CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Mobilitat

CIDAI-PAI 02-2021-D01

Suggested citation. CIDAI-PAI 02-2021-D01// Risk assessment of traffic accidents on urban and interurban roads in Catalonia by Elli Kartsakli, Eduard Quiñones – Barcelona Supercomputing Center (BSC), Albert Herrando – Centre de Telecomunicacions i Tecnologies de la Informació CTTI, Jordi Casals, Daniel Gonzalez, Rohit Kumar, Lali Soler – Eurecat, Josep Escrig, Ivan Huerta, Angel Martin – Fundació i2CAT, CIDAI, 2021. Licence: CC BY-NC-SA 4.0

Executive Summary

This project has contributed to the digital transformation of smart mobility, leveraging the use of Artificial Intelligence (AI) and big data analytics to achieve two specific objectives with a direct impact on society. The first objective aimed to enhance road safety and reduce the number of accidents by detecting hazardous situations involving both vehicles and pedestrians in urban and interurban scenarios. The second objective focused on the analysis of traffic flows, enabling the identification of traffic congestions and other related situations, which can lead to efficient traffic management policies.

The execution of the project has been organized in four technical work packages (WPs). First, the most relevant use case scenarios in both urban and interurban environments have been identified, along with their specific performance requirements (WP1). A prototype architecture has been designed, for the storage of input and output data, the analytics processing and the visualization of the project's results (WP2). Then, the necessary modules for the video and image analytics have been developed, consisting of two main blocks (WP3). The first block was responsible for the detection and tracking of road users (vehicles and pedestrians), providing information on their type, geolocation, speed, etc., as well as the status of traffic lights in urban environments. This information was further processed by the second block of analytics, applying the semantic annotation to detect the specified risk and traffic flow analysis events. The implemented methods have been validated and demonstrated (WP4), using recorded videos from street cameras targeting two highways and one urban intersection, covering several days, traffic and light conditions, etc.

Finally, a replicability study has been conducted, to identify the factors that affect the replication of the project's analytics methods to different scenarios and maximize the impact of the proposed solutions. Different factors have been identified, such as location, type of road, camera features and positioning, climate and lighting conditions, etc. Furthermore, some potential improvements of the proposed solutions have been described.

Overall, this project demonstrated the high potential of the proposed AI and big data-based methodology towards a safer and more efficient mobility. The specific solutions were oriented towards providing an offline analysis of the risk situations and traffic conditions observed in different urban and suburban environments. The quantitative and qualitative evaluation of the obtained results, supported by their visualization on the project's dashboards, enabled the extraction of useful conclusions that can lead to proactive measures, such as high-risk intersections.

Summary of the problem

The European Union is placing a lot of effort in improving road safety, with the ultimate goal of bringing the number of fatalities and serious injuries due to traffic accidents down to zero by 2050, an initiative known as "Vision Zero" [vision_zero]. The use of technology, and specifically Artificial Intelligence (AI) and big data analytics, can significantly contribute towards this direction, providing the tools and methodologies to implement both preventive and reactive solutions that can improve safety and efficiency of mobility.

In this context, this project has implemented a novel strategy to provide smart mobility solutions towards achieving two specific objectives. The first objective has been to enable the proactive prevention of traffic accidents, by identifying specific risk situations that are often observed in both urban and interurban scenarios. The second objective of the project focused on the traffic flow analysis of urban and interurban road segments, offering insights on the conditions that lead to traffic congestion and other anomalous road behavior (e.g., vehicles circulating with very high or very low speed).

In order to fulfill these objectives, the project first collected the necessary data, consisting of recorded videos covering three locations in the autonomous community of Catalonia. The first two locations represented an interurban scenario, targeting different segments of the C31 (one of Catalonia's primary highways), whereas an urban intersection within the city of Barcelona was selected as the third location. A set of representative use cases has been identified for the urban and interurban scenarios, covering typical risk situations highly correlated with traffic accidents, e.g., vehicles changing lanes over solid lines or pedestrians crossing with red.

Using these reference scenarios and recorded videos as an input, the project has elaborated and optimized the necessary AI and data analytics modules for the risk event detection and the traffic flow analysis. The analytics modules were organized in two main blocks, which were all executed on a prototype software platform deployed for the purpose of the project both on premise and at the cloud. First, the object detection and tracking block was responsible for detecting the different types of road users (cars, trucks, motos, pedestrians, etc.), following them across consecutive frames and extracting useful information such as their geographical position and speed. In addition, a traffic light detection module was elaborated, to derive the status of traffic lights in urban scenarios. In the second place, the appropriate semantic maps and rules were defined, allowing the detection of specific risk and traffic events, as required by the use cases. The output of the analytics modules was further processed, filtered and visualized in a dashboard, enabling the extraction of useful insights and observations from the analyzed data.

This document will provide an overview of the implemented methodology and the conducted work, describing in detail the defined use cases, the design of the software platform, the implemented analytics modules and the validation and analysis of the obtained results. Furthermore, this report will also highlight potential improvements and lessons learnt, as well as discuss the potential replicability of the proposed solutions in different settings.

Operational/ Technical Context

3

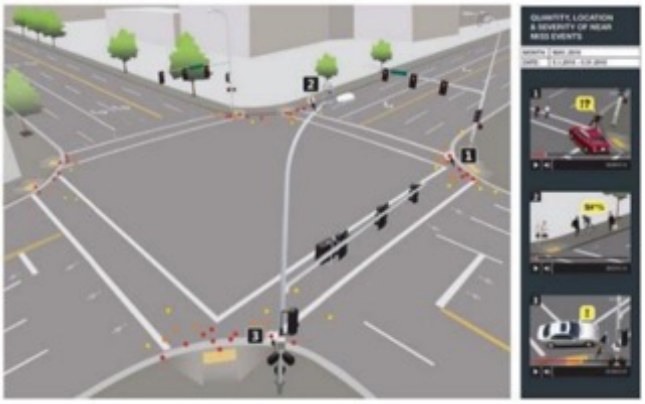
The autonomous vehicle, new forms of transport, traffic planning and management, such as the adaptation of real-time traffic light regulation, are a small sample of the fields of application of data convergence. The **prevention of accidents,** especially involving vulnerable road users such as pedestrians and cyclists, as well as the **analysis of traffic flows** to improve mobility and reduce pollution, are two important priorities in this field, motivating the two use cases targeted in this project.

According to www.barcelona.cat, there were 9,251 accidents in Barcelona in 2019 (around 25 accidents per day). Though different measures have been already implemented to reduce risks, achieving zero victims from traffic accidents is still a challenge faced by city authorities.

In the context of **preventing accidents,** the use case goal is to modernize a traditional reactive traffic safety approach, in which governmental agencies intervene only after enough police crash reports are filed to trigger a high crash corridor designation.

However, it is known there exist many other conflictive zones in which accidents may not finally occur, but risky situations are very frequent. The approach adopted in this project aimed to proactively determine risks using Video Analytics, Big Data and cloud computing.

On the other hand, the goal of the use case of **traffic flow analytics** is to analyse the traffic flow in order to recommend some actions that will improve mobility and reduce pollution.



Project Scope and Objectives

4

The overall scope of the project is to leverage the use of AI methods to extract valuable information in the context of smart mobility use cases, aiming to achieve a tangible social impact by enabling the design of enhanced traffic management policies and increasing the safety of road users. In this context, two general objectives have been defined, both based on video analytics, big data and edge, and cloud computing:

1. The first objective is to detect hazardous situations involving both vehicles and pedestrians in different scenarios (i.e., urban and interurban), fully aligned with the long-term aim of Vision Zero multi-national road traffic safety program towards preventing fatalities and serious injuries involving road traffic.
2. The second objective is to demonstrate the feasibility of identifying traffic congestions and other related anomalous situations through traffic flow analytics. This analysis can provide valuable insights on the conditions leading to traffic congestions, and enable the design of more efficient policies to mitigate their occurence.

In the context of this project, existing traffic camera systems have been considered to obtain the sample videos for the traffic flow and risk analysis, stressing the potential of the proposed solutions to leverage existing infrastrures.

To achieve the two overall project objectives, the project has defined six specific technical objectives (TO) that need to be fulfilled:

- TO – 01: Integrate heterogeneous data sources such as video cameras, traffic rules, semantic and traffic light data, to enable a specific analytics scenario.
- TO – 02: Define a series of known hazardous situations for both urban and interurban settings, considering a variety of scenarios, involving different types of vehicles, vehicle trajectories and speeds, pedestrians, traffic light status, etc.
- TO – 03: Design and implement an AI-driven approach to detect the occurrence of the risk events defined in TO – 02, based on the identified scenarios and data provided by TO – 01 for both urban and interurban scenarios, providing an indication of the risk factors of the analyzed zone.
- TO-04: Design and implement an AI-driven approach and the most appropriate visualization tools to provide a traffic flow analysis of the scenarios defined in TO – 01, also leveraging the video/image analytics modules developed in TO-03.
- TO – 05: Analyse the replicability of the models to other zones.
- TO – 06: Define the most suitable computing and communication infrastructure (including edge, and cloud resource) upon data analytics defined in OT-3 will execute

Work Performed

5

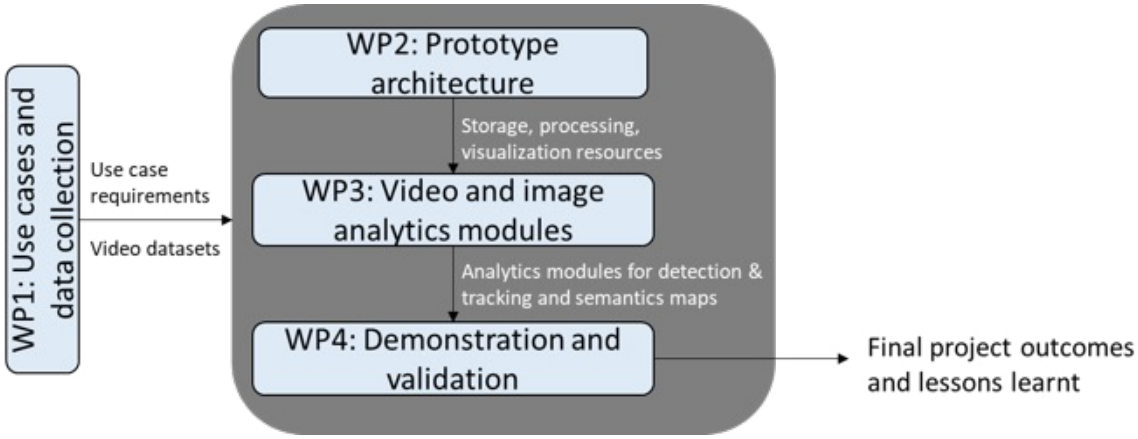
Pai Mobility explored the feasibility of anticipated traffic risk identification and traffic flow analytics for mobility improvement with a structure composed of four technical Work Packages (WPs). The WP structure and their interactions is shown in the next Figure 5.0. WP1 identified the project use cases and collected the most appropriate datasets (i.e., recorded videos). WP2 has designed a prototype architecture to handle the storage of both raw data and generated datasets, and the execution and visualization of the implemented data analytics modules. WP3 delivered the necessary modules for the video and image analytics, responsible for the detection and tracking of road users and the semantic annotation to enable the detection of specific risk and traffic-related events. Finally, WP4 was responsible for the final validation and demonstration of obtained results with respect to the risk event and traffic flow analysis. The work performed in each WP will be further detailed in the remaining of this section.

T1.1 Use case definitions.

This task was focused on the definition of the use cases that will benefit from the video analytics modules. Two different use cases were foreseen. The first use case focused on the detection of risk situations involving vehicles and pedestrians. The second use case refers to the analysis of the traffic flow from the information extracted from the video images. The context, objectives, sources of data and analytic methods applied in the use cases were also studied in this task.

Two different scenarios, interurban and urban, were defined for the risk use case because of the different nature of the risks that take place in each of them. Situations with pedestrians and crosswalks are only devised in urban scenarios, while forbidden lane changes or high-speed situations are more common in interurban scenarios.

Figure 5.0: Work Package structure



5.1 WP1: Project set-up, alignment and data sampling

During this work package, the teamwork, committees and resources were set-up. The worksite for experimentation was established based on its representativity and data availability and the stakeholders were requested to provide use cases examples and data samples of historical data to the different data sources (video cameras, traffic lights).

T1.2 Traffic data collection and pre-treatment.

The focus of this task was to identify and collect traffic images and videos. The footage needed to be similar to the ones expected in the defined use cases, similar scenarios, camera positioning (angle and distance) and image quality. This task was to check the availability of data among the different stakeholders, but also search for open data, open live streams, and data belonging to third parties. In this task we will also record new databases in case there is no data available for the specific use cases. Other types of traffic data different from images and videos were also gathered in this task.

In this task, the data was explored, and a first pre-treatment of the data was included (cleaning and detecting outliers). The structure of the database for the data storage was also defined, to be further used by WP3.

T1.3 Functional requirements and platform architecture design.

This task focused on the functional requirements of the platform to ensure a correct system execution, preserving all the time the privacy of the users visible on the traffic cams. This task also provided an initial design of the architecture and the required components of the platform, from a software and hardware perspective.

5.2 WP2: Prototype architecture setup

This work package focused on the design and execution of the overall prototype architecture for the storage, processing and visualization of the project's results. The design took into account the functional requirements of the analytics and the initial design provided in WP1. Furthermore, WP2 has been closely linked to WP4, in which the final setup of the prototype and the execution of the project's outcomes have taken place.

T2.1 Data/knowledge repository.

This task focused on specifying the data lake for the storage of all the project data. This data included:

- the raw videos covering both interurban and urban settings,
- the output of the detection and tracking module applied to each video frame, stored in the json format determined in T1.2 of WP1
- the output of the event detection module applied to each video frame, also stored in a predetermined json format, to be further leveraged by the visualization dashboard.

This task also involved setting up the data lake in Azure cloud and providing access to all partners to upload data and process data from the data lake.

T2.2 Data pipeline.

This task set up a scalable data stream for the ingestion and processing of the data, towards implementing the project use cases. This task mainly focused on the raw video processing and detection of events required in the later phase for a detailed analysis.

The data pipeline was set up using docker containers for each job.

T2.3 Analytical platform.

This task designed the analytical platform, along with all the functionalities needed. Different analytical platform options were analyzed and based on the use case requirements, a particular setup was selected for use.

As part of this task, the required components of the analytical platform were installed and configured in the Azure cloud environment.

5.3 WP3: Video and Image analytics modules

The third work package was dedicated to the development of the video and image analytics modules, including the definition of the semantics needed for the risk event detection and the data flow analysis.

T3.1 Object detection and tracking / tracking.

The focus of this task was to research the best algorithms for the detection of the different types of vehicles and pedestrians appearing in recorded video streams. T3.1 also researched the most suitables algorithms to track / trace the object's trajectory across consecutive frames of the same video. The task also implemented a traffic light detection algorithm, used for detecting the color of the traffic light from the video images, needed for the urban use case scenarios. The output of this task was a module able to detect and track a multitude of vehicles and pedestrians, as well as the status of traffic lights, providing information such as the vehicle type, position, orientation, etc.

Figure 5.3.1: Detection and tracking on a highway road

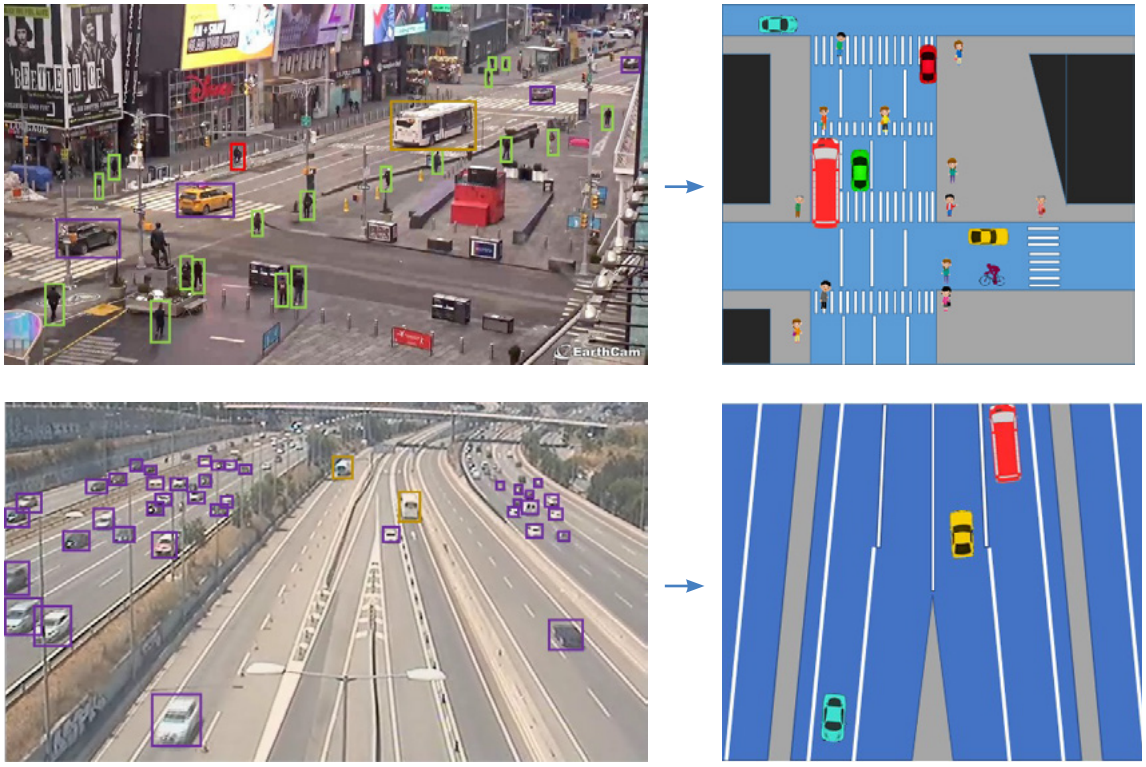
T3.2 Calibration and semantic model generation.

This task was focused on producing a module to transform the relevant position of the objects in the images (in pixels) into a real geographic positioning. This required the deployment of calibration methods considering the camera type, position and orientation and marker positioning on the scene.

Additionally this task was focused on researching the best tools to generate semantic models of the traffic scenarios from the images. These models included the necessary elements to deploy the use cases, such as vehicle lanes, directions of the lanes, bus stops, zebra crosses, pavements, cycle lanes, etc.

The output of this task was a module converting the video images with the detected and tracked objects into a semantic model, similar to the ones represented in the next figure.

Figure 5.3.2: Objects positioned on the semantic map. Vehicles and pedestrians detection on the left image and their position on the right image for urban and interurban scenes.



5.4 WP4: Demonstration of the use cases.

WP4 focused on the final implementation of the two use cases of the project, for risk detection and traffic flow analysis, demonstrating and validating them using real datasets. Within this WP, the more appropriate videos for processing have been selected for both urban and interurban scenarios, for the application of the analytics and semantic modules delivered by WP3. An iterative process has also been established between WP4 and WP3, through the continuous validation of the obtained results and the respective improvement of the analytics methods.

T4.1 Traffic flow analytics use case.

This task covered the traffic flow analysis, based on the results provided by the analytics modules developed in WP3. Leveraging the information on the position and speed of tracked vehicles, this task considered two relevant events for the traffic flow analysis, namely:

- the presence of congestion in specific areas, defined as the accumulation of vehicles within the areas of interest, which were delimited through the semantic annotation by T3.2,
- and the detection of vehicles moving with abnormal speed, i.e., speeds outside the expected range for the road segment under observation.

T4.2 Risk detection use case.

This task tackled the use case of detecting risk situations caused by the traffic in urban and interurban locations. Based on the output of the analytic modules developed in WP3, this task detected the occurence of specific risk situations, such as non-permitted changes of lane, or pedestrians crossing outside the designated areas. A quantitative and qualitative analysis of the obtained results have taken place, to validate the performance of the implemented analytics and identify any potential improvements beyond the lifetime of the project.

T4.3 Analytical platform set up.

This task implemented the prototype of the analytical platform designed in WP2, considering the heterogeneous ingestion of data (cameras, traffic lights, etc.), the execution of the risk identification models, the corresponding KPIs visualization, and potentially, vehicular communications to provide prescriptive management of risky situations.

T4.4 Demonstrator setup.

This task focused on the final demonstration of the work implemented in this project, showcasing the capability of the deployed platform and the implemented analytics and semantic modules to deliver the functionalities envisioned in the project use cases. The final demonstration included the implementation of dashboard with five different panels, used for:

- the visualization of the road user detection events, implemented in T3.1
- the visualization of vehicle speed data, provided by the tracking module implemented in T3.1
- the traffic flow analysis, implemented in T4.1
- the visualization of the detected risk events, provided by T4.2, for interurban scenarios
- the visualization of the detected risk events, provided by T4.2, for urban scenarios

In addition, short reference videos demonstrating examples of the correct detection of risk situations in both urban and suburban scenarios have been delivered.

Key Project Results

6

The key project result is a system that can perform a risk and traffic flow analysis of the scene based on the detection, tracking and geoposition of multiple mobile objects captured by RGB sensors.

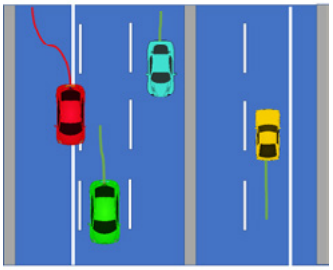
6.1 WP1: Project set-up, alignment and data sampling

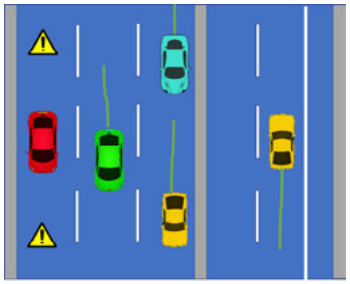
Use cases definition

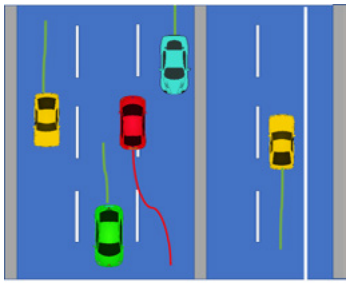
Two use cases have been defined focusing on (1) risk event detection, and (2) flow analysis. With respect to the risk use cases, two different cases have been considered, depending on the type of scene, distinguishing between urban and interurban scenarios. The type of the road, its users and their behavior are different between the two types of scenes, resulting in different definitions of risk events, which will be described below.

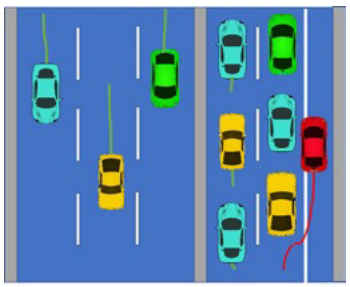
Interurban risk description

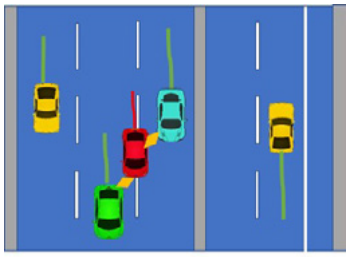
The following tables describe the different interurban risk cases contemplated in this project, providing a definition of the risk conditions and some visual examples.

Interurban Risk 1		
Risk description	Illegal lane change	
Conditions to be considered risk	Diagram	Camera exemple
The illegal lane change will always be considered a risk situation, regardless of the traffic intensity.		

Interurban Risk 2		
Risk description	Stopped vehicle on the roadway	
Conditions to be considered risk	Diagram	Camera exemple
This event will be considered a risk when the following conditions are met: • The speed of the vehicle is zero • No traffic signs that may cause a vehicle to stop (e.g., stop sign, give way, etc.) • Vehicle stopped with no other vehicle stopped in front of it		 

Interurban Risk 3		
Risk description	Vehicle going the wrong way	
Conditions to be considered risk	Diagram	Camera exemple
The event of a vehicle circulating on the shoulder lane will always be considered a risk situation, regardless of the traffic intensity.		 

