of 80 test videos from the synthetic dataset (i.e., that created by means of the Rockstar editor), and 20 real footage videos, all of which have an average duration of between 20 and 120 s, were used to validate the effectiveness of the prototype and the proposed approach. We manually annotated the static anomalies in each video, with some videos having no anomalies. Moreover, we focus only on precision and recall as metrics, as the aim is to observe the effectiveness of identifying events in the `anomalies` class only (i.e., (the overall accuracy of the system in correctly identifying even *normal* behaviours is not of interest) (Figs. 8 and 9).

For what concerns the risk estimation module (REM), the proposed implementation showed, during the validation process, to have good recall values (>90%) but relative low precision (∼35%). However, as the function of the risk estimation module is only to filter the data that the license place recognition module has to parse and the system has to store, it is required to avoid the presence of false negatives, even if at the cost of having a high number of false positives. Indeed, this behavior is desirable: not detecting an anomalous event would have much more impact on the reliability of the system than the parsing of extra data has. On the other hand, with this solution, we achieve high system

performance and scalability, as the amount of interesting data parsed is maxed and the amount of useless data is significantly reduced when compared to the alternative (i.e., storing the full video stream).



(a) Traffic collision; (b) traffic congestion.

Fig. 8. Examples of static anomalies detected in real footage videos.

With regard to the license plate recognition module (LPRM), the prototype has been validated in two modalities: an operational validation (from the video buffers of abnormal events, extrapolated by the REM module from the 80 videos described above), and a static one (using a set of 164 real-world license plate images, mainly downloaded from *PlatesMania*[3]). Globally, we obtained an overall accuracy of 87.21%.

There are several causes that led to a failure in the recognition. For example, detection errors may be induced by crops that are too wide or narrow, or that do not fully include all the characters of the license plate. Segmentation errors are mainly due to characters that are not isolated correctly, while (very rare) reading errors may occur even if the image contains only well-segmented and isolated valid characters.

Notably, no errors has emerged during the object detection phase, in both datasets. Also the segmentation step showed to be susceptible to few errors, in part due to occlusions (usually not problematic, as the same number plate can be detected, and therefore recorded, in an earlier or later frame of the video stream) and in part to the connection with anomalous regions, even if for a few pixels (in this case, morphological operations with wider filters could solve the problem, but with the possible drawback of separating characters on damaged plates and consequent failure of the rejection sub-phase). Reading errors represent the main problem in the pipeline. Images with perspective distortions are particularly sensitive to errors, but also there exist some (infrequent) issues due to ambiguity between characters and numbers – e.g., for pairs like (S-5), (A-4), (Z-7), (G-C), (O-0) – that the voting system did not completely solve. However, as mentioned above, during the video stream it is still possible to detect the correct licence

[3] http://platesmania.com/.



(a) Object on the road; (b) traffic collision;

(c) prohibited stop; (d) traffic congestion.

Fig. 9. Examples of static anomalies detected in the synthetic test videos.

plate number in a different frame, so combinations that do not appear in a sufficient number of frames can easily be excluded (Fig. 10).

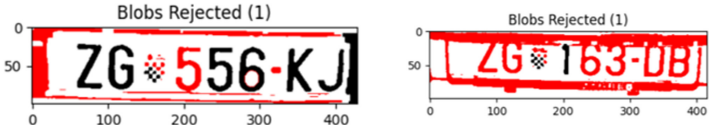


Fig. 10. Example of errors that may occur during the segmentation stage, which led to the rejection of valid characters. On the left, a character is linked to an invalid region; on the right, an occlusion disrupts the character separation.

5 Conclusions

In this paper, we have proposed a novel taxonomy of traffic anomalies, as a reference for artificial intelligence detection approaches, and we have presented an integrated solution for traffic control through automatic license plate recognition, based on Deep Learning and Computer Vision techniques, that leverages on the aforementioned taxonomy. The proposed solution consists of a network of smart lampposts, each equipped with a bullet camera and a System-on-Module (i.e., an NVIDIA Jetson board), capable of autonomously running two intelligent software components: a risk estimation module (REM) and a number plate detection module (LPRM). The solution aims at monitoring Smart City traffic automatically through a system of cameras placed on road lampposts, filtering

out normal traffic flows and marking as relevant only those in which - by means of a Faster R-CNN - an above-average risk index is detected, i.e. a static anomaly has occurred. Then, the license plate recognition approach, exploiting YOLO and Tesseract, enables the automatic identification of vehicles in the scene. Such a solution is, firstly, efficient, as it allows only relevant traffic events to be stored and processed, optimising space usage and allowing longer archiving; secondly, it is easily scalable, thanks to its decentralised nature. At this stage, however, there are still some open problems, towards which our future research work is oriented. First, the risk estimation module currently takes into account static anomalies only; the detection of dynamic anomalies is indeed more complex and, to deal with this limitation, a set of heuristics is currently being studied. Second, the use of a single camera makes each observation point susceptible to problems of occlusion, glare and perspective. In this sense, the combined processing of two parallel video streams, from different observation points, would improve the overall detection accuracy.

References

1. Ahad, M.A., Paiva, S., Tripathi, G., Feroz, N.: Enabling technologies and sustainable smart cities. Sustain. Cities Soc. **61**, 102301 (2020)
2. Al-Heety, A.T., et al.: Moving vehicle detection from video sequences for traffic surveillance system. ITEGAM-JETIA **7**(27), 41–48 (2021)
3. Al-Turjman, F., Lemayian, J.P.: Intelligence, security, and vehicular sensor networks in internet of things (IoT)-enabled smart-cities: an overview. Comput. Electr. Eng. **87**, 106776 (2020)
4. Albatish, I.M., Abu-Naser, S.S.: Modeling and controlling smart traffic light system using a rule based system. In: 2019 International Conference on Promising Electronic Technologies (ICPET), pp. 55–60 (2019). https://doi.org/10.1109/ICPET.2019.00018
5. Appathurai, A., Sundarasekar, R., Raja, C., Alex, E.J., Palagan, C.A., Nithya, A.: An efficient optimal neural network-based moving vehicle detection in traffic video surveillance system. Circ. Syst. Signal Process. **39**(2), 734–756 (2020)
6. Atzori, A., Barra, S., Carta, S., Fenu, G., Podda, A.S.: Heimdall: an AI-based infrastructure for traffic monitoring and anomalies detection. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), pp. 154–159. IEEE (2021)
7. Barra, S., Bisogni, C., De Marsico, M., Ricciardi, S.: Visual question answering: which investigated applications? arXiv preprint arXiv:2103.02937 (2021)
8. Barra, S., Carta, S.M., Giuliani, A., Pisu, A., Podda, A.S., et al.: FootApp: an AI-powered system for football match annotation. arXiv preprint arXiv:2103.02938 (2021)
9. Barra, S., De Marsico, M., Cantoni, V., Riccio, D.: Using mutual information for multi-anchor tracking of human beings. In: Cantoni, V., Dimov, D., Tistarelli, M. (eds.) Biometric Authentication, pp. 28–39. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-13386-7_3
10. Bock, F., Di Martino, S., Origlia, A.: Smart parking: using a crowd of taxis to sense on-street parking space availability. IEEE Trans. Intell. Transp. Syst. **21**(2), 496–508 (2020). https://doi.org/10.1109/TITS.2019.2899149