Interurban Risk 4		
Risk description	Vehicle on the shoulder	
Conditions to be considered risk	Diagram	Camera exemple
The event of a vehicle circulating on the shoulder lane will always be considered a risk situation, regardless of the traffic intensity.		 

Interurban Risk 5		
Risk description	Vehicle too close to other vehicles	
Conditions to be considered risk	Diagram	Camera exemple
This event will be considered as a risk when the vehicle exceeds a minimum permitted distance between vehicles. The distance will be defined depending on the traffic density and vehicle speed.		 

WP1 initially defined two more interurban risks, namely interurban risk 6 concerning vehicles that fail to stop before a stop sign, and risk 7, concerning vehicles that fail to yield priority.

Two specific cameras were selected for the detection of these events, with examples shown in the next figures. However, these cameras were not working properly at the time when the video sample databases were created, and as a result these two risks were not evaluated in the project.

Interurban risk 6: Vehicles that fail to stop a stop sign

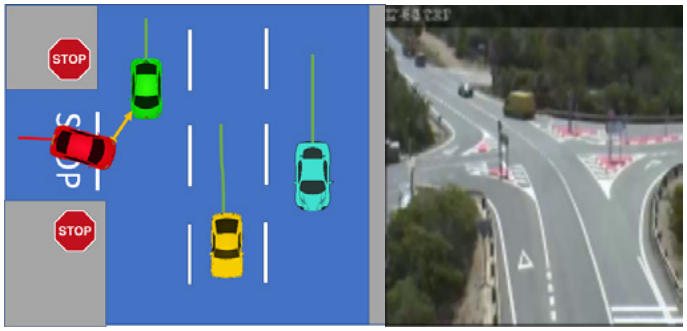


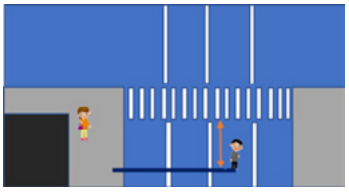
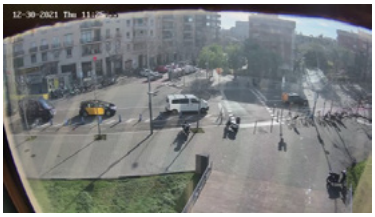
Figure: Diagram and camera example for the Interurban risk 6, vehicles that fail to stop a stop sign, not tested at the end because problems with the traffic camera selected

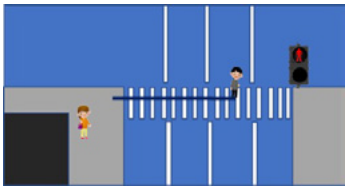
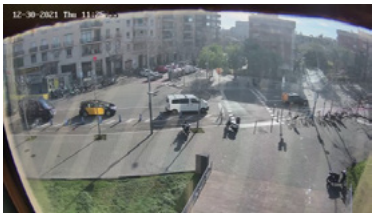
Interurban risk 7: Vehicles that fail to yield

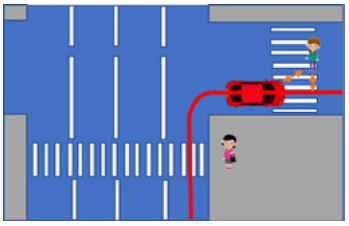
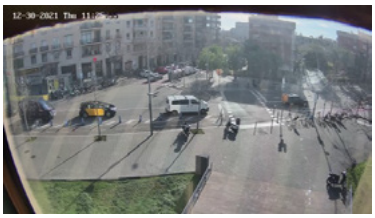
Figure: Diagram and camera example for the Interurban risk 7, vehicles that fail to yield. Not tested at the end because of problems with the traffic camera selected

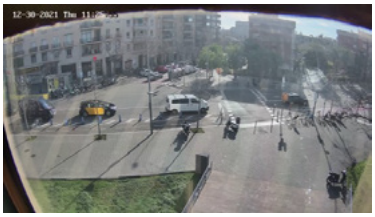
Urban risk description

The following tables describe the different urban risk cases contemplated in this project

Urban Risk 1		
Risk description	Pedestrians crossing outside of the crosswalk	
Conditions to be considered risk	Diagram	Camera exemple
It is consider risk when a pedestrian is crossing outside of the crosswalk		

Urban Risk 2		
Risk description	Pedestrians crossing at crosswalk when the signal is red	
Conditions to be considered risk	Diagram	Camera exemple
This event is considered a risk if the pedestrian is crossing the crosswalk when the traffic light is red for a minimum period of time		

Urban Risk 3		
Risk description	Car is making a turn and approaching too closely a pedestrian that is crossing at crosswalk of the street the car is turning into	
Conditions to be considered risk	Diagram	Camera exemple
This event is considered a risk when the car does not leave a minimum safe distance (frontal or lateral) with respect to the pedestrian, calculated by taking into account the vehicle's speed and the time needed to brake.		

Urban Risk 3		
Risk description	Car is making a turn and approaching too closely a bicycle that is crossing at crosswalk of the street the car is turning into	
Conditions to be considered risk	Diagram	Camera exemple
This event is considered a risk when the car does not leave a minimum safe distance (frontal or lateral) with respect to the bicycle, calculated by taking into account the vehicle's speed and the time needed to brake.		

Databases description

In order to proceed with the implementation of the project use cases, as defined in the previous section, two different types of scenes were required, referring to interurban and urban sequences. In order to acquire the necessary datasets, three different providers have been analyzed, as displayed in the next table.

Data type	Source
Interurban scenes private data	CTTI / SCT
Urban scenes private data	i2CAT
Urban scenes free data	Open data from internet

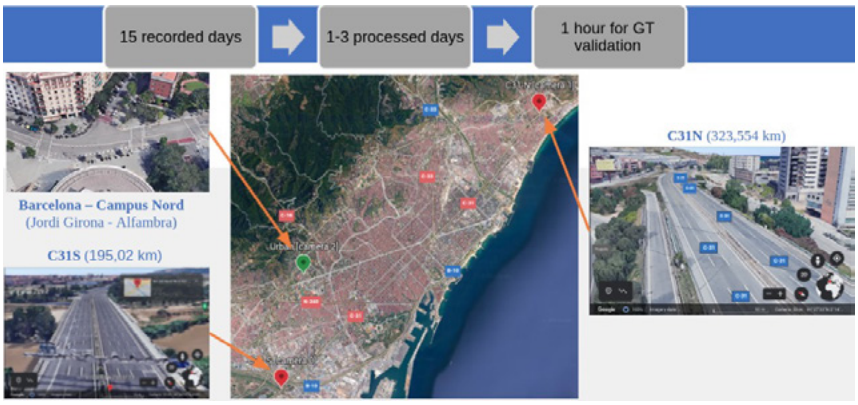
One of the stakeholders of the project, the "Servei Català de Transit (SCT)/Civicat" has around 600 interurban cameras (their own or third party cameras), with the possibility to select the best point of view from both directions and intersections inside of the camera field of view, resulting to approximately 1500 possible camera views. These cameras are able to capture different perspectives of the traffic, from frontal planes to lateral planes with different lateral angles, with enough detail and broad level to visualize multiple objects and their trajectories.

However, the stakeholders were not able to provide the project with access to urban cameras. As a first approach, an extensive research was done for free sources of urban video samples on the internet. However, no suitable sources were found to cover the defined urban risk use cases, and therefore urban sequences were recorded from one of the project partners.

In order to display the feasibility of the system for operation in interurban and urban scenes three cameras have been selected, two interurban cameras best matching the use case requirements and one urban camera provided by the project.

The next figure shows the location and 3D view of each of the three selected cameras.

Figure 6.1.1: 3D view and camera position



Video samples from the three cameras were recorded for approximately fifteen days. However we did not process the complete data set due to problems with the recorded videos (e.g., the camera not maintaining a fixed orientation) and the timing constraints for processing and analyzing all the data within the duration of the project. To that end, from these fifteen days, a subset ranging from one to three days for each camera was finally selected for processing, more than 150 hour analyzed. Interurban sequences have a framerate of 8 fps, and urban sequence of 25 fps but it is processed at 5 fps. The selected video segments included both day and night scenes.

With the purpose to validate the entire system quantitatively, one hour of recording per camera was selected to create the ground truth for the risk use cases.

The next figure shows an example of the three selected cameras:

Figure 6.1.2: Example of the three cameras selected for the databases. C31S, C31N, and Urban from left to right images respectively.

The sequences were recorded 24 hours in order to test the proposed solution on real conditions. In the next figures an example of the different illumination conditions can be seen :

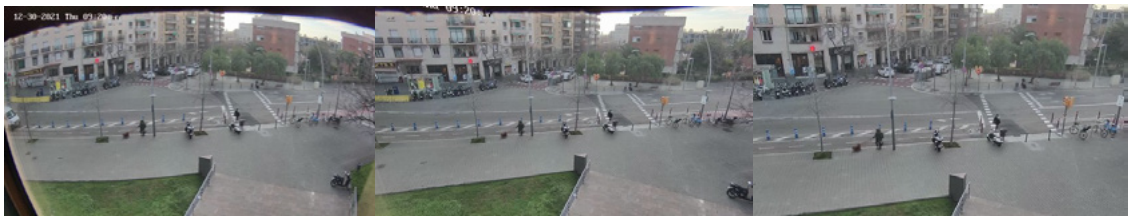
Figure: 24 hours example of C31N database

Figure: 24 hours example of C31S database

Figure: 24 hours example of urban database

For the urban scene recorded from one of the partners, a preprocessing step was proposed to improve the quality of the image and avoid image distortion.

Figure: Urban database preprocessing step. Original image, undistorted image and cropped image to 1080p resolution for urban database, from left to right respectively.



6.2 WP2: Prototype architecture setup

In this section we present the architecture of the PAI platform for analyzing the traffic video data and perform the required mobility analysis. In section 6.2.1 we will first present the logical design of the platform describing the different components of the architecture and finally in 6.2.2 we will present the physical design describing the implementation of the logical design in Azure cloud infrastructure.

6.2.1 Logical Design

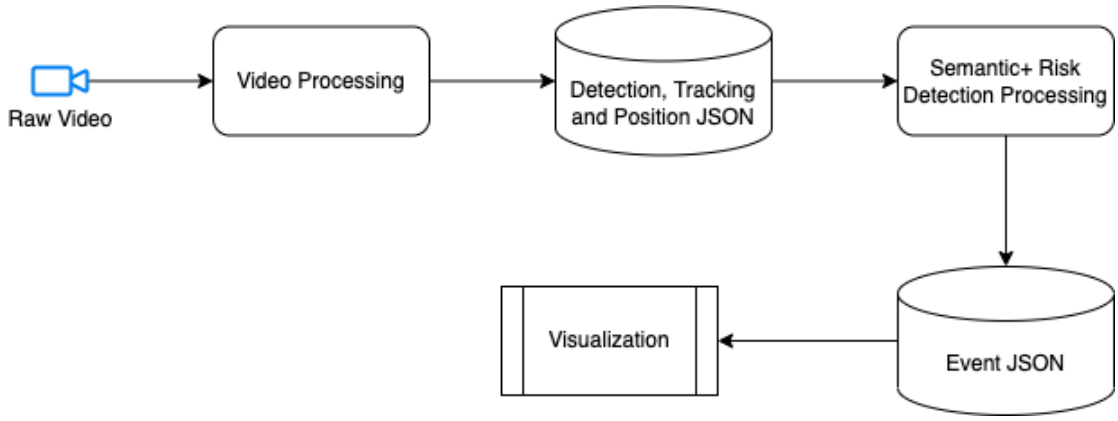
Figure 1 shows the logical Architecture design, consisting of three storage spaces:

- Raw Video: This storage layer was used to store the raw video recordings from the cameras.
- Processed Frame JSON: This layer was used to store the json data generated after processing the video frames. One json was generated to represent each frame of the video, stored in a specific storage path.
- Event Analysis Database: This layer stored the summarized and analyzed data from different processing modules, such as object detection and position tracking, and was also used for the final reporting and visualization.

The platform architecture defined two processing layers:

- Video Processing: This layer contained the GPU processors for processing the video data and storing the results by object detection, object tracking and object position. The output is stored as JSON.
- Frame JSON Processing: This layer was used for the execution of the different event processing workloads. such as using semantic data and video processing data to do risk detection and store the result in the final Event Analysis database for final reporting and visualization.

Figure 1 – Logical Architecture design.



6.2.2 Physical Design

Figure 2 shows the physical infrastructure setup for the logical design shown in Figure 1.

Storage:

- Azure blob Storage: The first two layers of Raw video and Processed Frame JSON were stored in separate containers in Azure datalake V2 storage account. Below is the folder structure used to store the data
- paidata: Root container for the storage account.
- paidata/rawvideo: Contains the raw video frame data as well the initial processed frame data in different folders
- paidata/processedjson: after using the semantics and frame data processed json for the analysis was saved here. The elasticsearch database is enriched from this data.
    - paidata/processedjson/track_res: data processed json for the detection, tracking and geoposition data
    - paidata/processedjson/risk_res: data processed json for the risk event detection
    - paidata/processedjson/traffic_res: data processed json for the traffic light color detection
- ElasticSearch: For event analysis Database elastic search was used. We hosted a self-managed virtual machine (VM) with ElasticSearch 7.14.2 database. We used the official docker image of the Elastic Search to host the database and mount the storage in Azure managed Disk attached to the VM. Kibana was installed in the same VM using the official docker image for Kibana.

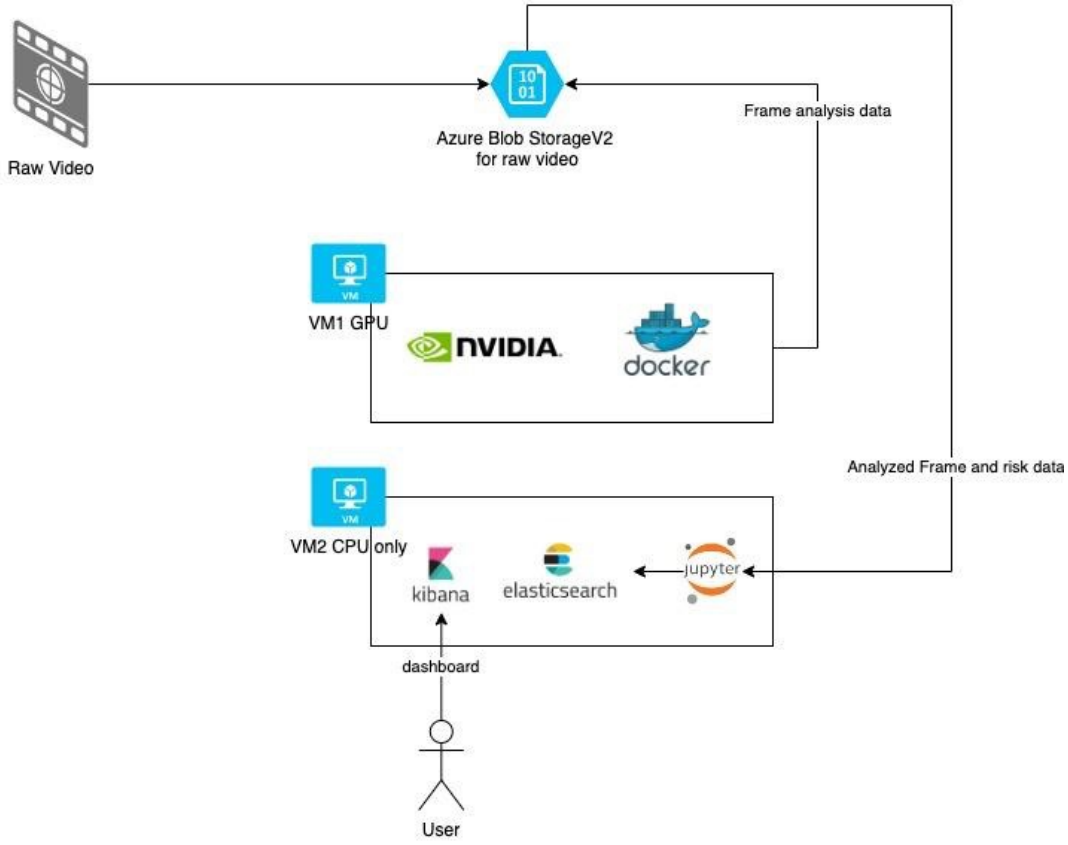
Processing:

- Video Processing: A GPU VM was configured to run the processing jobs. The jobs were dockerized and run on demand whenever new video files are uploaded.
- Frame JSON Processing: Jupyter Notebook was set up to process the Frame JSON and save the output into ElasticSearch for visualization using Kibana. The Jupyter Notebook server was installed in the same VM where elasticSearch and Kibana were installed.

Visualization:

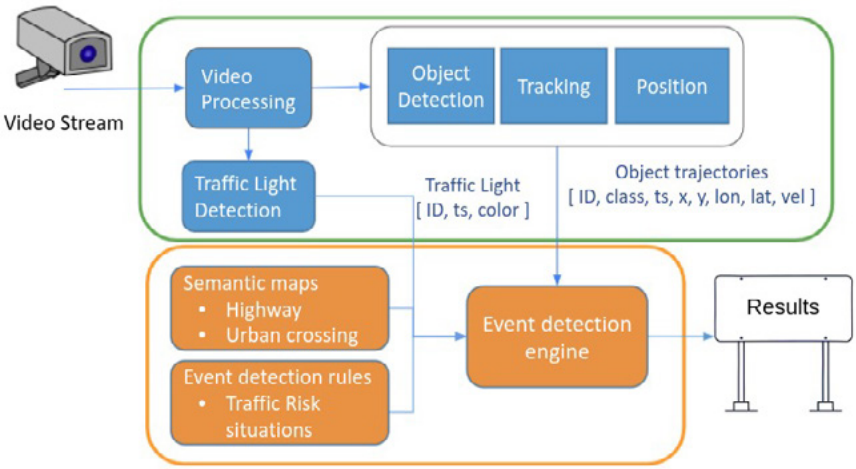
- Kibana: The VM created for ElasticSearch was also used to host the Kibana service, for the analysis and visualization of the data in ElasticSearch, used for the generation of reports and the dashboards.

Figure 2 – Physical Architecture design.



6.3 WP3: Video and Image analytics modules

Figure AI_over: Workflow AI modules for the risk event detection



Intro

In WP3 all the AI modules needed for the video and image analytics, and the semantics needed for the risk detection were developed. A general overview of all the components for the risk and traffic flow analysis can be seen in the figure AI_over, organized in two blocks marked by a green and orange box, respectively. In the first step (green box in figure AI_over), deep learning approaches were used to perform object detection, tracking and localization over the input video, in order to obtain the road user trajectories (ID and timestamp), with their class, dynamic information such as velocity (vel), their position in the image (x, y) and in the real world (longitude, and latitude). In this first step, we will also use image processing to evaluate the color of the traffic lights in the urban scene. In the second step (orange box in the figure AI_over), we defined the semantic map corpus for the two types of scenes needed for the risk event detection and the data flow analysis. Finally, in this second step we will also define the rules for the risk event detection.

Object detection and tracking

The first step for the risk and traffic flow detection was to analyze the video input and extract all the information about the objects in the scene. All objects in the scene should be detected and tracked, to determine where they are, when they are detected, which motion they exhibit, and which type of objects they are. In order to obtain this information a thorough review of the literature was conducted, selecting the best possible solutions and testing them for the type of scenes that we were evaluating (urban and interurban scenes).

Several requirements from the WP1 were taken into consideration in order to select the best solutions. The type of objects to detect, in our case, were non-rigid objects such as pedestrians, and rigid objects such as cars, motorbikes, bicycles and trucks. With respect to the quantity of the objects in the scene, in our case, multiple objects needed to be detected and tracked. An important factor to consider was how the existing approaches accurately solve the detection and tracking of small objects (in our case far away objects) and occlusions. There are multiple trackers in the literature [2, 3, Yilmaz06, Berclaz06, Evangelidis08, Huerta15, Kim15, Braso20]. But, they can mainly be divided into two types of trackers, online and offline trackers. In the online trackers [Bewley16, 50, Wojke17] the image sequence is handled in a stepwise manner, while in the offline trackers [Braso20] a batch of frames is utilized to process the data. Usually offline trackers are more computationally costly and do not perform near real-time. Even though the video analytics and the posterior risk data analytics did not require real-time performance, the big volume of data to process and the time constraints of the project led us to reject the option of offline trackers, since the obtained performance was too slow, not enabling us to process all the data on time for the end of the project.

The selected tracker follows the tracking-by-detection paradigm. It is a multi-object tracker that is part of the new family of trackers that learns the object detection task and appearance, embedding the task simultaneously (JDE) in a shared neural network [Wang20]. It is based on a siamese network for object tracking which has been employed on visual and multi-object tracking with very promising results [Berinetto16, Voigtlaender20, Shuai21]. The siamese tracker [Shuai21] models the motion of instances across frames and it is used to temporally link detection in online multi-object tracking. Detection is performed using an end-to-end detector (Faster-RPN [Shaoqing15]) with a motion model and adds a region-based Siamese tracker to model instance-level motion. The tracker is a region-based multi-object tracking network that detects and associates object instances simultaneously and uses deep features as visual cues for data association. The tracker improves the motion model of trackers such as [Bewley16, Wojke17].

However, at testing time we have seen that the joint detector with tracker was not enough to cope with small objects in the scene and it was not working correctly with low light conditions, as can be seen in the Figure Simamot_res. So, we decided to perform a previous object detection of the objects of the scene as input to the tracker. Several object detectors were analyzed and based on their performance and accuracy on the Coco database were selected for comparison [He17, Lin17, Redmon18, Long20, Bochkovskiy20, Tan20, ChienYao21]. Scaled-YOLO_v4 [ChienYao21] from the yolo family is the one that has provided better results in general as it can be seen in the figure cmp_yolo_efficentDet compared with EffiencientDet family detectors [Tan20].

Figure Simamot_res: Detection and tracking using siammot for C31N at daytime with backlight, and C31S at night time, small objects are not detected and the approach is not correctly working with poor light conditions.

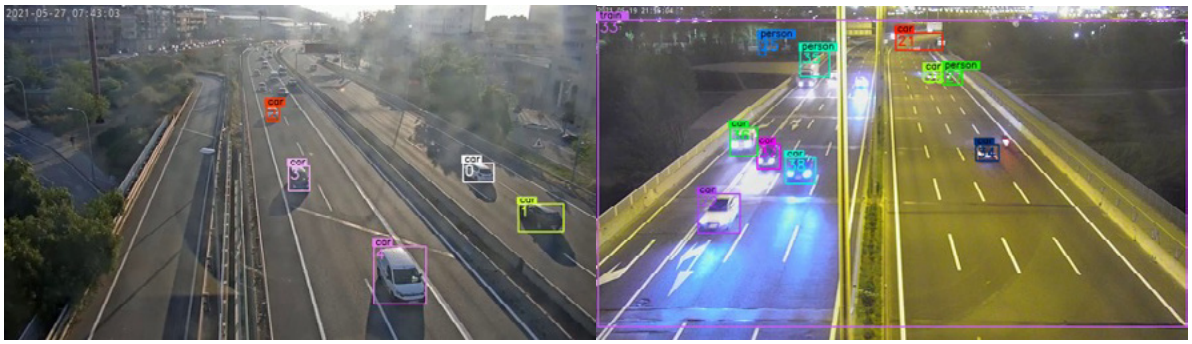


Figure cmp_yolo_efficentDet: Results comparison between EffiencientDet and Scaled-YOLO_v4 approaches for C31N at daytime with backlight, and C31S at night time, first and second row respectively. Better results are achieved using scaled-yolo_v4 in general.

Using the detection outputs as the inputs for the tracker has improved the accuracy considerably, especially for the small objects. However, we have detected some general problems, which are briefly discussed next. Motorbikes were not detected when they were far from the camera, but cars and people with the same size were detected. Also, performance is affected by the lighting conditions. In general, at night the detector is still working when there is some minimum light on the scene. The traffic lamps on the highway for the evaluated interurban scenes are enough to have some detections but the accuracy ratio has decreased considerably compared with the daytime, whereas the size of the vehicles to be detected at night needs to be bigger than at daytime. The detection for the urban area was not possible at night because the light was not enough. Regarding the tracker, the main observed problems were caused by occlusions and the switching of the objects' ids. More qualitative results of the detection and tracking will be presented at the risk event analysis section and their direct influence with the risk and traffic flow data analysis.

The algorithms that were finally employed were selected for their general accuracy. However, the project's pipeline is modular, allowing the selected components to be interchanged with more suitable approaches in the future without affecting the other parts of the project. Better object detection methods could enable the detection of smaller objects and better tracking methods would be able to solve abrupt appearance changes and severe object occlusions in the future. The modularity of the solution presented allows the exchange of these detectors and trackers for better ones which could increase the accuracy of the final risk event detection and traffic flow analysis.

Geoposition and dynamics

In order to compute the real geo-position for the objects detected on the pixel image we needed to follow two steps. First, compute the mapping between pixels from camera view to Earth's surface view (bird's eye view), and second, compute the mapping between pixels from bird's eye view to real world longitude and latitude coordinate system.

For the first mapping we needed to compute the perspective transformation (3x3 matrix) from the camera view to the Earth's surface (bird's eye view). This matrix was used to map the pixel coordinates from camera to bird's eye point of view. To compute the perspective transformation [1], four virtual markers were selected for each of the cameras and the corresponding virtual points were selected on the bird's eye view, as it can be seen in the next figure for the C31S camera. The same process has been done for each of the selected cameras .

Figure 6.3: Perspective transformation from C31S camera point of view to bird's eye view. Virtual markers are blue markers and orange marker is the camera position

Once the objects were situated on the earth's surface image, an approximate mapping between the 2D pixel image and the real world longitude and latitude was computed using the four virtual markers and their real world position. With the real object position computed previously and the id data and timestamp obtained for each of the objects provided by the tracking module, we were able to compute the dynamics for this object, such as the velocity, following the next equation:

$$\underline{v} = \frac{pos_2 - pos_1}{t_2 - t_1}$$

, where the object displacement between two positions (pos) is divided by the time (t) needed to make this displacement.

Some results from each of the evaluated sequences can be seen on the next figures, showing the correct detection of all objects on the scenes.

Figure: Detection, tracking, velocity and geoposition for each of the elements of the C31S sequence. Left image: camera view, right image: Bird's eye view. Bounding box for each of the elements detected and tracked with their class, identifier, and velocity over each of the bounding boxes respectively.

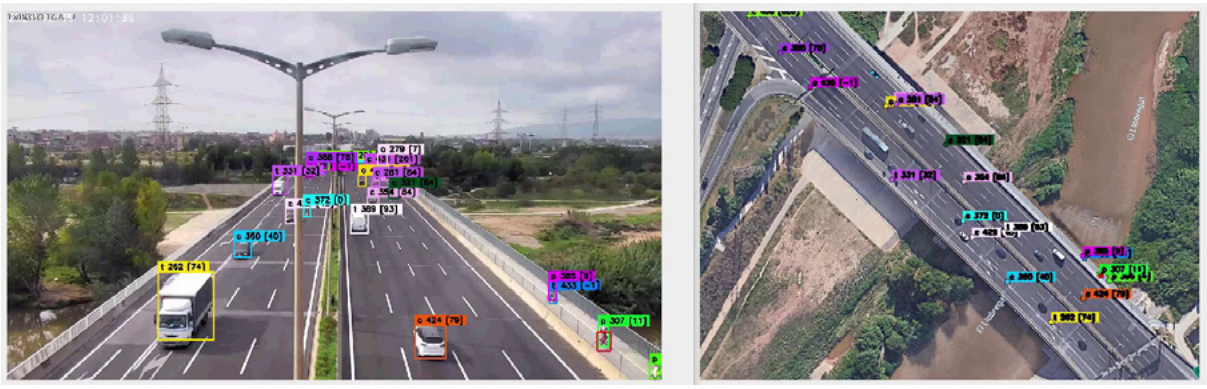


Figure: Detection, tracking, velocity and geoposition for each of the elements of the C31N sequence. Left image: camera view, right image: Bird's eye view. Bounding box for each of the elements detected and tracked with their class, identifier, and velocity over each of the bounding boxes respectively.

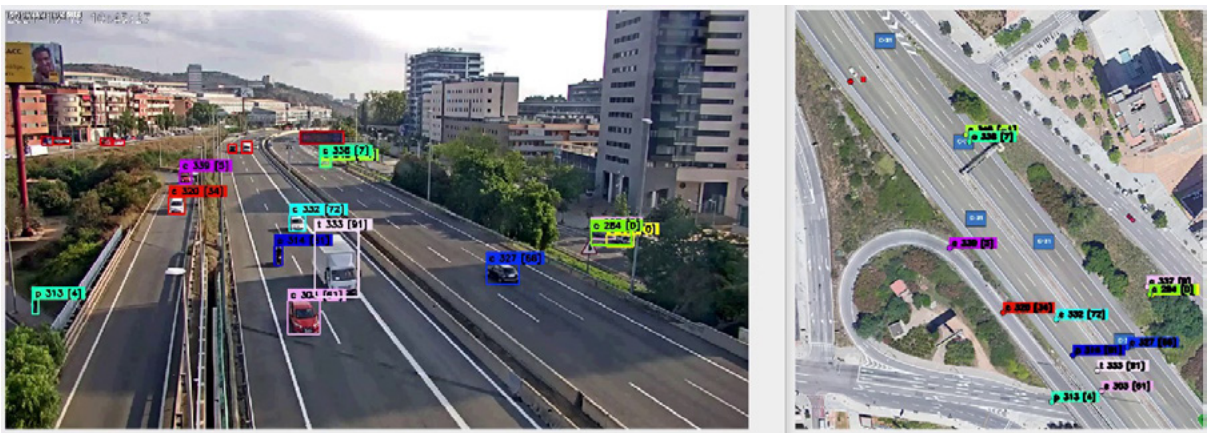


Figure: Detection, tracking, velocity and geoposition for each of the elements of the urban sequence. Left image: camera view, right image: Bird's eye view. Bounding box for each of the elements detected and tracked with their class, identifier, and velocity over each of the bounding boxes respectively.

Traffic light detection

For the urban scenario, the traffic lights status must also be obtained, in order to implement the urban risk case 2 regarding pedestrians crossing at crosswalk when the signal is red. In our case the traffic light sequence with a timestamp associated is not provided with the recordings, so computer vision and image processing must be employed to detect the traffic light color in the provided video sequences.

The algorithm for the traffic light status detection is based on detecting which of the colors of the traffic light is more intense and the temporal intensity difference between colors. The traffic light position is based on the semantics of the scene for easiness, but it could also be detected from the detector. Once the traffic light position is established, the position of the different lights is based on the changes in the Value channel of the HSV (Hue, Saturation, Value), because we want to detect which light is more intense between all the possible colors (in our example red, and green) and their position. Firstly, the position is established for all colors of interest, in our example red is in the higher position and green is in lower position. Secondly, we will compare the mean Value channel of both region positions computing the percentage difference between them to know which one is more intense.

Figure 6.3.TF_night is an example of the correct traffic light detection at night, where the position and the intensity can be seen more clearly for the traffic light evaluated. On the corner of each of the images is displayed a zoomed version of the traffic light and next to it the color detected.

Figure 6.3.TF_night: Example of correct traffic light detection at night, at the upper left corner there is a zoom of the traffic light evaluated and the corresponding color detected. Green and red color traffic light images detection, left and right images respectively.

The validation of this approach has been done by testing it all day long with all possible lighting conditions from direct sunlight to absence of light, and checking them with a manually created traffic light color sequence with timestamp associated, detecting less than hundred fails from the approximately million of frames evaluated, most of them caused by occlusion from buses. Some examples of this validation test are displayed on the figure 6.3.all_day.

Figure 6.3.all_day: Example of the traffic light detection correctly working along the day with the different illumination changes. At the upper left corner of each of the images there is a zoom of the traffic light evaluated and the corresponding color detected.

Semantic maps

In order to be able to detect the risk cases explained in the WP1, it was necessary to define a series of semantic information, such as vehicle lanes and their directions, bus stops, zebra crosses, pavements, cycle lanes, etc., which will be associated with the camera view image. This semantic information was used jointly with the event detection rules, and the information provided from the previous task of this WP to detect the risk events.

First, we defined a corpus with the semantic information needed for the type of scene and use cases evaluated, in our case interurban and urban sceneries. The corpus was defined with all the information needed independently of the camera view used.

So, two different corpus were defined, one for urban scenes and one for interurban scenes, since different areas of interest were relevant to each scenario. For example, urban scenes needed more information regarding crosswalks, traffic lights, etc., which were not relevant for the interurban scenarios. The two different corpus were applied to the three selected cameras (two for the interurban scenarios and one for the urban).

The data for the interurban corpus contained the region identifier in the image (id), the region class (road_lane, tracking_area, etc.), the traffic lane direction rule (if the car can change to the left or right lane), the type of user

can use the region, the orientation of the lane, and extra attributes if needed. For each of the defined fields there is only one possible selection from a group of already predefined possibilities. Id and attribute fields are the only ones that don't have a predefined choice, and they need a real number and free text respectively. The data for the urban corpus was similar but with more region class types such as crosswalk, intersection area, traffic light, etc., as well as an extra parent attribute associating the crosswalks with their corresponding traffic lights ids.

Secondly, once both corpus were decided, we manually annotated the different camera views available with the VGG Image Annotator tool [cite_tool]. This tool is commonly used for ground truth creation for machine learning training. We will be using this tool because it allows us to easily annotate the semantic regions using polygons and associate them easily with the corpus created previously. Automatic semantic annotation is possible but it will not give neither the precision nor the amount of information needed for the risk event detection, and every camera only needs to be annotated once.

In the next figure can be seen an example of the semantic annotation of the C31S using [cite_tool]. The annotated region polygons can be seen in yellow color and under the picture can be seen the semantic annotation for the first three regions (road lanes in this case).

Figure: Example of the semantic labeling for the C31S image employing [cite_tool] tool. Yellow polygons define the regions and the defined corpus is associated with each of these regions.

In the next figures the region tracking area for the C31S, C31N and urban sequences is displayed with blue and cyan colors.

Figure: Right figure : C31S semantic labeling where each of the yellow polygons corresponds to a labeled area. Left figure: Same picture but with some areas in different colors as example: lane 1 in green, and tracking areas in blue and cyan.

Figure: C31N and urban semantic labeling, left and right images respectively. C31N image with lane 1 area in green, and tracking areas for the scene in blue and cyan colors. Urban image with tracking area in blue and crosswalk 1 area in green.

Event detection rules

Once all the information about the semantics was defined, the rules for the event detection needed to be specified, corresponding to the five interurban and four urban detection events considered in this project. We used the information provided by the detection and tracking module to situate the objects spatially and temporally in the scene. Then, for each object, the algorithm first checked whether it lied within the tracking area defined in the semantic maps and, if yes, proceeded to compute its distance from each of the other detected objects. The risk events were evaluated for every frame of the analyzed databases. Temporal information was provided analyzing the difference between current and previous frames. In our experiments, a window of 500 ms was more than enough to correctly process the event changes, as it will also be discussed later in the quantitative evaluation. The rules are explained in more detail in continuation:

First interurban risk case, illegal lane change. We will consider a vehicle at risk when it is performing an illegal lane change. Note that the road lanes and the corresponding prohibited changes are defined in the semantic map of the scene. By tracking the position of the vehicle between consecutive frames, it can be determined if a lane change has occurred, and based on the lane orientation, whether this change is permitted or not.

Second interurban risk case, stopped vehicle on the roadway. In this case there are two possible situations when a vehicle is stopped on the roadway: traffic jam or car problems. For the risk analysis we are interested in the situation when there is no traffic jam involved. So, we check if the velocity of a vehicle is around 0 (not using absolute zero value to avoid noise) and the distance of the vehicle with other objects in front of it (with the direction determined by the orientation of the lane). If no objects exist, then it can be concluded that no traffic jam is present, and the stopped vehicle is considered a risk.

Third interurban risk case, vehicle going the wrong direction. In this case we estimated the trajectory direction of the vehicle by comparing its change of position from the previous to the current frame, and compared this information with the lane orientation. In case they diverge, the vehicle is assumed to be going in the wrong direction.

Fourth interurban risk case, vehicle on the shoulder. This is an easy case to detect because it only requires the evaluation of the position of the vehicle in order to know the road lane and the semantic info associated with this lane. In case that the vehicle is detected to be on a shoulder, this risk is detected.

Fifth interurban risk case, vehicles too close to other vehicles. This case was considered as a risk when the vehicle exceeded a minimum permitted distance between vehicles. The minimum vehicle distance is based on the official traffic minimum distance rule from the spanish “Dirección General de Tráfico” entity which depends on the vehicle speed [DGT_news]. In order to not be so restrictive and consider only the cases where it can be considered a risk, we used a 30% of the official minimum distance between vehicles and a minimum velocity of 50 km/h. The minimum distance was computed based on the vehicle speed on runtime. The risk will be considered true if the vehicles in front of the evaluated vehicle are a lower distance than the minimum distance. For easiness, we have only considered the distance to the car in front, and not the lateral distance.

First urban risk case, pedestrian crossing outside of the crosswalk. In this case we first need to correctly identify the type with which the person is detected, which can either be pedestrian, motorcycle rider, or cyclist. To improve this detection, a person was considered a rider if more than 50% of the detected person's bounding box is overlapping with the motorcycle or bicycle bounding box. Once a pedestrian is detected, based on their position and the crosswalk area determined in the semantic map, it can be determined whether this risk occurs.

Second urban risk case, pedestrian crossing at crosswalk when the signal is red. In this case we also need to distinguish between pedestrians and riders. We also need to determine whether the pedestrian is on the crosswalk, as well as the corresponding traffic light, using the position and the semantic maps. Once we have this information, if the corresponding traffic light color is detected as red, then this event is considered true.

Third urban risk case, a car is making a turn and approaching too closely a pedestrian that is crossing at a crosswalk of the street the car is turning into. This event is considered a risk when: i) the vehicle that enters an intersection area does not leave a minimum safe distance, which in our implementation is defined as less than 6m to a pedestrian crossing the crosswalk, and ii) the vehicle exhibits some velocity, higher than 3 km/h, thus filtering out avoid stopped cars in the area.

Fourth urban risk case, a car is making a turn and approaching too closely a bicycle that is crossing at a crosswalk of the street the car is turning into. The approach for this event detection is similar to the previous one, but focusing on bicycles instead of pedestrians.

6.4 WP4: Demonstration of the use cases

The last WP implemented the risk event and traffic flow analysis, based on the outcome of the detection and tracking module provided by WP3. Regarding the risk event analysis, the rules for the detection of the events, combining the type and position of the detected road users with the semantic mapping were first implemented, according to the definition of the suburban and urban risk situations, defined in WP1. The obtained results were validated through both a quantitative and qualitative evaluation, which also helped to improve the implemented methodology and derive several valuable technical lessons. With respect to the traffic flow analysis, the filtering rules applied to determine the congestion conditions and detection of vehicles with abnormal speeds were implemented. Finally, WP4 provided a dashboard to enable the visualization and quick analysis of the obtained project results, implementing dedicated panels to show the output of detection and tracking, speed estimation, risk event detection for urban and suburban scenarios and the flow analysis.

6.4.1 Risk event analysis

In this section the risk event detection rules, semantic corpus and all the modules defined in the previous WP3 will be tested to provide the risk events in the interurban and urban scenarios defined in the WP1.

Quantitative evaluation

In order to assess the validity of the implemented analytics, a quantitative evaluation has been conducted based on three reference 1-hour videos for the three monitored locations, i.e., the two interurban highways C31-S and C31-N, and the urban location in Barcelona. The selected reference videos were recorded on:

- C31-S: Friday 15/10/2021 from 12:00 to 13:00 CET
- C31-N: Friday 15/10/2021 from 15:00 to 16:00 CET8_00h a
- Urban: Thursday 30/12/2021 from 13:00 to 14:00 CET

The analysis consisted on establishing the ground truth for the considered risks, generating the confusion matrix and evaluating the precision, recall and F1 score metrics, defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

For the interurban scenarios, the quantitative analysis was limited on risk 1 (illegal lane change) and risk 4 (vehicle on shoulder lane). Risks 2 (vehicle stopped) and 3 (vehicle circulating in the wrong direction) were not materialized within the reference videos, whereas only two incidents of true positive detection of risk 3 has been detected for the total duration of the processed videos (for both highways), thus providing very few data to make a meaningful analysis. Finally, it has not been possible to objectively evaluate the detection risk 5 (safety distance) due to the lack of ground truth data.

The results are summarized below. Table 1 and Table 2 are the confusion matrix and qualitative analysis for C31-S, respectively, whereas the corresponding results for C31-N are given in Table 3 and Table 4. High performance has been obtained for the suburban scenario and especially for the C31-S where there was observed a high number of occurrences for risks 1 and 4. On the other hand, the scarcity of events in C31-N led to results with lower statistical significance.

Table 1. Confusion matrix for risks 1 and 4, for the C31-S highway

C31-S	Risk 1 (illegal lane change)	Risk 4 (vehicle on shoulder)
True Positives (event occured, detection)	74	14
False Negatives (event occured, no detection)	9	0
False Positives (no event, detection)	5	6

Table 2. Quantitative analysis results for risks 1 and 4, for the C31-S highway

C31-S	Risk 1 (illegal lane change)	Risk 4 (vehicle on shoulder)
Precision	89.16"%	70.00"%
Recall	93.67"%	100.00"%
F1-score	91.36"%	82.35"%

Table 3. Confusion matrix for risks 1 and 4, for the C31-N highway

C31-N	Risk 1 (illegal lane change)	Risk 4 (vehicle on shoulder)
True Positives (event occured, detection)	3	1
False Negatives (event occured, no detection)	0	0
False Positives (no event, detection)	0	0

Table 4. Quantitative analysis results for risks 1 and 4, for the C31-N highway

C31-N	Risk 1 (illegal lane change)	Risk 4 (vehicle on shoulder)
Precision	100.00%	100.00%
Recall	100.00"%	100.00"%
F1-score	100.00"%	100.00"%

Finally, Table 5 and Table 6 summarize the analysis of the three urban risks, namely the pedestrians crossing outside the designated areas (risk 1) or with red (risk 2), and cars making a turn and approaching pedestrians (risk 3). Risk 4 regarding cars turning towards bicycles did not materialize in the considered scenario, so it was not included in the evaluation. The performance in the urban scenario is again satisfactory, although a higher num-

ber of false positives was observed, compared to the suburban scenario, leading to a lower precision score. The next section will discuss in more detail the reason behind these false event detections, mainly caused by errors of the detection and tracker module.

Table 5. Confusion matrix for risks 1, 2 and 3, for the Urban scenario

Urban Scenario	Risk 1 (pedestrian crossing outside crosswalk)	Risk 2 (pedestrian crossing with red)	Risk 3 (vehicle turning towards pedestrian)
True Positives (event occured, detection)	48	33	15
False Negatives (event occured, no detection)	0	0	1
False Positives (no event, detection)	21	22	8

Table 6. Quantitative analysis results for risks 1, 2 and 3, for the Urban scenario

C31-S	Risk 1	Risk 2	Risk 3
Precision	69.57"%	75.00"%	65.22"%
Recall	100.00"%	100.00"%	93.75"%
F1-score	82.05"%	85.71"%	76.92"%

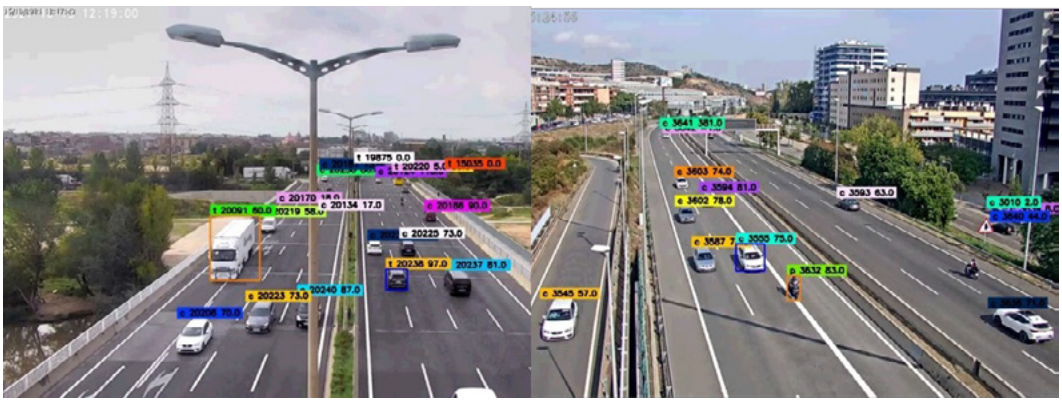
Qualitative evaluation

This section will provide a qualitative evaluation of the obtained results on the risk event detections for both interurban and urban scenarios. The analysis will focus on showing examples of how the developed analytics detect the different risk categories, as well as discussing the identified sources of errors.

Interurban risk 1: Illegal Lane Change

As mentioned before, risk 1 refers to the detection of vehicles crossing solid lines, which can be the cause of multiple traffic accidents. Figure 1 shows an example of a true positive (TP) detection of vehicles crossing the solid line in both C31-S (left) and C31-N (right), indicated by orange bounding boxes.

Figure 1. Example of correct interurban risk 1 (orange bounding box) detection in C31-S (left) and C31-N (right). Left image : motorbike crossing continuous line (risk1 - orange bounding box) and security distance problem (risk4 - blue bounding box).



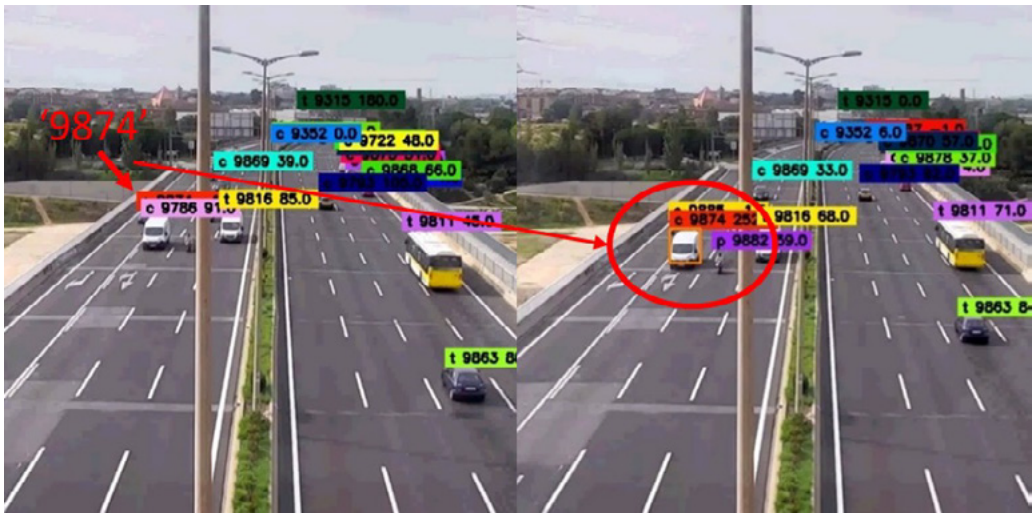
In some cases, the vehicles did not complete the "change of lane" maneuver, but were circulating on the solid line, as in the examples shown in Figure 2. These cases were also counted as TPs.

Figure 2. Example of correct TP interurban risk 1 detections – vehicles on the solid line

As shown in the quantitative analysis, there were some cases of False Positive (FP) detections, reporting the occurrence of the risk when no real event took place. These false detections were mainly due to three different types of errors, all coming from the detection and tracker module, which are further discussed in continuation.

The first type of error is caused by the switching of the tracker id, usually occurring when a vehicle is partially hidden, either by another vehicle in its vicinity or other object on the scene (e.g., lamp post, trees, etc.). As a result, the tracker identified a previously tracked vehicle as a new object and assigns a different id across consecutive frames. In the example depicted in Figure 3 (left), a white van was detected as object "9786", whereas there was another partially hidden car circulating in the adjacent lane, detected as object "9874" (red label). In the next frame, the tracker mistakenly assigned ID "9874" to the front vehicle, as shown in Figure 3 (right). Due to this error, the vehicle with id "9874" was considered to have crossed the solid line, thus reported as a Risk 1 event detection.