11. Bock, F., Di Martino, S.: On-street parking availaibilty data in San Francisco, from stationary sensors and high-mileage probe vehicles. Data Brief **25**, 104039 (2019)
12. Braun, T., Fung, B.C., Iqbal, F., Shah, B.: Security and privacy challenges in smart cities. Sustain. Cities Soc. **39**, 499–507 (2018)
13. Carta, S., Podda, A.S., Recupero, D.R., Saia, R.: A local feature engineering strategy to improve network anomaly detection. Future Internet **12**(10), 177 (2020)
14. Chakraborty, M., Pramanick, A., Dhavale, S.V.: MobiSamadhaan—intelligent vision-based smart city solution. In: Gupta, D., Khanna, A., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds.) International Conference on Innovative Computing and Communications. AISC, vol. 1165, pp. 329–345. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-5113-0_24
15. Cho, Y., Jeong, H., Choi, A., Sung, M.: Design of a connected security lighting system for pedestrian safety in smart cities. Sustainability **11**(5) (2019). https://doi.org/10.3390/su11051308, https://www.mdpi.com/2071-1050/11/5/1308
16. Choi, S., Kim, J.T., Choo, J.: Cars can't fly up in the sky: improving urban-scene segmentation via height-driven attention networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020
17. Combs, T.S., Sandt, L.S., Clamann, M.P., McDonald, N.C.: Automated vehicles and pedestrian safety: Exploring the promise and limits of pedestrian detection. Am. J. Prev. Med. **56**(1), 1–7 (2019)
18. Deng, J., Li, L., Zhang, B., Wang, S., Zha, Z., Huang, Q.: Syntax-guided hierarchical attention network for video captioning. IEEE Trans. Circ. Syst. Video Technol. (2021, in press)
19. Dhingra, S., Madda, R.B., Patan, R., Jiao, P., Barri, K., Alavi, A.H.: Internet of things-based fog and cloud computing technology for smart traffic monitoring. Internet Things, p. 100175 (2020). https://doi.org/10.1016/j.iot.2020.100175, https://www.sciencedirect.com/science/article/pii/S2542660519302100
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
21. Sri Jamiya, S., Esther Rani, P.: An efficient method for moving vehicle detection in real-time video surveillance. In: Suresh, P., Saravanakumar, U., Hussein Al Salameh, M.S. (eds.) Advances in Smart System Technologies. AISC, vol. 1163, pp. 577–585. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-5029-4_47
22. Khan, L.U., Yaqoob, I., Tran, N.H., Kazmi, S.M.A., Dang, T.N., Hong, C.S.: Edge-computing-enabled smart cities: a comprehensive survey. IEEE Internet Things J. **7**(10), 10200–10232 (2020). https://doi.org/10.1109/JIOT.2020.2987070
23. Khan, M.A., et al.: Human action recognition using fusion of multiview and deep features: an application to video surveillance. Multimedia Tools Appl. 1–27 (2020)
24. Malik, K.: Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking. In: Proceedings Ninth IEEE International Conference on Computer Vision, vol. 1, pp. 524–531 (2003). https://doi.org/10.1109/ICCV.2003.1238392
25. Li, W., Guo, H., Nejad, M., Shen, C.C.: Privacy-preserving traffic management: a blockchain and zero-knowledge proof inspired approach. IEEE Access **8**, 181733–181743 (2020)
26. Li, Y., et al.: Multi-granularity tracking with modularlized components for unsupervised vehicles anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2020

27. Longo, R., Podda, A.S., Saia, R.: Analysis of a consensus protocol for extending consistent subchains on the bitcoin blockchain. Computation **8**(3), 67 (2020)
28. Makhmutova, A., Anikin, I., Minnikhanov, R., Bolshakov, T., Dagaeva, M.: Detection of traffic anomalies for a safety system of smart city. In: Information Technology and Nanotechnology (ITNT-2020), pp. 638–645 (2020)
29. Michalopoulos, P.G.: Vehicle detection video through image processing: the autoscope system. IEEE Trans. Veh. Technol. **40**(1), 21–29 (1991). https://doi.org/10.1109/25.69968
30. Neves, J.C., Moreno, J.C., Barra, S., Proença, H.: Acquiring high-resolution face images in outdoor environments: a master-slave calibration algorithm. In: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8 (2015). https://doi.org/10.1109/BTAS.2015.7358744
31. Nguyen, K.T., Dinh, D.T., Do, M.N., Tran, M.T.: Anomaly detection in traffic surveillance videos with GAN-based future frame prediction. In: Proceedings of the 2020 International Conference on Multimedia Retrieval, pp. 457–463 (2020)
32. Otsu, N.: A threshold selection method from Gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
33. Pagliara, F., Mauriello, F., Di Martino, S.: An analysis of the link between high speed transport and tourists' behaviour. Tourism Int. Interdisc. J. **67**(2), 116–125 (2019)
34. Pang, G., Shen, C., Cao, L., Hengel, A.V.D.: Deep learning for anomaly detection: a review. ACM Comput. Surveys (CSUR) **54**(2), 1–38 (2021)
35. Piccinelli, L.: Raddrizzare il contenuto di un'immagine, November 2016. https://luca-picci.medium.com/raddrizzare-il-contenuto-di-unimmagine-37f9bbc16207
36. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
37. Sreenu, G., Durai, M.S.: Intelligent video surveillance: a review through deep learning techniques for crowd analysis. J. Big Data **6**(1), 1–27 (2019)
38. Suzuki, L.R.: Smart cities IoT: enablers and technology road map. In: Rassia, S.T., Pardalos, P.M. (eds.) Smart City Networks. SOIA, vol. 125, pp. 167–190. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61313-0_10
39. Yang, Y.T., Chou, L.D., Tseng, C.W., Tseng, F.H., Liu, C.C.: Blockchain-based traffic event validation and trust verification for VANETs. IEEE Access **7**, 30868–30877 (2019)
40. Fu, Z., Hu, W., Tan, T.: Similarity based vehicle trajectory clustering and anomaly detection. In: IEEE International Conference on Image Processing 2005, vol. 2, pp. II-602 (2005). https://doi.org/10.1109/ICIP.2005.1530127