Figure 3. Example of false positive detection for interurban risk 1 – objects switching tracker ID



The second type of error is the detection of "ghost objects", produced when the detection module detects nonexistent objects. In this case, the detection and tracking module reported objects that were not really present in the frame, sometimes also resulting in false event detections, as in the example shown in Figure 4.

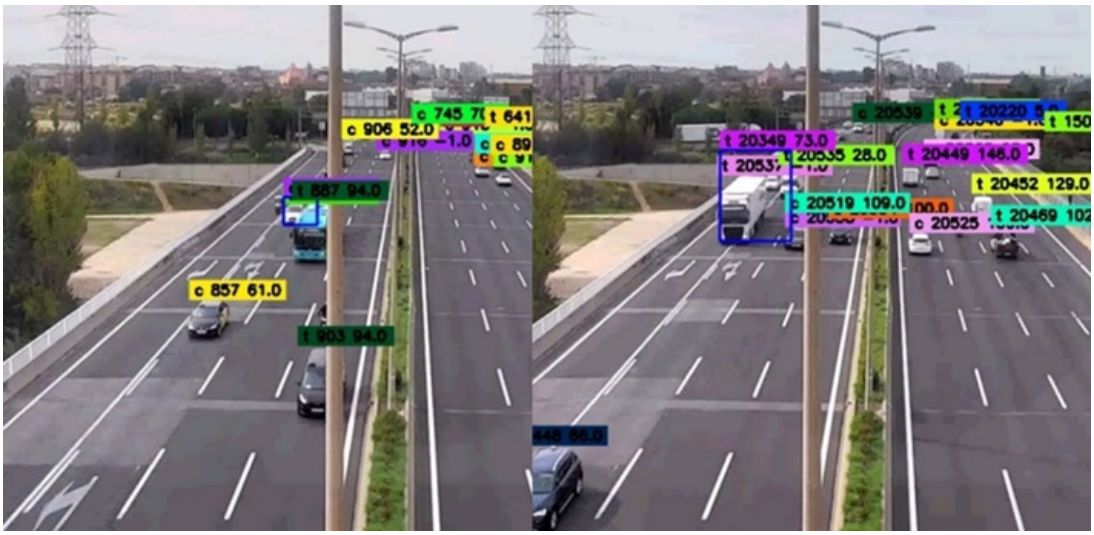
Figure 4. Example of false positive detection for interurban risk 1 – detection of "ghost objects"

The third cause of FPs was the imprecise detection of very large vehicles, often having complex structure, as the ones depicted in Figure 5. In these examples, the size and position of the bounding boxes, combined with the perspective of the camera, resulted in the false detection of the event.

Figure 5. Example of false positive detection for interurban risk 1 – large vehicles

Finally, in some cases the analytics failed to detect a real occurrence of the event, i.e., a vehicle crossing the solid line, resulting in a false negative (FN). The majority of FNs resulted from vehicles that were not properly detected by the detection and tracking module. This included cases where vehicles were not detected at all, which mostly happened with motorcycles circulating very close to other vehicles, as in the example shown in Figure 6 (left). Another example, shown in Figure 6 (right), refers to big trucks that were not properly detected (e.g., in the example, the bounding box incorrectly includes partially hidden vehicles behind the truck). Since in the considered approach, the vehicle's position is determined by the middle point of the lower part of the bounding box, in the depicted example, the truck is considered to circulate within the second lane from the left (and not crossing the solid line).

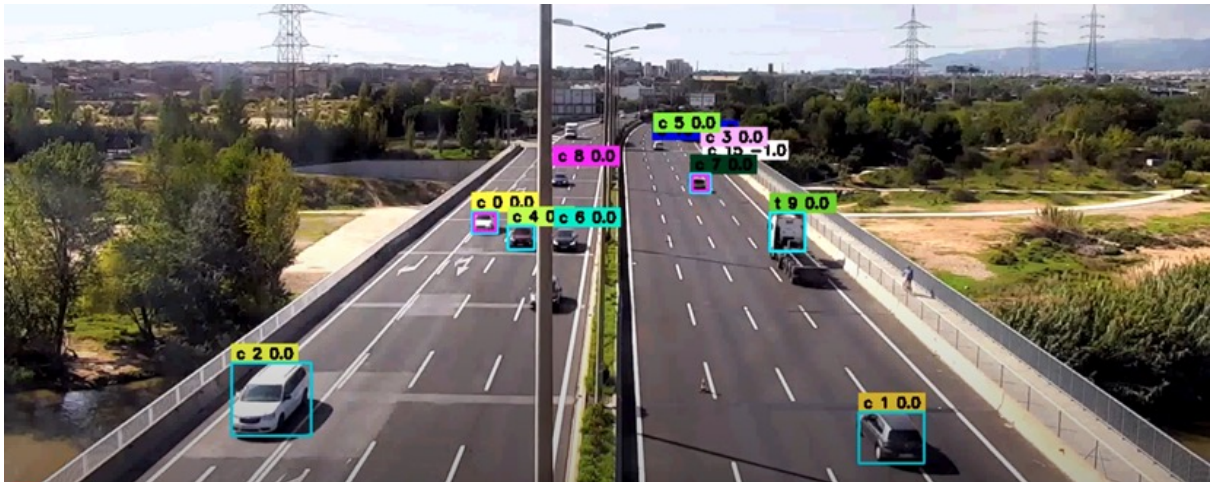
Figure 6. Example of false negative detection for interurban risk 1 – detection errors



Interurban risk 2 – Stopped vehicle on the roadway

The second interurban risk refers to the detection of vehicles stopped on the highway, with no apparent reason (e.g., the presence of a traffic jam). During the three days of processing, no recorded incident for this risk took place, resulting in zero True Positive or False Negative detections. However, six False Positive detections were observed within one video frame, depicted in the example of Figure 8 (with events marked by light blue bounding boxes). This caused by the freezing of the recording video for around 1 min, causing the tracker to return a zero speed estimation for all vehicles and consider them correctly as stopped vehicles in the highway (risk 2)

Figure 8. Example of false positive detection for interurban risk 2 – error due to malfunction of the camera. The event was correctly detected because the camera was frozen for around 1 min but it was considered FP detection because was not a real event



Interurban risk 3 – Vehicle circulating in the wrong direction

Interurban risk 3 refers to the detection of vehicles (and pedestrians) circulating in the wrong direction. During the three days of processing, there were only two recorded incidents of this risk, depicted in Figure 9, showing a maintenance vehicle circulating on a closed highway lane (left) and a pedestrian walking in the opposite direc-

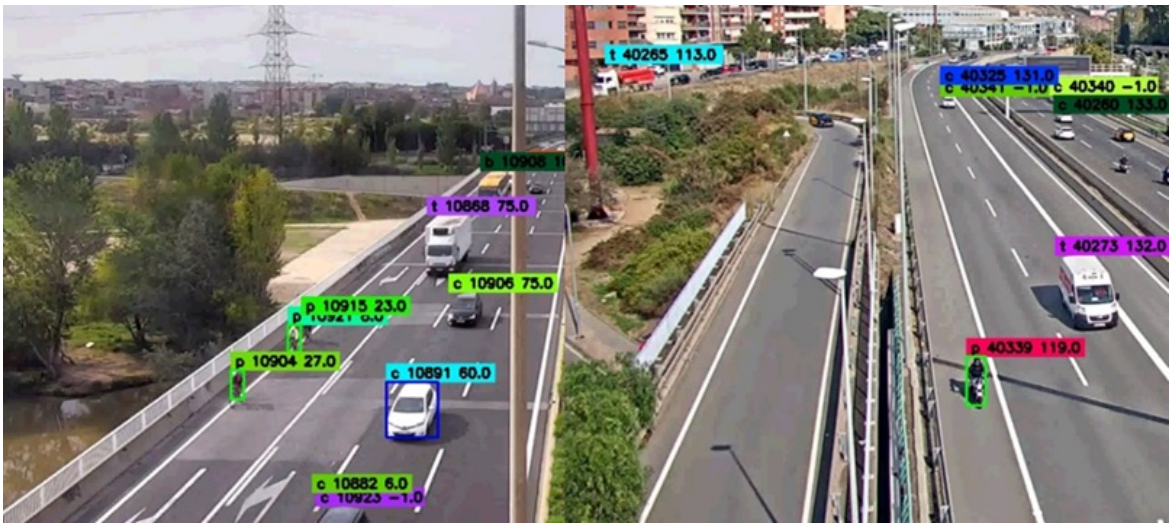
tion on the shoulder lane (event marked by a pink bounding box). There have also been observed two cases of false positive detection due to the problem of "ghost object" detection.

Figure 9. Example of interurban risk 3 – road users moving on the wrong direction

Interurban risk 4 – Vehicle circulating on the shoulder

Interurban risk 4 refers to the detection of vehicles (and pedestrians) circulating on the shoulder lanes. Two examples of correct detection are shown in Figure 10, depicting three cyclists circulating on the safety lane in C31-S (left) and a motorcycle in C31-N (right), with the detected events marked by green bounding boxes.

Figure 10. Example of interurban risk 4 in C31-S (left) and C31-N (right), events marked by green bounding boxes



There have been a few cases of false positive detections regarding the presence of vehicles in the shoulder lanes, with two representative examples shown in Figure 11. In the left part of the figure, the truck was not correctly detected, generating a bounding box that lies within the shoulder lane, triggering a risk detection. In the right part of the figure, a car circulating very close to the solid line resulted in a bounding box within the shoulder lane, due to the perspective of the camera. This latter error was corrected by adjusting the semantic annotation of the shoulder lane to account for the image perspective.

Figure 11. Example of false positive detection for interurban risk 4 due to wrong detection (left) and image perspective (right). The latter error was corrected by adapting the semantic annotation, thus improving the precision of the analytics.

Interurban risk 5 – Safety distance not respected

Interurban risk 5 refers to the detection of incidents in which the safety distance of two vehicles has not been respected. Establishing the ground truth for this risk was not feasible, since it would require knowledge of the real speed and distance between the vehicles. Given this limitation, the qualitative analysis only focused on identifying examples of the correct detection of this risk, as shown in Figure 12, depicting cars circulating very close together while circulating with relatively high speeds (marked by blue bounding boxes).

Figure 12. Example of interurban risk 5 - vehicles not respecting the safety distance

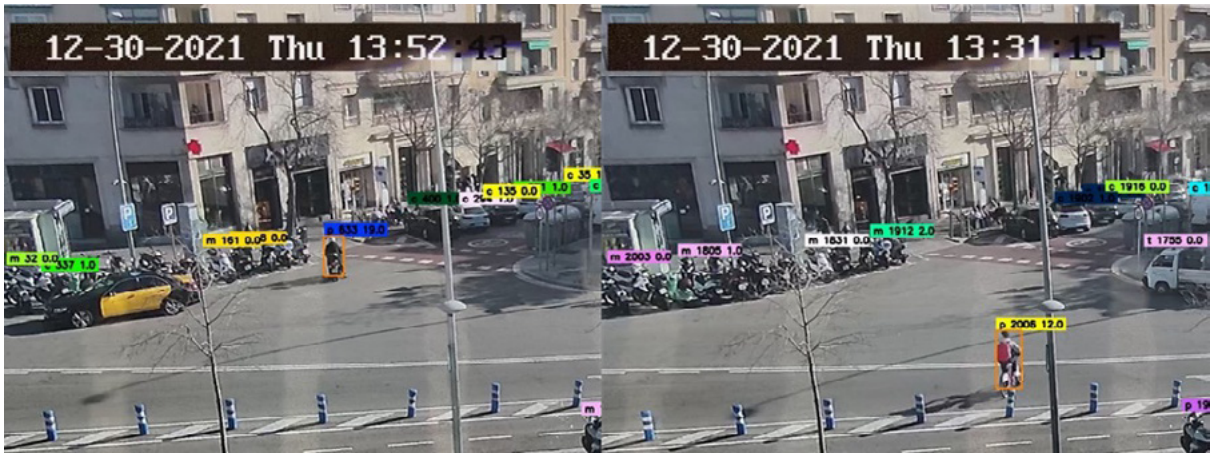
Urban risk 1: Pedestrian crossing outside the crosswalk

The first urban risk refers to the detection of pedestrians crossing the street outside the designated crossing area. Figure 13 shows two examples of true positive detections, marked by orange bounding boxes.

Figure 13. Example of urban risk 1 detection - pedestrians crossing outside the designated crossings, marked with orange bbox

In some cases, the riders of motos or bicycles were wrongly detected as pedestrians, leading to false positive detections. In order to distinguish between real pedestrians and riders, the implemented analytics correlated the presence of people with detected vehicles. However, in some cases, this approach failed due to the incorrect detection of motorcycles or bicycles, as in the examples of Figure 14.

Figure 14. Example of urban risk 1 false positive detection – motorcycle and bicycle riders detected as pedestrians because of the incorrect detection of the motorcycle



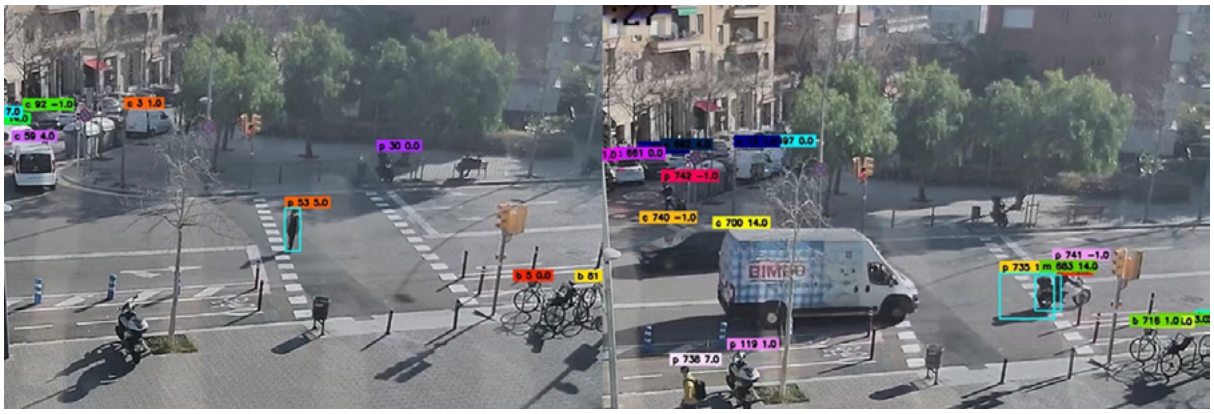
Another special case was the detection of scooter riders as pedestrians, as shown in Figure 15. This type of error was expected, since the detection module was not trained to detect scooters. Such problems could be avoided in future deployments by extending the training set to include more vehicle types and improve the detection of smaller vehicles.

Figure 15. Example of urban risk 1 false positive detection – scooters detected as pedestrians

Urban risk 2: Pedestrian crossing with red

The second urban risk refers to the detection of pedestrians crossing within the designated area when the traffic light is red. For the detection of this event, the detected status of the traffic light was correlated by the detection of pedestrians on the crosswalk. An example is shown in Figure 16 (left), with the detected event correctly marked by a light blue bounding box. There were some false positive detections due to the incorrect detection of shadows as a person, as shown in the right part of the figure. In this case, the long shadow of the objects is detected as a person, thus showing the impact of the lighting conditions on the detection precision. It should also be noted that in all the analyzed videos, the traffic light detection was performed correctly.

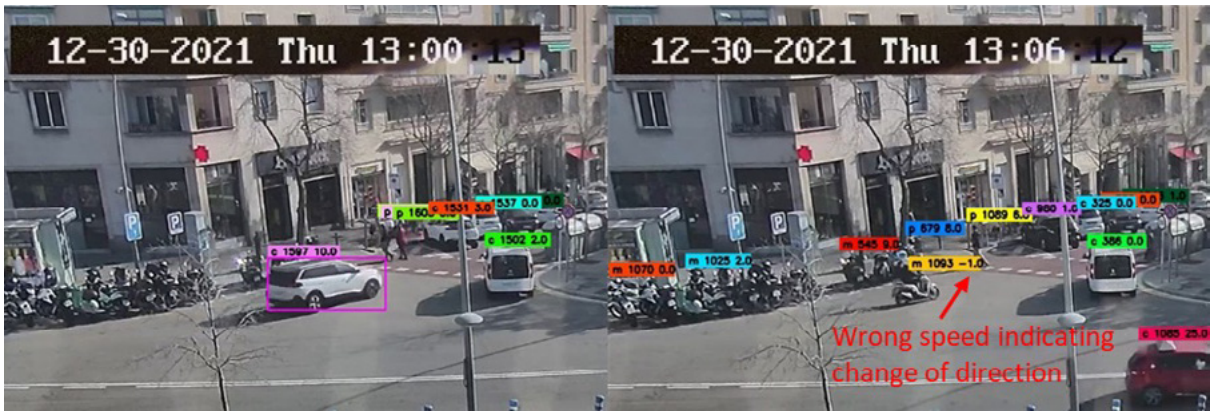
Figure 16. Example of urban risk 2 detection – pedestrians crossing with red, showing a true positive detection (left) and a false positive detection due to the incorrect detection of a shadow as person (right)



Urban risk 3: Vehicle turning while pedestrian is crossing

The third urban risk refers to the detection of vehicles turning from the main street and approaching the pedestrian crosswalk at relatively high speed when people are crossing. Figure 17 shows an example of a true positive detection (left), as well as an example of a false negative detection (right), marked by pink bounding boxes. The false negative was caused by the inconsistent detection of the motorcycle, resulting in a switching of the object's id. This, in turn, prevented the correct detection of the vehicle's speed, thus failing to detect the occurrence of the risk event.

Figure 17. Example of urban risk 3 – true positive (left) and false negative (right) detection



6.4.2 Traffic flow analysis

This section will provide the specification of flow analysis events. Firstly, it details how congestion has been computed. Furthermore, the speed of the detected vehicles will be analyzed in three scenarios: vehicles with abnormally high speeds, vehicles with abnormally low speed and vehicles accessing the speed limit.

Congestion

Congestion is a condition in transport that is characterized by longer trips times, slower speeds and increased vehicular queuing. Our purpose is to detect potential congestions. For this reason, the following values are calculated:

- average detections per hour (d)
- standard deviation of detections per hour (std)
- threshold value, which was defined as “two” in this implementation (threshold)
- number of detections for each hour (n)

Then, the following rule is applied:

$$n >= \underline{d} + std \; x \; threshold$$

If the number of detections per hour (n) is greater than the sum of the average number of detections (d) and two times (threshold) the standard deviation (std), it is marked as congestion. Otherwise it is considered that no congestion even has taken place.

Speed analysis

The speed is a very strong indicator for the flow analysis, since it can reveal a lot of information about what is happening on the road. Furthermore, it can also help to detect congestion, when vehicles are moving at very low speeds. Our purpose is to analyze the speed of vehicles on the road. For this reason, the following values are calculated:

- average speed (s)

- standard deviation of speed (std)
- threshold value, which was defined as “two” in this implementation (threshold)
- speed of the detected object (v)
- road speed limit (lim)

With the aim to detect vehicles with abnormally low speeds, the following rule is applied:

$$v <= \underline{s} - std \ x \ threshold$$

if the speed of vehicle (v) is lower than the average speed (s) minus two times (threshold) the standard deviation (std), it is marked as a vehicle with abnormally low speed.

Moreover, in order to detect vehicles with abnormally high speeds, the following rule is applied:

$$v >= \underline{s} + std \ x \ threshold$$

if the speed of vehicle (v) is greater than the sum of average speed (s) and two times (threshold) the standard deviation (std), it is marked as a vehicle with abnormally high speed.

Finally, in order to detect vehicles crossing the speed limit, the following rule is applied

$$v > lim$$

if the speed of vehicle (v) is higher than the speed limit, it is marked as accessing the speed limit. Otherwise, it is not.

6.4.3 Visualization of risk and flow analysis

This section will provide an overview of the implemented Kibana dashboards used for the visualization of the obtained project results. As previously explained, five different panels were created, to show the output of the detection and tracking module, the congestion analysis, the speed variations and the urban and suburban risk events. The five panels will be discussed in more detail next.

PAI - Detections:

The first dashboard, Figure 1, provides a holistic view of the outcome of the detection and tracking module, for both urban and interurban scenarios obtained by the three selected cameras.

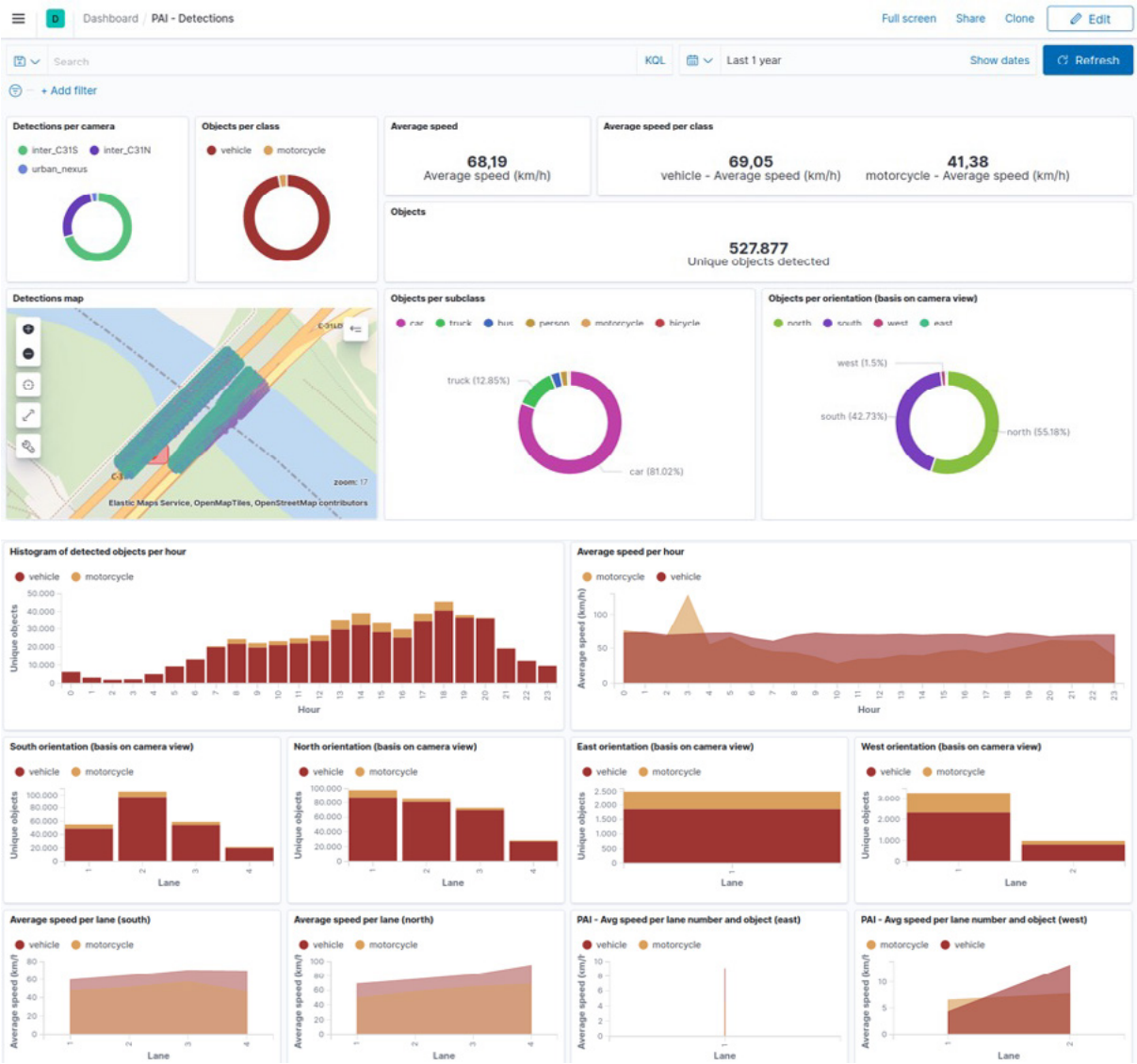


Figure 1: PAI - Detections dashboard panel

In the first row you can see the number of detections for each camera and the number of detections for each vehicle class, which will be further explained in Section 6.4.4. In addition, there are metrics about average speed.

In the second row, there is a map with camera location and all detections, enabling us to verify the output of the geolocalization module (in this example corresponding to detection on the C31-S segment). Also, the number of detections per subclass and per orientation is reported.

The third row includes two visualizations. On the left side there is an histogram of the detections per hour, whereas the average speed per hour is shown on the right, with different colors representing the different vehicle classes.

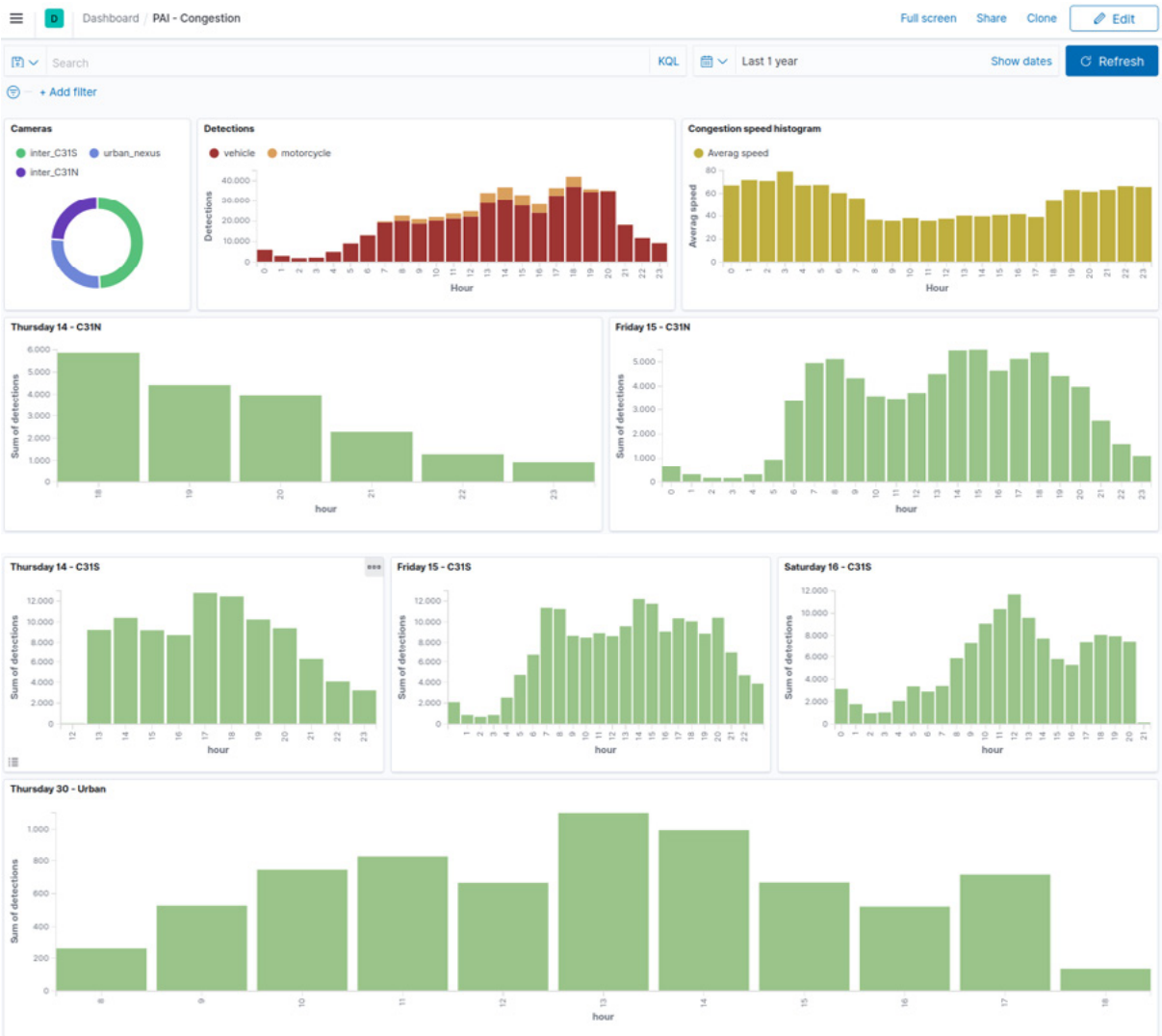
The fourth row contains an histogram of detections for each orientation, also detailing the lane number (assigned in the semantic mapping). It should be noted that the orientation is computed based on camera view, with permitted values defined as north, south, east and west.

Finally, the last row shows the visualizations on the average speed per lane.

PAI - Congestions:

This dashboard, Figure 2, contains all the information about the congestion on the different road segments.

Figure 2: PAI - Congestion dashboard



In the first row, we can see a pie chart with the ratio of detections for each camera. Moreover there are two histograms: detections per hour and average speed per hour.

Furthermore, the second row contains the histogram of detections per hour, aggregated per day for each camera. Specifically, the visualized data includes histograms of the following dates:

14th and 15th of October 2021, for the interurban camera C31N (second row)

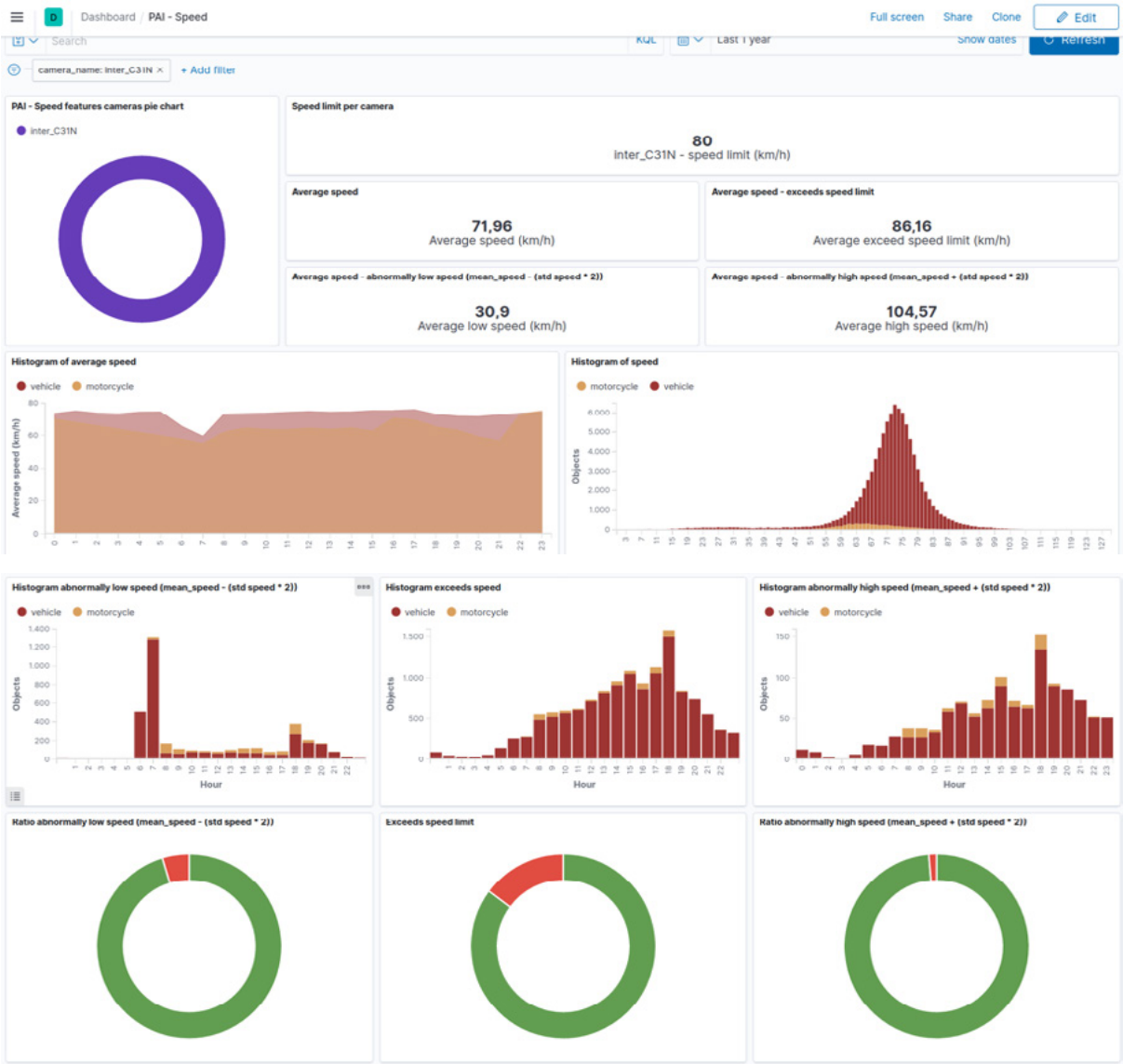
14th, 15th and 16th of October 2021 for the interurban camera C31S (third row)

30th of December 2021 for the urban camera (fourth row).

PAI - Speed:

The aim of this dashboard, Figure 3, is to analyze the estimated speeds of the detected vehicles on the different roads, based on the filtering rules described in the previous section.

Figure 3: PAI - Speed dashboard



In the first row there is a ratio of detections for each camera. Furthermore, we can see the speed limit for each camera, the average speed, the average speed of vehicles passing the speed limit and the average speed of vehicles going at very low/high speed (defined in the previous section).

In the second row there is a visualization of the average speed per class and the histogram of speed values for each hour.

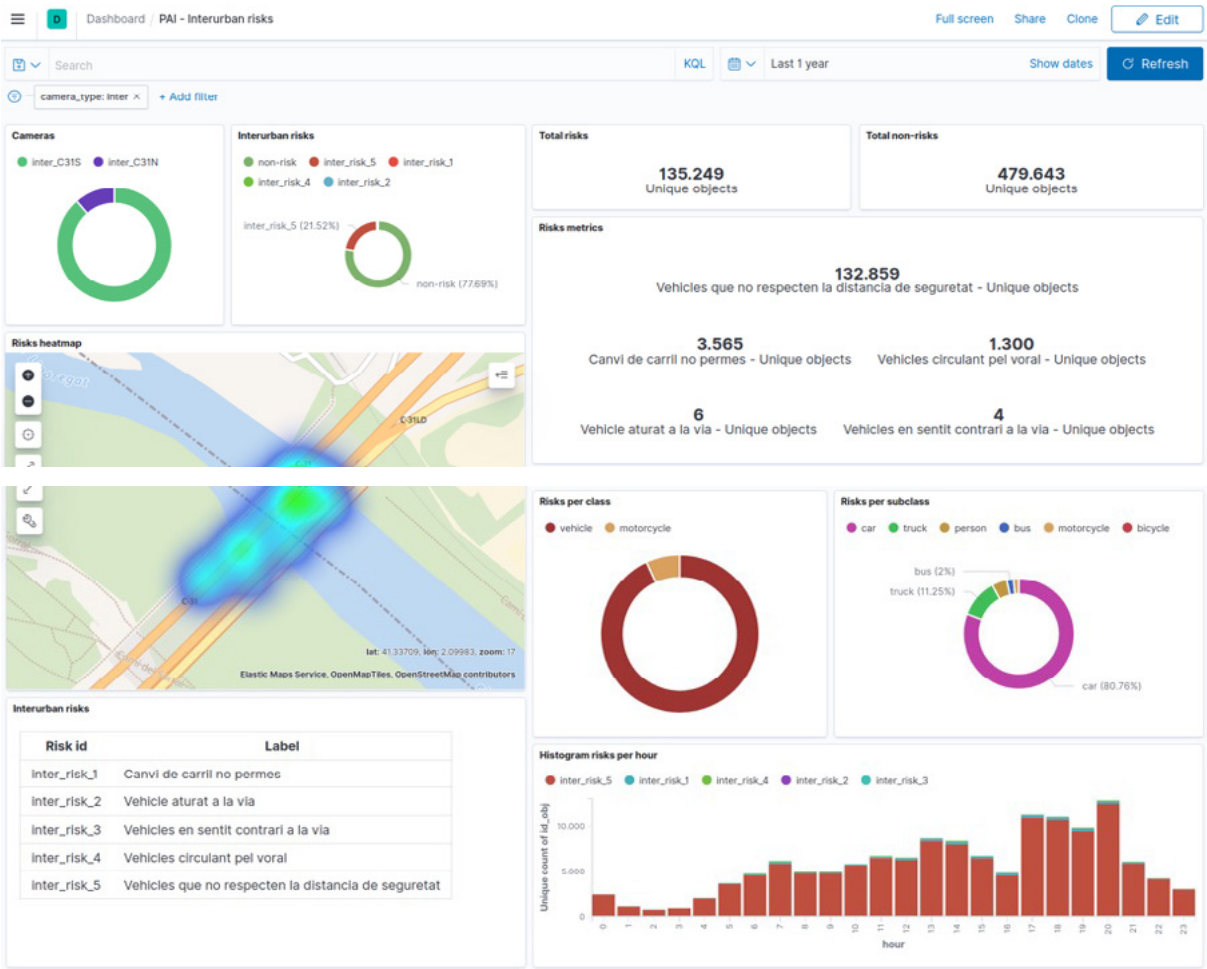
Then, in the third row there is an hourly histogram of detections at very low, very high speed and over the speed limit of the road.

Finally, there are three pie charts relating to the ratio of vehicles going at very low, very high speed and over the speed limit of the road.

PAI - Interurban risks:

The goal of this dashboard, Figure 4, is to monitor potential interurban risks.

Figure 4: PAI - Interurban risks



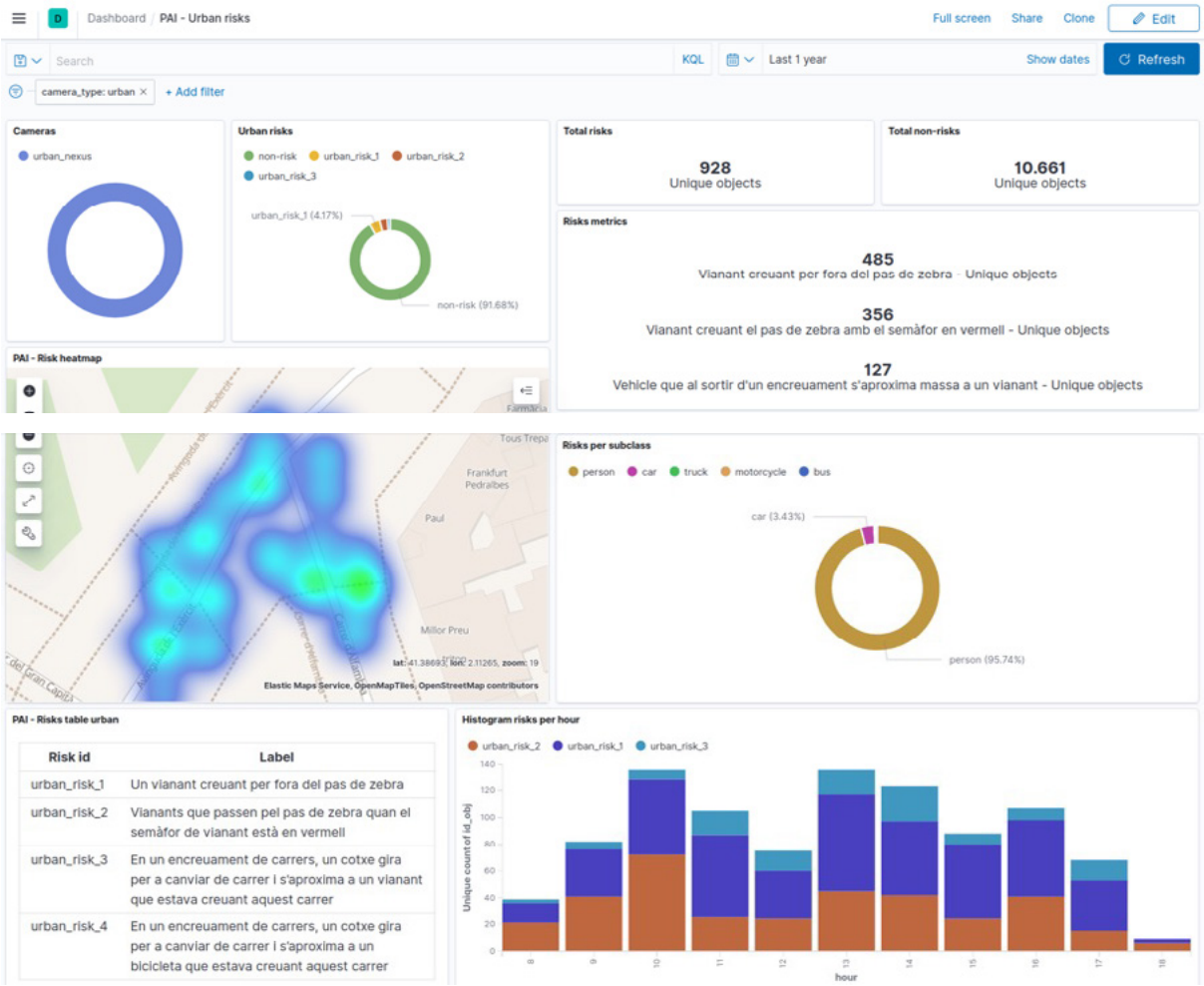
In the left-hand side of the dashboard you can see two pie charts. The first one is about the ratio of detection for each interurban camera (C31S and C31N), while the second is the ratio of detections for each interurban risk. Then, there is a heatmap for the occurence of the risks. Finally, we can see a table with the interurban risk id and its definition.

On the other hand, on the right-hand side the number of unique event detections for each interurban risk is shown. Moreover, there are two pie charts about interurban risks per class and subclass. Finally, there is an histogram of interurban risks per hour.

PAI - Urban risks:

The aim of this dashboard, Figure 5, is to monitor potential urban risks.

Figure 5: PAI - urban risks



As in the previous case, two pie charts are shown on the left-hand side, containing the ratio of detections for the urban camera and for each urban risk, respectively. A heatmap of risks and the table with their definition is also depicted.

On the right-hand side the number of detections for each urban risk is shown. In addition, a pie chart showing urban risks per subclass and an histogram of urban risks per hour is plotted.

6.4.4 Data analysis

This section will provide insights on the traffic detections, traffic flow and risks in urban and interurban environments, facilitated by the visualization of the project outcomes on the previously presented dashboards. In order to perform the data analysis, two pre-processing steps were carried out.

The first pre-processing step consisted in a simple data cleaning. Detections with an invalid object identifier or speed were deleted. Specifically, only detected vehicles with a speed below 300 km/h were kept. Moreover, any detections belonging to the airplane, train and boat classes were removed. In addition, the date and time was computed for each of the video detections.

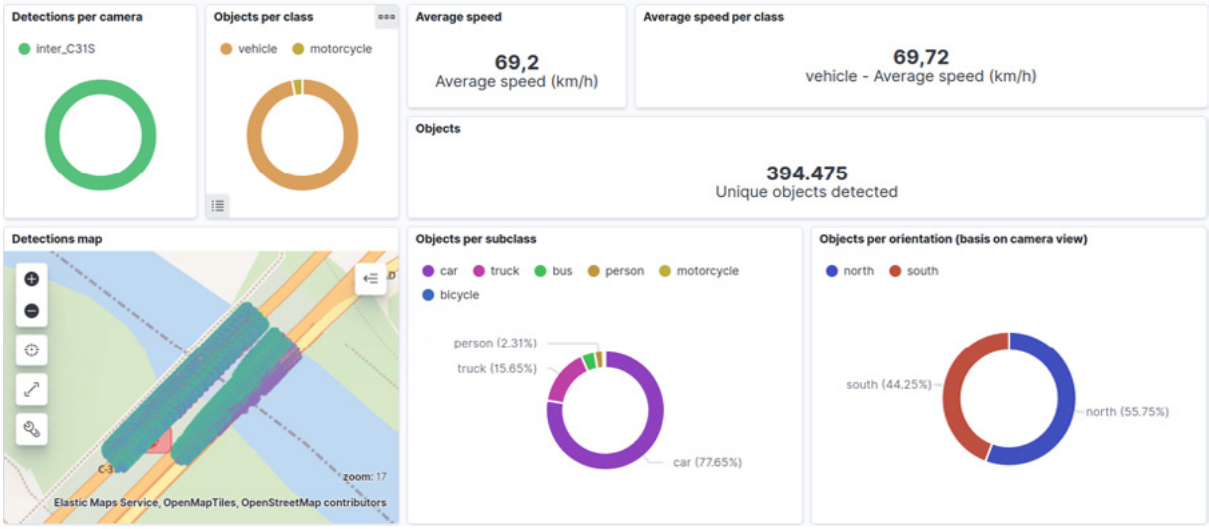
The second pre-processing step consisted in adding the semantic features. For each detection, it was computed whether the road user lied within the region of interest, otherwise the object was removed. Furthermore, in order to mitigate the problem of false detections, the maximum size of the bounding box was computed as a threshold. The maximum size of the bounding box was computed based on the distance to the camera. Detections

with greater bounding boxes than the threshold were removed. Finally, each detection was assigned to a lane, hard shoulder or bicycle lane with the corresponding orientation (based on the camera view).

The dashboards mentioned in the previous section have been used for the data analysis step. In continuation, some representative examples for each dashboard panel will be discussed, showing how the visualized data can lead to valuable insights.

PAI - Detections dashboard

Figure 6: Visualizations of PAI - Detection dashboard about objects classes and subclasses, orientations and detections map.



The image above shows the average speed of all detected objects from camera C31S. Also, the average speed is splitted by object class: vehicle and motorcycle. The mean speed of vehicles class is greater than motorcycles class.

The "objects per subclass" pie chart shows the ratio between whole subclasses. The most common subclass is "car". In addition, the objects class has the following division:

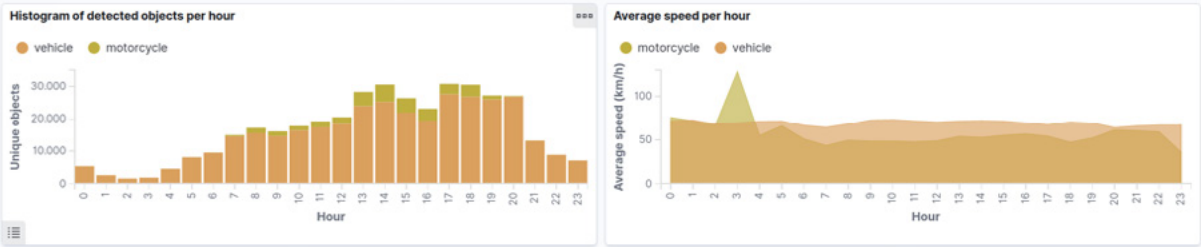
class name	subclass name
vehicle	car
	truck
	bus
motorcycle	person
	motorcycle
	bicycle

It is worthwhile to mention that in the interurban roads (C31-S and C31-N) it is forbidden that there are pedestrians and bicycles. Furthermore, when a motorbike is detected far away it is sometimes detected only as a person and not a motorbike with a person, it is because the model is trained to detect small people but not trained for small motorbikes. It could be improved with a better detector trained specifically for motorbikes. For these reasons, in these interurban roads the subclasses of person and bike are incorporated to the motorcycle class because they would not exist (although a small number of cases were detected).

The left side of Figure 7 below shows the histogram of detection at each hour splitted by class. The first peak is observed at 14h, and the other peak from 17h to 19h.

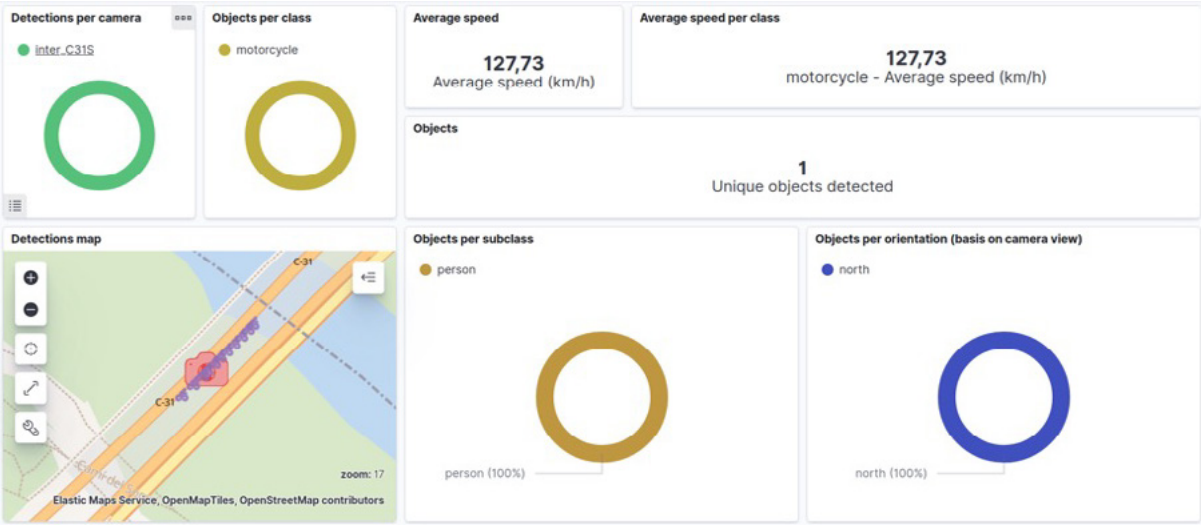
The right side of Figure 7 shows the average speed separated by object class. At 3 o'clock there is a peak for the motorcycle class. Likewise, as mentioned before, it is important to consider that the subclasses of person and bicycle are not included in this average speed.

Figure 7: On the left side histogram of detections per class. On the right side histogram of average speed per class



Going more in depth, the dashboard has been filtered to show only data for the motorcycle class obtained from the C31S camera at 3 o'clock. As shown in the next figure 8, the same motorcycle was detected 11 times, reported as a unique object detection, since the same id was assigned by the tracker module. This motorcycle circulates at high speed, above 120 km/h.

Figure 8: PAI - Detections dashboard filtered by camera C31S, hour at 3 and class equal to motorcycle.



The figure 9 below shows the histogram of detections and average speed per lane for each camera orientation, defined based on the camera's perspective. In addition, the number of lanes depends on their position on the image, with ascending lane numbers assigned from the left part of the image to the right.

With respect to the south orientation, four lanes are captured, with lane 1 being a slow lane and a road exit. By observing the histogram, it can be seen that lane 1 has a much lower speed than lanes 2 and 3, as well as a lower number of detected vehicles (who tend to circulate in the other lanes). It should also be noted that the average speed on lane 4 was not accurately estimated, due to the partial occlusion of the lane by a street light post.

On the other hand, in the north orientation, the whole lanes are accurately tracked. In this case we can see that the average speed increases from lane 1 to lane 4. We can observe In both orientations as the vehicle class has higher speed when using a left lane.

Figure 9: Visualizations of PAI - Detections dashboard about lanes and orientation.

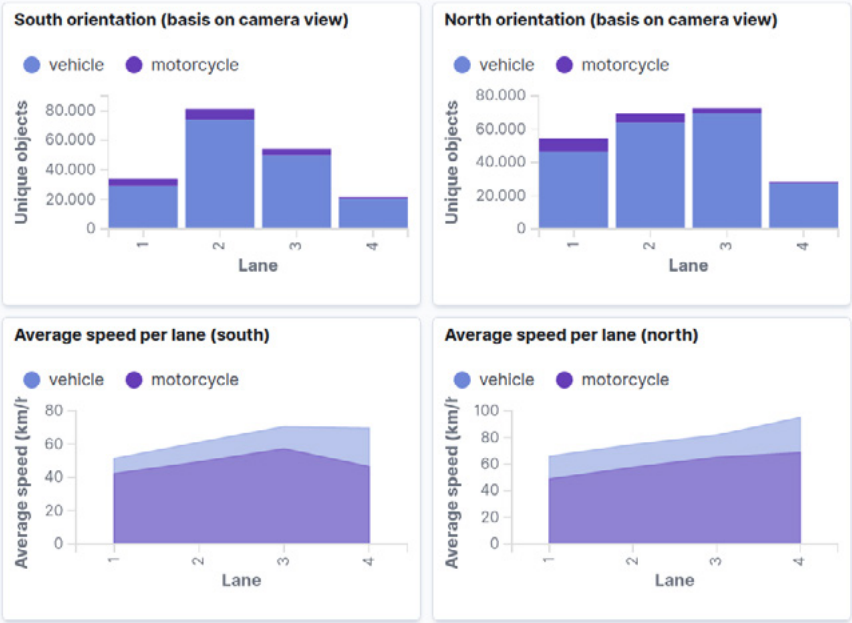


Figure 10: C31S camera view. The C31S road has four lanes for each direction. In addition, on the left side of the picture, you can see how the left lane does not allow lane changes.

To conclude the PAI - Detections analysis, the dashboard has been filtered to show only detected persons (as a subclass) in the urban scenario (filtered by the urban camera as a source).

Figure 11: Visualizations of PAI - Detections filtered by urban camera and person subclass.

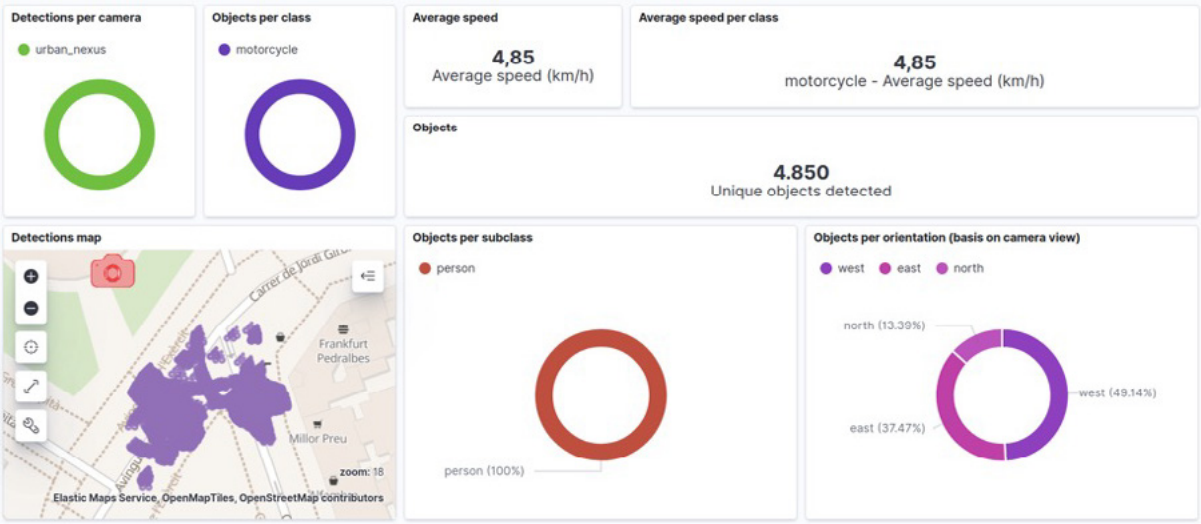


Figure 12: Urban camera view. The urban camera has one zebra crossing on the right side and the other one at the top of the image.

The figure above shows two zebra crossings, with a lot of detections inside. Furthermore, there are two clear lines of detections outside the zebra crossings. This means that some people cross outside the zebra crossings because the path is shorter.

PAI - Congestion dashboard

The figure 13 below shows the histogram of detections and average speed per hour for the C31N camera for the two days analized. It should be noted that the recorded samples from the first day (Thursday 14) are limited between 18h to 23h. This results in a peak in the accumulated detections at 18h, as shown in the histogram.

By observing the "Congestion speed histogram", there is a fall in average speed at 7h, suggesting a potential congestion problem. Focusing on detections histogram of Friday 15th, the number of detections at 7h is high, although not the highest with respect to the other hours. To better interpret this observation, a visual inspection of the video revealed that there was some light congestion for a small amount of time, from 7:02 to 7:06, as can be seen in the figure 15. This has led to the conclusion that more fine-grained filtering rules (e.g., aggregating detected vehicles in 10-minutes intervals instead of 1 hour), could provide more useful insights of the congestion conditions.

Figure 13: Visualizations of PAI - Congestion filtered by C31N camera.

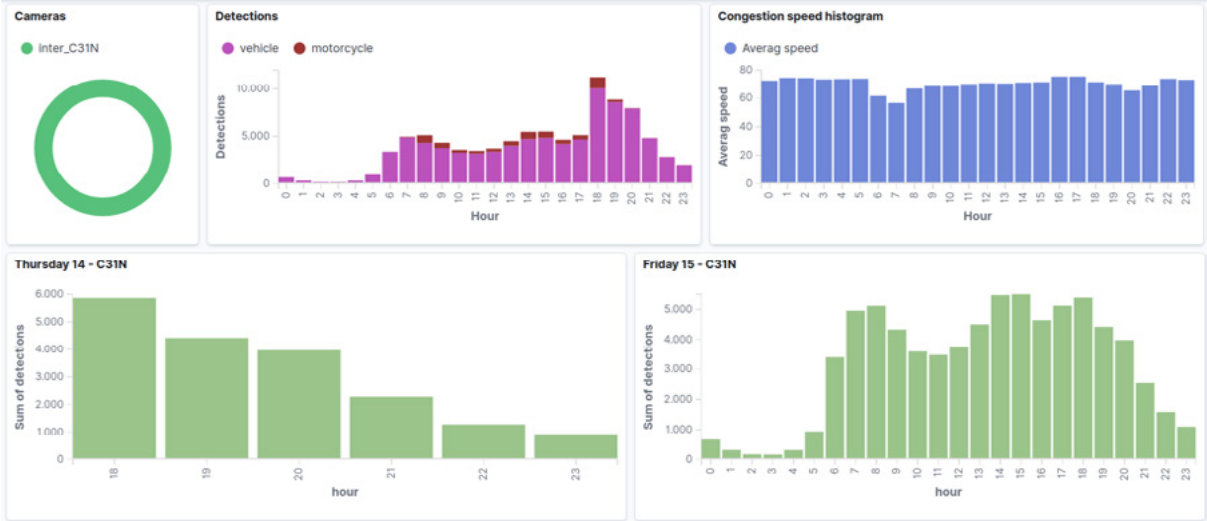
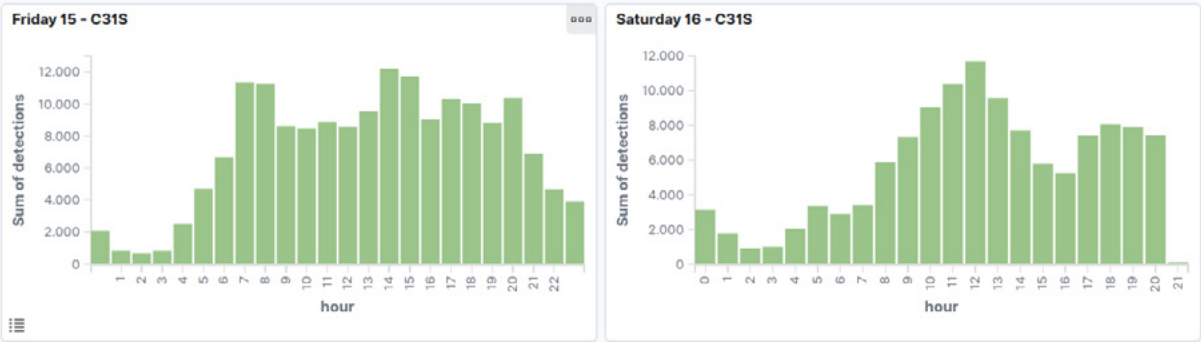


Figure 15: Image of a congestion on Friday 15th at 7h for C31N camera.

The figure 14 below shows the histogram detection for each hour for the C31S camera for two of the three days analyzed (Friday 15 and Saturday 16). The two most pronounced peaks are observed on Friday morning, from 7h to 8h, corresponding to the time most people go to work, and from 14h to 15h, matching the typical end of working day on Fridays.

Figure 14: Histogram of detections for each day of C31S camera.



On the other hand, the detections on Saturday increase in a smoother way, with two main peaks at 12h and in the afternoon, from 17h to 20h, corresponding to typical times when people travel for shopping or leisure during the weekend.

PAI - Speed

The figure 16 below shows some insights that can be obtained by analyzing the vehicles' speed with the corresponding dashboard panel. The speed limit for the C31N highway is 80 km/h, which matches the average speed of all detected objects (estimated at 79,96 km/h). In addition, the speed interval with more detections is between 65 km/h - 80 km/h. Vehicles exceeding the speed limit move with an average speed of 86,16 km/h. Focusing only on extreme values, the average speed of vehicles driving at an abnormally low speed for this highway is 30,90 km/h, whereas vehicles driving at an abnormally high speed show an average of 104,57 km/h.

Figure 16: Visualizations of PAI - Speed dashboard for C31N camera

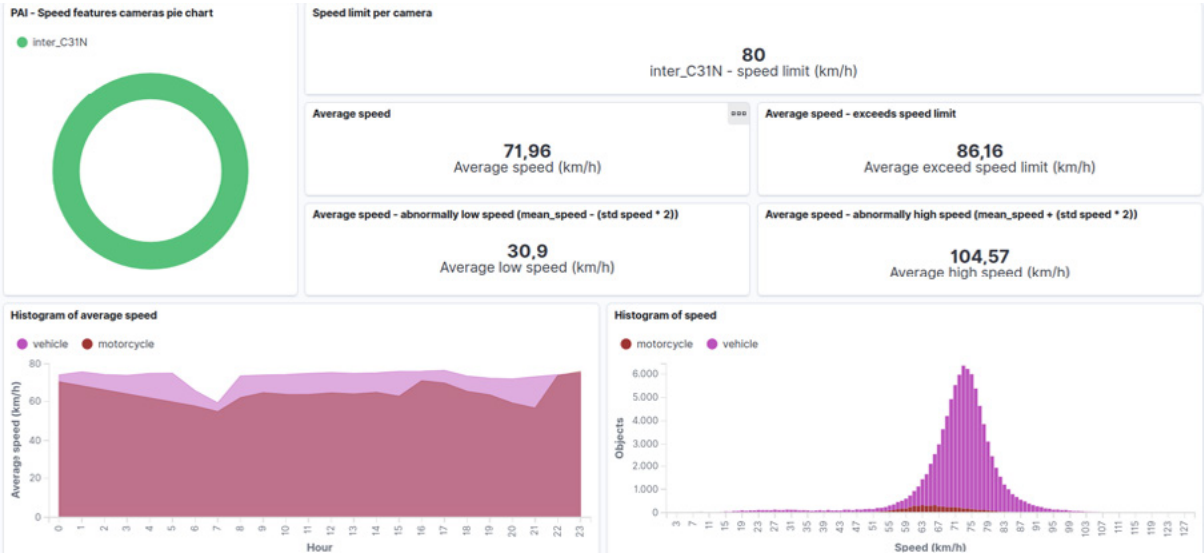
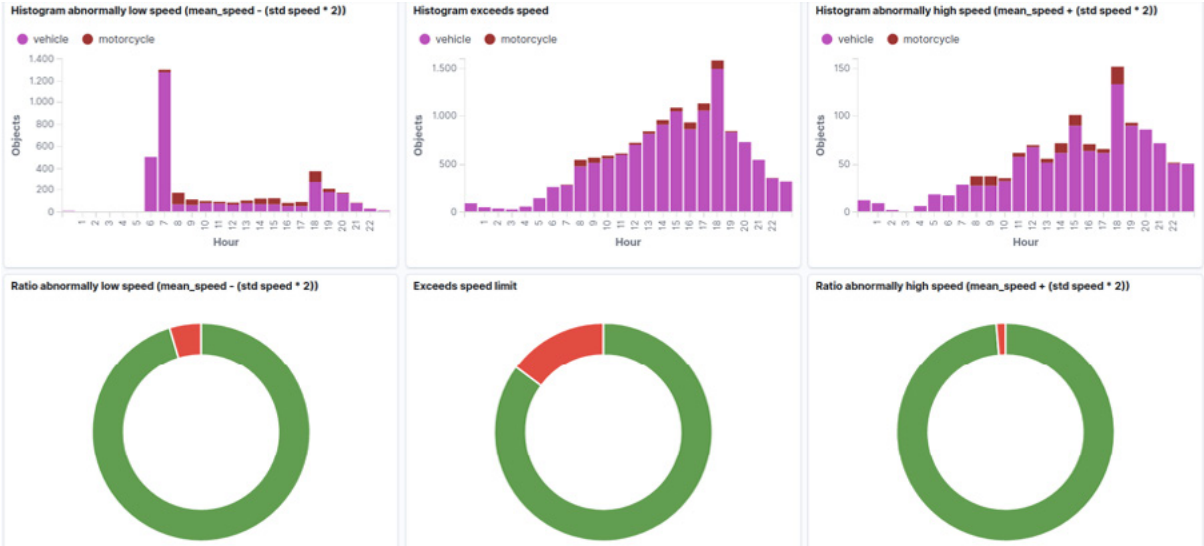


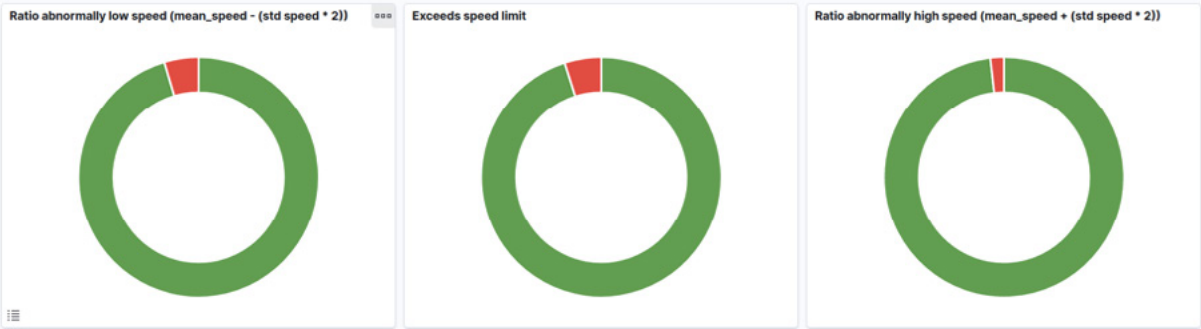
Figure 17: Histogram of abnormally speed and ratios for C31N camera



As we previously saw in the PAI - Congestion dashboard, figure 13, the average speed for the C31N had a fall at 7h. In the above image we can see that there is a peak in the number of vehicles with abnormally low speed, showing how the speed metric can serve as another indicator of congestion. On the other hand, almost 15% of vehicles exceed the speed limit.

By comparing with similar metrics observed on C31S, in the figure below, we can see that the ratio of vehicles exceeding the speed limit is approximately 5%. The reason for this lower percentage is because C31-S has a higher speed limit of 100km/h, compared to the 80km/h allowed in C31-N.

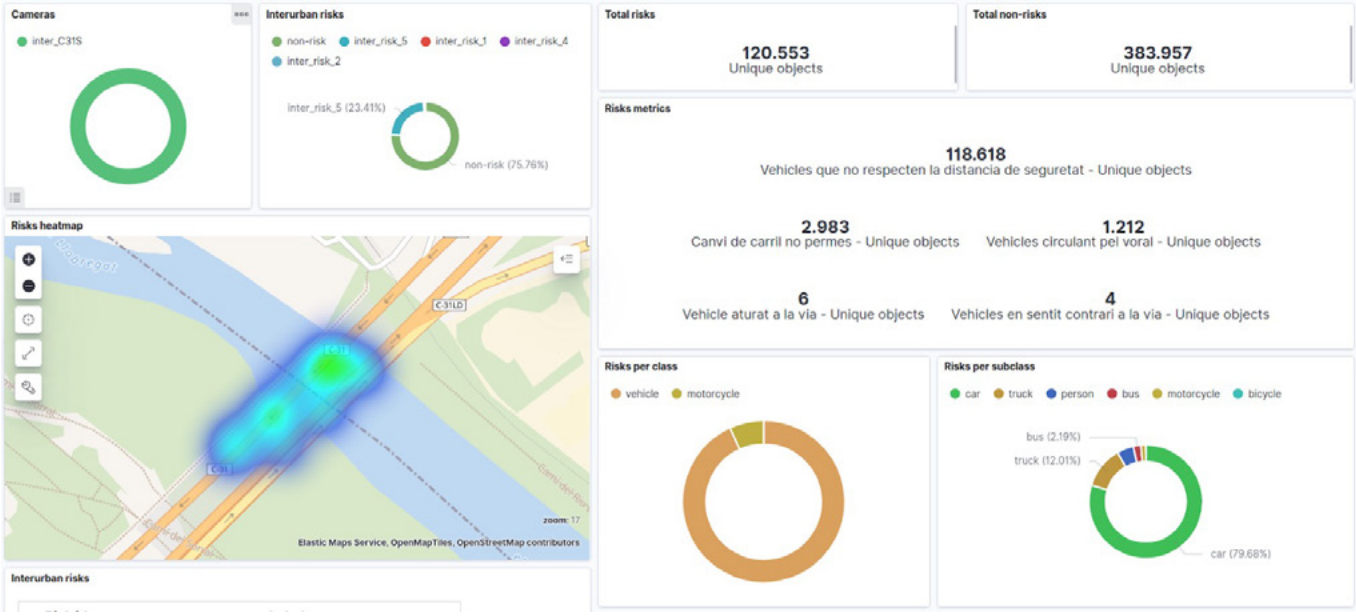
Figure 18: Speed ratios for C31S camera



PAI - Interurban risks

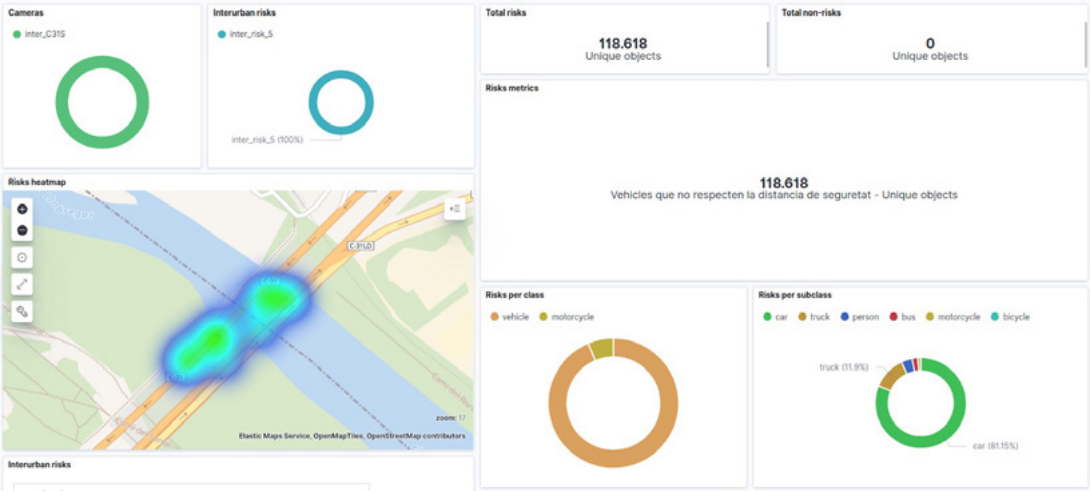
The below figure 19 shows the interurban risks for the C31S highway. As we can see, almost 25% of the detected vehicles are involved in interurban risk events.

Figure 19: PAI - Interurban risks filtered by camera C31S



The next figure 20 shows the panel for the most common interurban risk of vehicles not respecting the safety distance (risk 5, “Vehicles que no respecten la distància de seguretat”). As we can see, this risk takes place in the whole tracking area.

Figure 20: PAI - Interurban risks filtered by camera (C31S) for risk 5 (safety distance)



Filtering by risk 1 for the illegal lane change (“Canvi de carril no permès”), we can see on the map of the following image that this risk only takes place at the left lane in south orientation. This makes sense since this is the only lane with a solid line, forbidding the change of lane, as shown in figure 22. The majority of objects involved in this risk is of class vehicle (93%), with the remaining 7% of violations done by objects of class motorcycle.

Figure 21: PAI - Interurban risks filtered by camera C31S and inter_risk_1

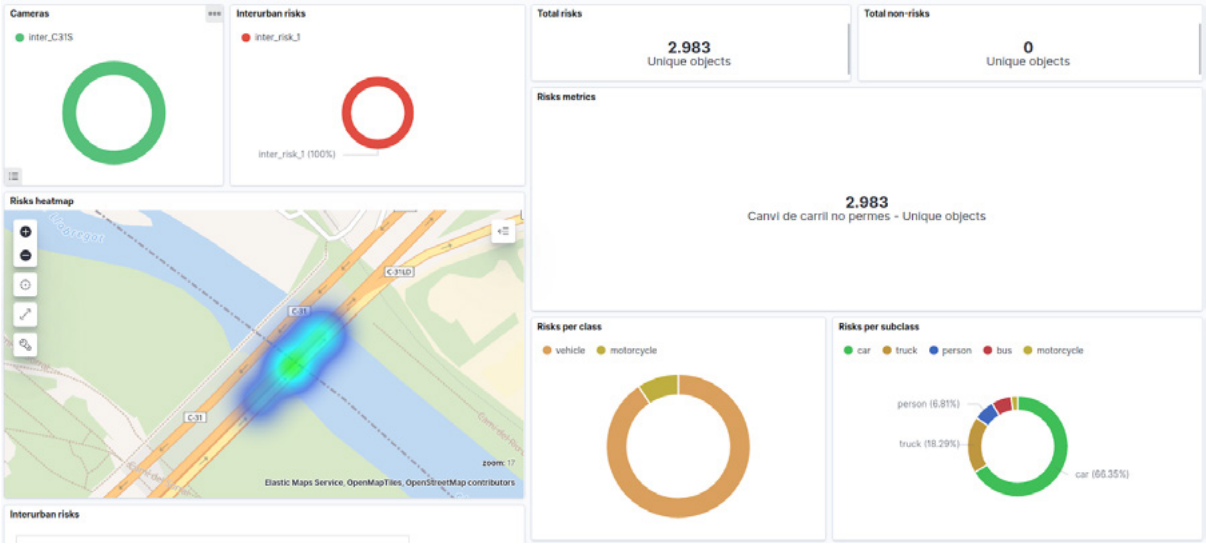
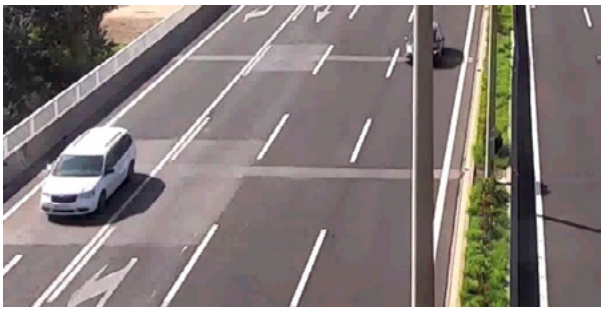
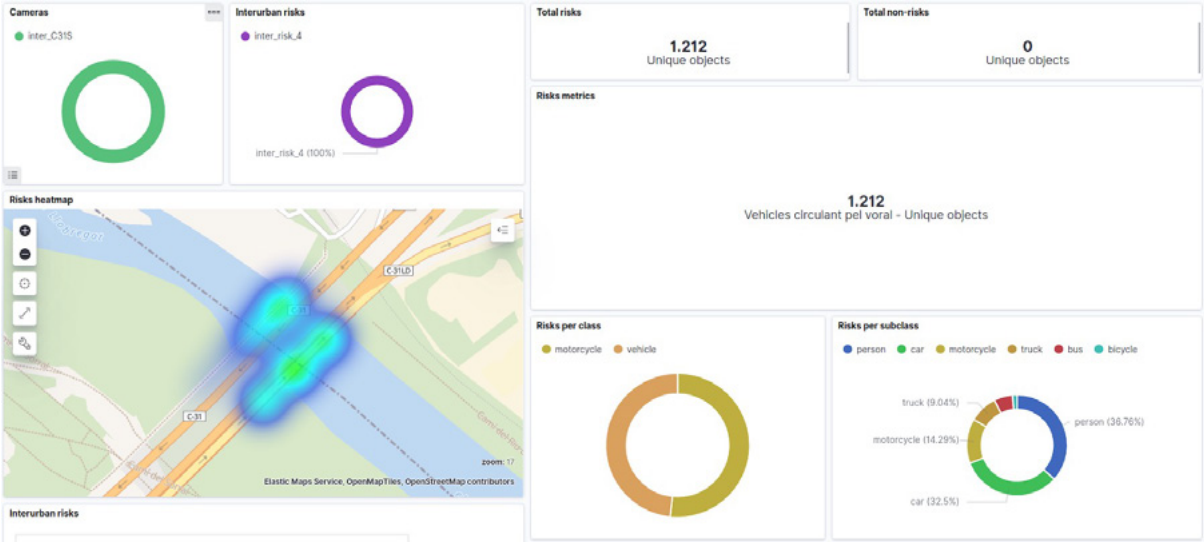


Figure 22: Lanes in south orientation for the C31S camera, showing the road lane with continuous line at the left part of the image



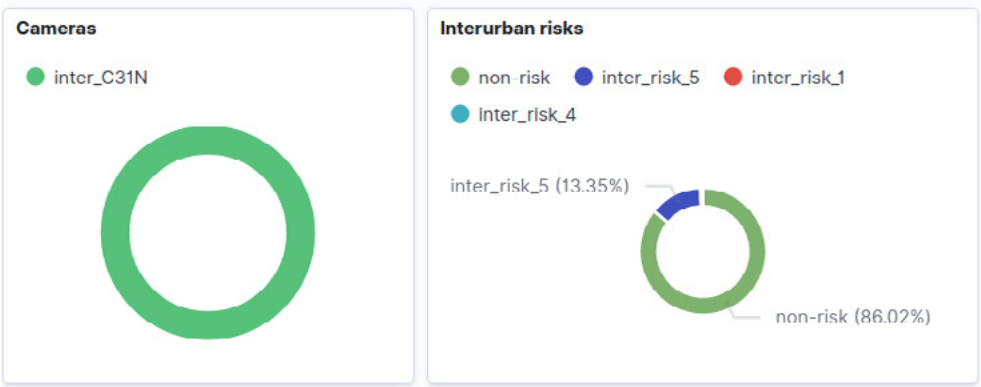
The below figure 23 shows the risk 4 for vehicles circulating on the shoulder lane ("Vehicles circulant pel voral"). The heatmap shows it is more common in the south orientation but it is happening in both orientations. The majority of objects involved in this risk is of class motorbike. Although, we have seen on the qualitative experiments that there are people walking on the shoulder of the lane, most of these detections on the subclass person are in real motorbikes due to the problem with the detector and the motorbike and person commented before.

Figure 23: PAI - Interurban risks filtered by camera C31S and inter_risk_4



Finally, to conclude the interurban risks comparing the data from the two highways, the ratio of observed risks in C31N is approximately 15% of total number of detections, as shown in the figure 24, whereas in C31S this percentage is higher by almost 10% (around 23.4%, as shown in Figure 19).

Figure 24: PAI - Interurban risks filtered by camera C31N



PAI - Urban risks

The following figure 25 focuses on the visualization of the detected urban risks. Almost 5% of the detected objects generate some urban risks. There is a lot of difference between subclasses. Unlike the suburban scenarios, in urban environments 96% of the detected risks involve persons.

Figure 25: PAI - All urban risks

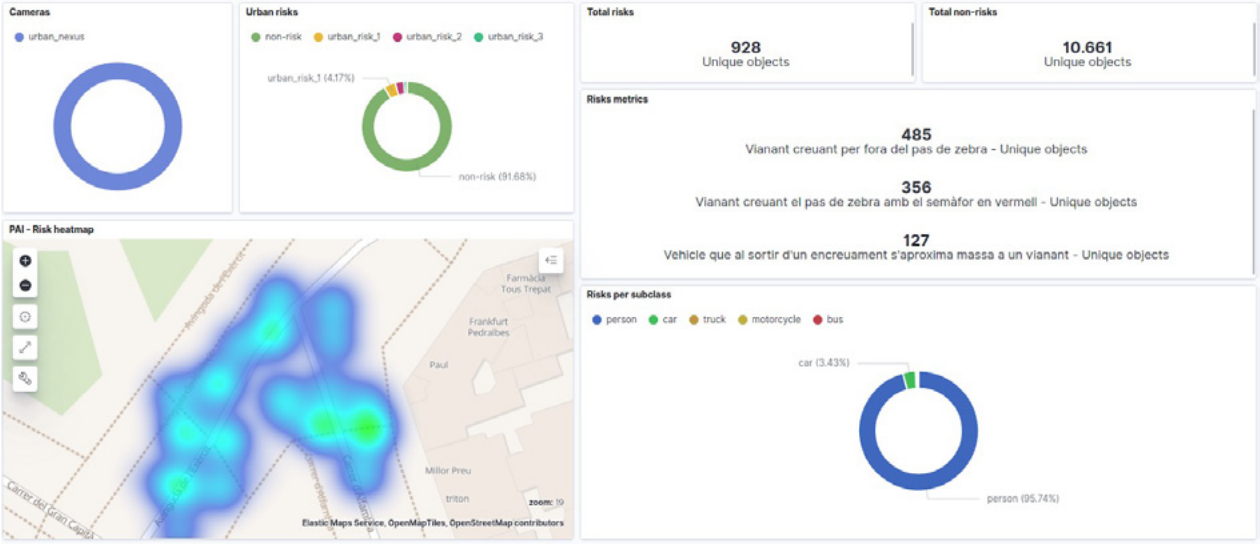


Figure 26 shows the urban risk 1 of pedestrians crossing outside the crossway ("Vianant creuant per fora del pas de zebra"), which matches the results shown in the heatmap with detections being outside the zebra crossing. In addition, in figure 27 you can see the detections of the urban camera, where the two pedestrian crossings are clearly visible. Clearly, we can see as the people cross in diagonal (heatmap of the figure 26) instead of using the two crosswalks to go to the opposite corner.

Figure 26: PAI - Urban risks filtered by urban_risk_1, heatmap showing where the people is crossing for outside of the the crosswalk

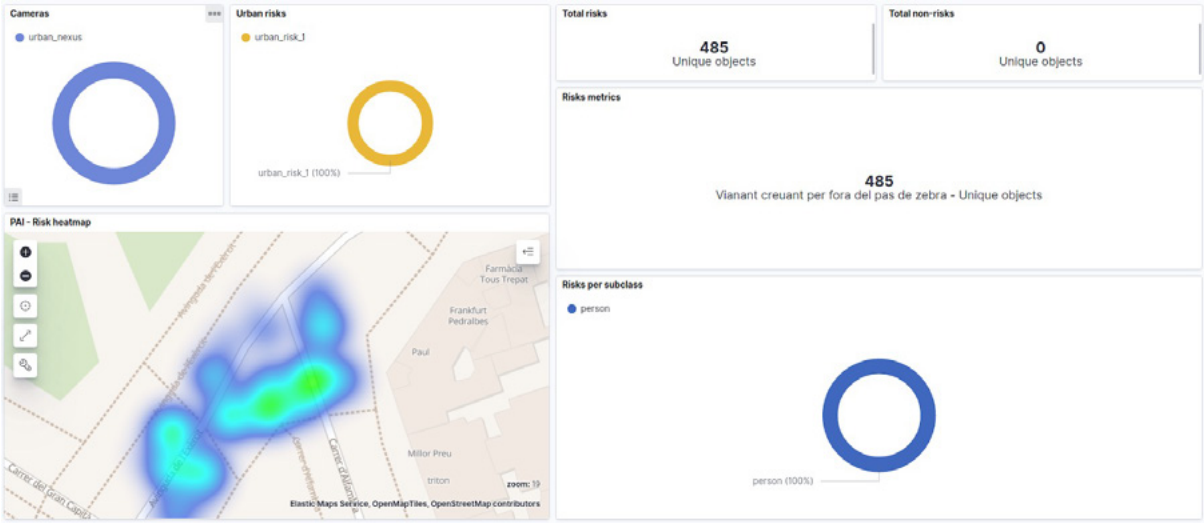


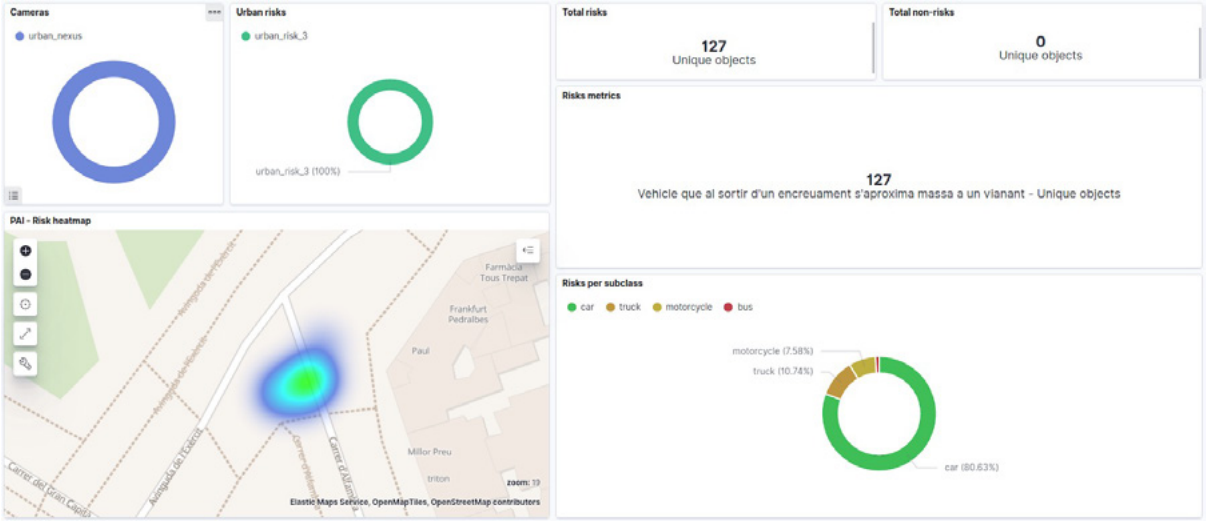
Figure 27: The urban camera detections. The two groups with the highest number of icons are the two zebra crossings.

The figure 28 below shows the urban risk 2 of pedestrians crossing with red ("Vianant creuant el pas de zebra amb el semàfor en vermell"). Unlike the heatmap of the previous figure, here it can be seen that all detections are concentrated in the same area that corresponds to the zebra crossing.

Figure 28: PAI - Urban risks urban_risk_2

To conclude the urban risk analysis, the figure 26 below shows the urban risk 3 of vehicles turning towards pedestrians crossing the street ("Vehicle que al sortir d'un encreuament s'aproxima massa a un vianant"). The subclass mainly responsible for this risk is of type "car", almost 81%., followed by "truck" and "motorcycle" types responsible for almost 11% and 8% of the cases, respectively.

Figure 26: PAI - Urban risks urban_risk_3



Technical Deliverables

Reference	Title	Delivery Date	Dissemination Level
Description			
D1.1	Use case definition and requirements	30/06/2021	CO
The first deliverable defined the overall project objectives and identified the specific urban and interurban use case scenarios, to be considered throughout the project. Furthermore, a number of functional requirements has been determined, regarding the performance of the analytics processes to be implemented.			
D2.1	Platform architecture	30/11/2021	CO
The second deliverable provided the overall platform architecture to be used for the execution of the developed analytics modules. Specifically, the key components of the platform have been specified, for storage, processing and visualization of the results. The overall data pipeline has also been defined, consisting of different stages, i.e., the video processing for the detection and tracking of objects, the application of semantics and finally the visualization of the results on the project's dashboard.			
D3.1	Video and image analytics modules	30/01/2022	CO
D3.1 provided a detailed description of the video and image analytics modules, implementing all the functionalities needed by the project's use cases. In the first part, the object and detection modules have been described, including functionalities for the geolocation of the detected objects and the detection of the traffic light status, needed by the urban use case scenarios. In the second part, the development of semantics maps required to annotate the areas of interest on the selected urban and interurban videos has been presented.			
D4.1	Use case demonstration and validation	30/03/2022	CO
The fourth deliverable focused on the demonstration and validation of the project's use cases in urban and interurban scenarios, focusing on the detection of specific risk events and the traffic flow analysis. This deliverable provided a qualitative and quantitative analysis of the performance of the implemented analytics modules. With respect to the visualization of the results, the project's dashboard has been implemented, offering different panels to show the outcome of detections, risks and traffic flow analysis.			
D5.1	Replicability study	30/03/2022	CO
D5.1 provided a replicability study aiming to identify the factors that affect the replication of the project's analytics methods to different scenarios and to maximize the impact of the proposed solutions. Different factors have been identified, such as location, type of road, camera features and positioning, climate and lighting conditions, etc. Furthermore, some potential improvements of the proposed solutions have been described.			
DF	Final project report	10/05/2022	PU
The final project report provided a global overview of the project objectives, methodology, implementation and validation of obtained results. This public report provides the reader with the necessary technical details regarding the project's contribution, as well as a discussion of the lessons learnt and next research steps.			

Conclusions

This project developed a methodology and a proof of concept solution based on AI and big data technologies to achieve the two key objectives of improving road safety and enhancing traffic management. The proposed software framework, which was successfully deployed both on premise and on the cloud, was used to process a large amount of raw data corresponding to more than 150 hours of recorded videos taken from cameras in urban and suburban scenarios.

With respect to the detection and tracking, the most appropriate DNN models have been selected and combined to provide the accuracy needed for the implementation of the use cases. Even though the evaluation of the detection and tracking processes was beyond the scope of the project, a very good performance was obtained, enabling the detection and tracking of multiple vehicle types at all times of day, despite the non-optimal video recordings and lighting conditions. Furthermore, most identified weaknesses, e.g., in the accurate detection of motorcycles, bicycles and very complex vehicles (e.g., trucks carrying cars), could be overcome by more extensive and targeted training of the employed models. The accurate tracking also enabled the estimation of vehicles' speed, which was a critical metric to determine several traffic related events, such as the detection of stopped vehicles or the recommended safety distance. The detection of the traffic light status was also a key component for the urban scenarios, enabling the identification of traffic violations without the need to acquire additional data from the city (e.g., from a traffic light management system).

The main contribution of the project focused on the risk event and traffic flow analysis. To that end, the project developed a corpus with the semantic information needed for the defined use cases, providing a methodology to describe a multitude of traffic related events that may be of interest in both urban and suburban settings. This methodology was then applied to annotate three specific examples, corresponding to the camera views of the three selected locations, along with a set of rules to detect the specified events. A thorough quantitative and qualitative evaluation of the obtained results was conducted, validating the proposed approach and leading to very valuable insights and lessons learnt.

As a representative example of the quantitative analysis for the suburban scenario. Performed manually annotating 1h for each of the interurban and urban cameras analyzed. The event of illegal lane change (i.e., crossing the solid lines) was detected with a precision of 89% and a recall of 93.67%. Similarly, for the urban scenario, the event of pedestrians crossing with red traffic light was detected with a precision of 75% and a recall of 100%. Another very interesting result was the very low number of false negative detections (the lowest observed recall was 93.75%), meaning that almost all risk situations were successfully detected. Nevertheless, it should be mentioned that some events had very low occurrence, or were not observed at all (e.g., vehicles stopped in the highway or vehicles turning towards pedestrians) within the processed video samples, stressing the need for more extensive datasets for future analysis.

Furthermore, the qualitative analysis of the results, also supported by the visualization of the data on the implemented dashboard panels, led to some very interesting conclusions from both a technical and a use case perspective. On the one hand, through an iterative process, the qualitative analysis enabled the refinement of the implemented detection and tracking modules, as well as the semantic annotation and rules, identifying the most common sources of error. On the other hand, the qualitative analysis led to the detection of some very interesting and high-risk situations that were not contemplated in the initial definition of the use cases, for instance the circulation of bicycles and pedestrians on the shoulder lane of the highway.

The visualization of the data on the dashboards also led to some interesting conclusions regarding the behavior of the vehicles. For example, up to 25% of the detected vehicles in interurban scenarios were involved in some type of risk, with the most common violation concerning the safety distance. Another observation was the fact that vehicles are more likely to exceed the speed limit when this is set to a lower value. Furthermore, there was an evident difference between the type of risks expected in a suburban setting, mainly involving cars and larger vehicles (trucks, buses, etc.) and urban scenarios, where pedestrians are mostly at risk in the vast majority of scenarios.

Finally, the traffic flow analysis focused on the identification of traffic congestion and vehicles circulating with speeds outside the expected limits (i.e., too high or too low for the type of road under examination). This analysis was facilitated by the design of a dedicated dashboard, and the most appropriate filtering mechanisms to combine the available data into useful plots (e.g., correlating number of vehi-

cles with their type and time of day). The obtained results were consistent with the traffic observed during the recorded videos, even though no significant traffic jams occurred within the recorded video samples. It was also shown that correlating speed analysis with detected vehicles is a good indicator of the presence of congestions.

Overall, the high level of performance obtained by the proposed prototype has demonstrated the strong potential of the developed AI and big data enabled approach to extract valuable information from the offline analysis of a large volume of recorded videos from urban and interurban locations. As an immediate next step, the developed methodology and platform could be exploited to process a larger video sample, in order to extend the analysis over a larger period of time and extract more statistically meaningful conclusions. Finally, the next section will highlight some lessons learnt and potential improvements that can further enhance the capabilities of the proposed system.

Technical Lessons Learned

During the project implementation, by following an iterative process between the development of the analytics modules and the validation of the obtained results, several technical challenges were identified. Some of them were tackled within the project, whereas others require further effort and research, in order to further refine and improve the obtained performance of the proposed approach.

With respect to the detection module, after a thorough state of the art review, it was determined that the initially considered detection and tracking module was not capable of detecting small objects located far from the camera. Since this was a necessary requirement, given the nature of the recorded videos with the cameras covering large road sections, a bigger model was selected for the object detection process. Even though the performance of the detector generally met the project requirements, there were still some cases in which specific errors were observed. Specifically, the detection was prone to failures during night, when light conditions were very poor. This problem can be mitigated by considering specific models trained under low light conditions. Another issue was the difficulty in correctly detecting motorcycles and bicycles, which can also be improved by extending the training set, taking into account the camera angle and distances.

With respect to the tracking module, the most frequent problems were concerning the occlusion of objects, when vehicles especially at the back part of the frame were hidden by other vehicles (suburban scenario) or trees and traffic signs (urban setting). In such cases, the tracking of objects often failed, with the tracker assigning different ids to the same object across consecutive frames, thus resulting in multiple detections of the same event. Such problems are expected to be less frequent by the use of an improved tracking module.

With respect to the semantic mapping, the most limiting factor was the movement of the camera. In some cases, the camera position was reset, leading to a drastic change of the recorded view, whereas in other occasions the angle was slightly altered due to the wind. However, since the semantic annotation was designed with respect to a specific video snapshot, any changes in the recording angle resulted in a mismatch of the semantics, affecting the correct detection of events. In the context of the project, this issue was tackled by selecting video samples where the image was stable, however, additional efforts would be needed to overcome this issue in a more extended deployment. Furthermore, a valuable insight acquired within the project was the need to take into account the perspective and angle of the video when defining the semantic maps for the cameras.

Regarding the detection of risk events, most false detections were caused by the errors introduced by the detection and tracking module. However, an additional inconvenience was the lack of sufficient samples corresponding to the defined use cases (e.g., showing vehicles stopped in the highway or moving in the wrong direction), which limited the validation of the proposed solutions.

Another issue that should be taken into consideration regarding the deployment of the software solution is the operational cost. Specifically, the plan at the beginning of the project was to fully host the software platform for the storage, processing and visualization on the cloud. However, due to the heavy computational load required for the big data processing, this solution was not feasible within the budget of the project, leading to the adoption of a hybrid approach, with some of the processing being done on premises. As a next step, other alternatives such as the use of edge computing could be considered, to reduce the amount of computation needed to take place in the cloud.

Finally, the implementation of the panel dashboards was proven to be a very valuable tool for the traffic flow analysis, as well as the visualization of the detected risks and other relevant parameters from the detection and tracking module. However, the selected platform based on Kibana and elasticsearch lacked two features that would further facilitate the analysis and could be considered in future implementations: i) the possibility to include videos in the analysis, enabling the correlation of observed metrics with the visual inspection of the obtained results, and ii) more advanced map visualizations, for the more detailed inspection of the tracker output. Furthermore, with respect to the implemented filters for the traffic flow analysis, even though the selected approach of computing congestion in 1-hour intervals is appropriate for the detection of highly congestion events, it cannot capture concentration of vehicles for smaller periods of time (e.g., a congestion lasting only a few minutes). To that end, adding more fine-grained filtering capabilities (e.g., computing congestion in 10-minute intervals) could provide more valuable information.

Recommendations for future activities

10.1 Benefits of replicating this project

The products obtained in this project represent a solid basis for the construction of risk and traffic flow detection systems using artificial intelligence. These can be useful in designing strategies for road safety based on technology, as well as in improving the quality of traffic information.

The main benefits of replicating this project include:

- Basis of knowledge and experimentation: new projects can use this project as a basis of knowledge and experimentation, which includes a list of use cases, a data repository, a technological architecture, analytical and artificial intelligence modules and the results of use cases.
- Preventive actions: the information obtained allows the authorities to guide on preventive measures before accidents or major traffic delays occur.
- Improving road safety: the product can automatically detect risks and traffic flows, reducing the number of accidents at strategic points and improving the safety of people and vehicles.
- Improvements in traffic and congestion: the results allow making decisions in order to minimize traffic jams and delays.
- Reduction of travel time: carrying out actions for the reduction of the will minimize the travel time of the drivers.
- Optimization of the waiting time at traffic lights on urban roads: it can be used to improve the times of traffic lights in cities and make traffic smoother.
- Reducing emissions of polluting gases: optimizing the flow of traffic, both on urban and interurban roads, can lead to a reduction in emissions of gases such as CO2 and NOX. Therefore, it would also have a positive environmental impact.

Over the next few years, the administrations and traffic authorities of many countries are considering a digital transformation strategy that will allow them to exploit information and data from different sources.

In order for decisions and processes to be more precise, it will be necessary to implement technological solutions such as the one developed in this project. Therefore, the reuse, replication or expansion of the products obtained in this project can give a great advantage for the development, implementation and production of such systems.

10.2 Factors affecting the replicability

In case we decide to scale this project and put it into production, different factors must be taken into account. This section explains the variables that affects the replication of the project. These factors include concepts such as location, type of road, type and number of traffic cameras or climate and lightning.

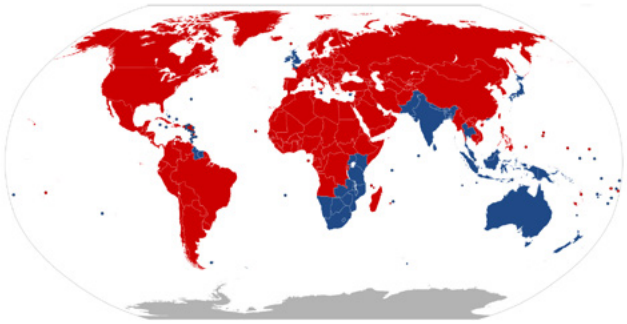
10.2.1 Location

It is important to consider the location where we will replicate the project. This factor is relevant as it not only determines the characteristics of the road, but also determines the rules of traffic and its signs.

The first factor to consider is the country where you want to replicate. Risk and traffic flow events may be similar in different countries, but in others, they may be completely different. It all depends on the traffic rules of each country, its road signs and the type of driving, which can be on the right or left.

Catalonia and Spain share similarities in traffic routes and have the same traffic rules and signs. Therefore, putting the solution into production would not present major changes in any location in the territory.

On the other hand, we find countries whose driving is on the left, such as the United Kingdom, India, Australia, among others. The figure below shows the countries where you drive on the right in red, and the countries where you drive on the left in blue.



Each camera has to be set up semantically, that is, manually set what a solid line is, what a zebra crossing is, and so on. It is a question of defining and making the algorithm understand what the signs of the route are. Next, on this semantics, the rules that indicate when there is a violation of traffic rules apply.

The location affects the semantics in the sense that it can vary the signals of the road. Therefore, there is the need of defining different types of lines and their meanings. For example, in some areas, white zigzag lines are used and in others a yellow solid line. In the case of Spain, or even most areas of the European Union, it would not be a big job. On the other hand, we should perform specific studies for other foreign countries. However, the algorithms and automations that detect these signals are of general purpose and could be adapted to these new environments.

On the other hand, location has also an impact on the rules that use these semantics. For example, in some foreign countries, we would need to define new rules for detecting the risks or traffic flow.

Another important point to have in mind is the case where you want to replicate the project in multiple locations. Each of these can have different cameras, semantics, rules, resolutions, encodings, and image extraction processes. This will involve different levels of effort depending on the scale of the new project. Finally, the geolocation of the objects in the scene would require a new mapping for each new camera. The mapping corresponds to a GPS identification (longitude and latitude) of the point of contact of four elements with the ground where the objects move. These points must be visible from the point of view of the camera.

10.2.2 Type of road

The second factor that may affect the replicability of this project is the type of mobility road. That is, whether it is an urban or interurban route, and their sub-types. The urban route has different characteristics from the interurban one. In the first, traffic lights, pedestrians, lighter vehicles and less speed predominate. On the other hand, in the second, there are large vehicles, at high speeds and the dimensions are larger.

This project has defined different types of risk and traffic flow on urban and interurban roads, which include the majority of cases. However, it is necessary to keep in mind that the proposed model has limited coverage and, therefore, it is possible that there are traffic situations that are not covered.

For example, multi-level road scenarios, tunnels or bridges have not been included. These scenarios would need special treatment and we should define them in future developments.

Another aspect to consider is that depending on the type of road, unexpected or extraordinary situations may occur that had not been contemplated. For example, on an urban road, there may be a high concentration of people waiting at a traffic light because it is the rush hour of a public holiday. In these cases, we would have to study the impact on the system, the speed of execution of the processes, etc. We can only obtain these answers by analysing each type of specific case study. However, there are not a considerable amount of this kind of situations and they can be treated specifically.

Finally, although the algorithms apply to multiple roads, there are components that we can optimize according to the type of road. For example, we could calibrate the tracking system according to the average speed of the cars, or we could optimize it for straight roads, curved, roundabouts or other cases.

10.2.3 Type of traffic camera

Another factor to consider is the type of traffic camera used to record the videos. It will be important to get a clear and detailed picture of the road so that you can analyze the risks and traffic flows.

Each camera has certain characteristics and specifications that may or may not favor the development of the project. These variables include, for example, the resolution of the camera, which is the number of pixels generated by the image of the digital sensor behind the optical lens. The format of the video, which may depend on the manufacturer and the codecs used. On the other hand, we have the zoom and the recording angle, which are adjustable settings at the recording location, and in addition, image stability can play an important role in recording, as a shacking video can prevent the proper functioning of the solution obtained. Finally, we can also consider the connection and data transmission of the camera.

The solution obtained in this project is not limited to a certain type of camera, with fixed characteristics. However, there will be conditions in which they will work better or worse. We should have to check that the videos in the camera respond well to the models and algorithms used.

The camera resolution and zoom will affect the detection of people, motorcycles and small cars. We will need approximately 8x25, 12x30 and 30x25 pixels for people, motorcycles and small cars respectively in order the detector to recognize them. On the other hand, if there is a lot of zoom and the vehicles occupy a large part of the image, then there will not be enough information to determine its trajectory.

The perspective influences the detection and tracking of people and vehicles in the scene. If the camera angle is too vertical, there can be objects that difficult the detection and tracking of our vehicles.

Image stability will influence the geo-location and semantics of the scene. Depending on the mobility of the camera and the distance from the objects to it, both the mapping of the geo-location and the mapping of the camera with its semantic map may contain a deviation. Analytics would apply anyway, but performance would be affected. In this project, we have seen that stability can be a problem in the accuracy of object positioning, but we can solve it through stabilizer software and we could re-use it for other cameras.

In terms of data connection and transmission, it would not play a very important role because the detection and tracking are done in batch after obtaining the videos. In the case that we want to evolve into a semi-real-time or real-time system, this would be an important aspect to consider. Therefore, we should define new data flows in real-time infrastructures.

Finally, we can consider the concept of a fixed camera or mobile camera. Fixed-type traffic cameras are at one point on the road and they capture images of the scene. Mobile cameras can be temporarily installed to capture videos at a given time, or they can capture moving images.

Cameras in movement have not been included in this project. For example, a mobile camera mounted on a patrol car. We should study how the movement affects the artificial intelligence module. In addition, what are the consequences of having a video with different levels of vibrations.

Another case of a mobile camera would be the capture of videos through drones or helicopters.

In this case, we can identify new semantics, new risks and new traffic flow scenarios. Studying the scene from a higher perspective will create videos with more vertical angles. However, the stability and vibration of the drone due to gusts of wind would affect the quality of the videos.

10.2.4 Number of traffic cameras

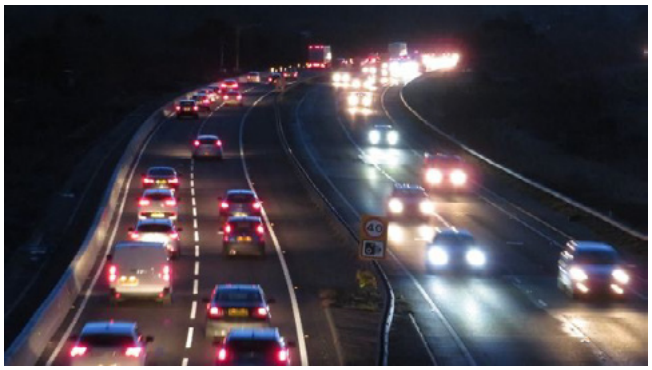
The number of traffic cameras can also play an important role in replicating the project, especially if you want to scale and use a large number of cameras of different types and in different locations.

Each camera uses different semantic maps, so you need to create a new semantic map for each new camera. This can be an impediment to scaling the number of cameras used horizontally. On the other hand, the geolocation of objects in the scene will also need a new mapping for each new camera.

10.2.5 Climate and lighting

Climate is an important variable to consider as it can affect the performance of algorithms and the results obtained. Precipitation such as rain, snow, hail can alter the image and affect the quality of the recorded videos. That is why it is important that the climate is stable and without precipitation, if possible. In addition, extraordinary situations can also occur, for example when there are large traffic jams due to heavy rainfall.

Another factor related to weather is the lighting of the scene. The hours of light available must be taken into account as the hours of sunshine will have a positive effect on the results obtained, while the hours with darkness can make them worse. In some cases, there are routes that have artificial lighting and the algorithms can detect the objects. However, in natural light conditions, the camera's sensors can better capture image information. It can also cause glare, affecting the system performance.



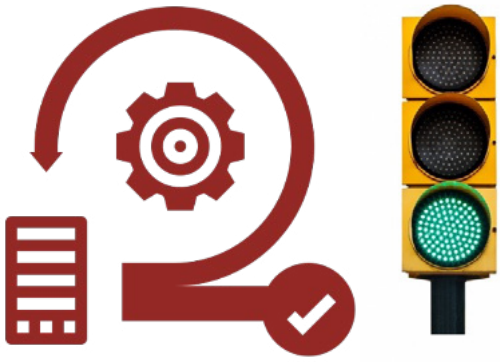
In this project, we have not deeply analysed the limits of the algorithms for the detection and monitoring of vehicles and pedestrians. Therefore, when replicating the solution, we should take into account that there can be performance changes in situations of special climates and lighting.

Finally, other weather or human activity scenarios can cause visibility issues and we should consider them. For example, fog, smoke, dust storms, or even air pollution.

10.3 Replicability requirements and procedures

This section aims to answer the question of what do we need to replicate the project.

In order to carry out the replication, it will be necessary to have staff and a human team with the appropriate knowledge, of both technologies and procedures to obtain good results. In this sense, they must have the appropriate training in artificial intelligence, infrastructure for the processing of massive data and traffic knowledge.



In addition, we will have to make a financial investment for the storage, information processing and analytics infrastructures. The costs may vary depending on the number of videos we have to analyse, and the degree of reuse of the solution developed.

There are three main scenarios to replicate the products obtained or part of them. The first is a direct replicability, where the whole solution is used without any new development, the second is a replicability where the model is re-trained or refined, and finally, the scenario where the project is replicated by making changes in the architecture, procedures or technologies of the solution.

The following subsections present the requirements needed in each of the above scenarios.

10.3.1 Direct replicability

In the case of direct replication, we will reuse the entire system for risk and traffic flow detection without making any changes to the architecture, procedures or technologies, and without any refinement or re-training of the model obtained.

To implement successfully the new analysis you will need to meet the following requirements:

- Use cases: The use cases of accident risks and traffic flow must be the same as those used for this project.
- Batch / Real-time: Video analysis have to be in batch mode, as done in this project.
- Video Features: You will need to provide new videos with features equivalent to those used:
  - Locations with signs and traffic rules
  - Number and duration of videos
  - Scenarios with light, without glare or precipitation
  - Single-level traffic lanes (no tunnels, bridges or underpasses or overpasses)
  - Minimum 8x25, 12x30 and 30x25 pixels for detecting people, motorcycles and small cars, respectively
  - Fixed and stable traffic camera
  - Scenes without visual obstruction objects
- Cloud platform credits: Credits will be required for data storage on the platform, for advanced video processing and analytics, and for the results reporting module.
- Video pre-processing: You will need to create new semantic maps and object mappings to geo-localise the objects. This process must be done for each camera you want to analyse.

In most cases, the performance and accuracy obtained with direct replication will be similar to that of this project. However, they may be affected by different video factors, such as scene type, object type, or weather conditions.

In case you decide to reuse the solution obtained without making any changes, a guide of the steps that should be followed is as follows:

- Recording and collecting new videos.
- Risk assessment and traffic flow to be analyzed. If the analyzes in this project are sufficient, direct replication can be continued.
- Upload videos to the storage platform.
- Make intrinsic image corrections for each camera if necessary.
- New semantic map for each new camera.
- New geolocation mapping for each new camera.
- Analyze the videos with the already trained model.
- Obtaining results, evaluation and presentation of the same.

10.3.2 Replicability with re-training or fine-tuning

The second case of replication is based on the re-use of the entire architecture and the data pipeline , but in this case a re-training or refinement of the model is performed. It should be noted that the model used is general and we can find situations and scenarios in which its performance and predictions are negatively affected. In these cases an optimization of the module of analysis and detection of the cases of use can be done.

An example of a situation in which the model could be optimized could be a scenario with objects too small to be identified. In this case, the re-training would take into account the new size of the objects and then the system would be able to detect them.

It is important to know that you can re-train and improve your model without starting from scratch. That is, the model will retain its knowledge base and will be able to incorporate new information. As an example, we could add new elements or types of vehicles to detect without having to train the model from scratch, which is a process that requires great computing power and considerable time.

It can also be re-trained to optimize certain scenes, for example, with a specific angle or height, or to detect different types of specific vehicles that were not covered in the data set used.

If you want to replicate the project through re-training or tuning, the following requirements must be taken into account:

- Use cases: New use cases can be added for accident risks and traffic flow. However, its detection will need to be developed and implemented.
- Batch / Real-time: Video analysis must be performed in batch mode, as done in this project.
- Re -training videos: You will need to provide a large number of tagged videos that contain information about the scenarios you want to optimize, whether it's to detect new objects, to include new detection cases, or to optimize the position of a particular scene. In addition, the quality of the videos must be equivalent to the videos to which the model inference will later be applied.
- Cloud Platform Credits: Credits will be required for data storage on the platform, to build the necessary infrastructure for re-training, for advanced video processing and analytics, and for the results reporting module.
- Video pre-processing: New semantic mapping and new mapping need to be created for geolocation of scene objects. This process must be done for each camera you want to analyze.
- Re-training: Dynamic clustering with GPUs and the appropriate libraries will be needed to fine-tune the model.

On the other hand, the following points are recommended to follow when replicating:

- In the case of cameras with different resolutions, it may be advisable to re-train with the cameras with the lowest resolution.
- If possible, use existing video sets that are public and contain tagged data. This can save effort and time.
- Different weather conditions can be tested, but the impact on model performance will need to be studied.

In case we need to optimize the analytics modules in order to deal with a new use case or a specific scenario, the following phases must be followed:

- Get new workout videos. There are two options:
  - Use existing video sets that contain tagged data.
  - Record videos and tag them.
- Upload training videos to the storage platform.
- Make intrinsic image corrections for each camera if necessary.
- New semantic map for each new camera.
- New geolocation mapping for each new camera.
- Re-training of the model with machines with GPUs and corresponding libraries.
- Evaluation of the performance metrics of analytics models.
- If the model improves, analyze the new videos you want to infer.
- Obtaining results, evaluation and presentation of the same.

10.3.3 Replicability with changes in modules, architecture or technologies

The latest replicability scenario includes changes in the technology architecture and components of the data pipeline developed. In this case, components of the back-end, ETLs, analytics and artificial intelligence modules, or front-end and reporting could be changed. We can carry this type of implementation when significant improvements are needed.

An example would be the evolution of the system to be able to analyse videos in real time. The entire data architecture needs to be redesigned in order to implement a real-time data pipeline. That is why certain technologies and developments are needed that have been left out of the scope of this project. In this case, the infrastructures that guarantee the capture, ingestion and processing of the videos will also be of vital importance. In addition, camcorders must be

compatible and have a good connection to ensure proper transmission.

On the other hand, the analytics module could also be redesigned to use other more efficient algorithms. There are different branches of algorithms designed to be more accurate in specific situations. For example, when the camera is very high on a highway and the size of the cars is very small. Therefore, different algorithms could be used for the various positions of the cameras and thus optimize the results.

Additionally, there may be new cases of risk or flow usage that require changes to the analytics module. Therefore, they must be modified or added in order to be detected.

Another aspect that may evolve is the front-end components. If you want to improve user interaction or include more features, some of the technologies used may need to be modified. In addition, additional information such as indicators or metrics may also be included.

Finally, a change of data storage and processing service provider can be made. Microsoft Azure has been used in this project. However, the solution could be migrated to other data platforms and the change would not be a huge effort. One recommendation is to check if a wrapper is needed to read the back-end videos with the new service provider.

There are no specific steps to follow in this scenario, however the following points should be considered:

- Design of the new solution, architecture and technologies to be used
- Cloud platform where you want to store and process videos
- Implementation of new more efficient algorithms
- Validation and tests of the correct operation of the entire pipeline

10.4 Risks and mitigations

When making use of this project, different risks and mitigations may arise that should be taken into account when replicating the project. The following table summarizes these circumstances and presents how to deal with them to ensure the success of the application.

Risks	Mitigations
• Scenarios with objects that hinder the visibility of people and vehicles in the scene	• Selection of new scene or change of camera orientation if possible • Modification of the monitoring module
• Scenario with objects too small for detection	• Use cameras with higher resolution or zoom in the desired area • Re-train with new object size if possible
• High number of cameras to analyse	• Improvement and further automation of the processes needed to obtain and build the necessary information from the cameras (semantics, homography, mapping, etc.)
• Scenarios with moving cameras	• Modification of the positioning module and the semantic map

Potential impact of the solution

This last section presents, on the one hand, the improvements that could be applied to the products obtained and, on the other hand, what functionalities could be included and what use cases could be developed in the future.

11.1 Potential improvements

At the end of the project, results have been obtained satisfactorily; however, there are always details that could be improved. A first improvement would be to make a performance study of the analytical models in different situations, whether it is weather conditions, lighting or other types of roads and / or scenarios. A study of the detection of false positives and false negatives could also be included to have a more complete assessment of what permissiveness is in each condition.

It could also include scenarios of more complex roads such as urban and interurban roads with different levels, tunnels and / or bridges.

Another significant improvement is the implementation of an architecture that handles videos in real time. This way, the video could be picked up directly from the camera, sent through the entire pipeline of analytics, and trigger real-time alerts for traffic controllers so that they could make decisions.

Additionally, the definitions of use cases of risk situations and traffic flow could be expanded in order to consider more casuistry. Below are events that could be considered as solution improvement.

Potential new risk scenarios:

- Number of times a vehicle (motorcycle, car, truck) does not respect the continuous line we see on the left, with smooth traffic or heavy traffic.
- Presence of pedestrians on the sidewalk.
- Obstacles to the road (loss of load).
- Separation between vehicles.
- Incorrect maneuvers.
- Climatic incidents such as rain, fog, vibration of the image by the wind, glare by sunrise and sunset, ...
- Drivers who talk on their cell phones or do not wear a seat belt.

Potential new traffic flow detections:

- Vehicles traveling in the middle lane (distinguishing between motorcycles, cars or trucks (*)) with “little traffic”. It could be defined as low traffic.
- Number of motorcycles circulating between vehicles, with vehicles in motion or with vehicles completely stopped.
- Travel time per direction of traffic within the section included in the display window (could be separated by lanes, but in principle would not be necessary).
- In relation to the detection of trucks, it would be desirable to be able to discriminate between transport vehicles in heavy tonnage (articular type) and low tonnage (distribution type).
- Formation of congestion of different levels of severity (for example depending on the density per section).
- Calculation of travel time between two or more cameras with AI functionality. In this way it will be possible to obtain a much more robust value of the travel time compared to that obtained in the “viewing window”.
- Number of occupants in a vehicle traveling on a BUSVAO lane.

References

12

[51] Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixe. Tracking without bells and whistles. In ICCV, 2019.

[vision_zero] European Comission, "EU road safety policy framework 2021-2030 – Next steps towards 'Vision Zero'", EU 2020, doi:10.2832/391271.

[1] Heckbert, Paul. "Projective mappings for image warping." Image-Based Modeling and Rendering 869 (1999).

[2] Luo, Wenhan, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. "Multiple object tracking: A literature review." Artificial Intelligence 293 (2021): 103448.

[3] Zhao, Zhong-Qiu, Peng Zheng, Shou-tao Xu, and Xindong Wu. "Object detection with deep learning: A review." IEEE transactions on neural networks and learning systems 30, no. 11 (2019): 3212-3232.

[Wang20] Wang, Zhongdao, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang. "Towards real-time multi-object tracking." In European Conference on Computer Vision, pp. 107-122. Springer, Cham, 2020.

[Berinetto16] Fully-convolutional siamese networks for object tracking (siamFC), L Bertinetto, J Valmadre, JF Henriques... - ECCV 2016 - Springer

[Voigtlaender20] Siam r-cnn: Visual tracking by re-detection, P Voigtlaender, J Luiten, PHS Torr... - CVPR 2020

[Shuai21] SiamMOT: Siamese Multi-Object Tracking, B Shuai, A Berneshawi, X Li... - CVPR21

[Shaoqing15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NeurIPS, 2015

[Wojke17] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In ICIP, 2017.

[Bewley16] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In 2016 IEEE International Conference on Image Processing (ICIP), pages 3464–3468. IEEE, 2016.

[DGT_news] https://revista.dgt.es/revista/num242/mobile/index.html#p=10

[Yilmaz06] Yilmaz, Alper, Omar Javed, and Mubarak Shah. "Object tracking: A survey." Acm computing surveys (CSUR) 38, no. 4 (2006): 13-es.

[Huerta15]Huerta, Ivan, Michael B. Holte, Thomas B. Moeslund, and Jordi Gonzàlez. "Chromatic shadow detection and tracking for moving foreground segmentation." Image and Vision Computing 41 (2015): 42-53.

[Braso20] Brasó, Guillem, and Laura Leal-Taixé. "Learning a neural solver for multiple object tracking." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6247-6257. 2020.

[Berclaz06] Jerome Berclaz, Francois Fleuret, and Pascal Fua. Robust people tracking with global trajectory optimization. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006.

[Evangelidis08] Georgios D Evangelidis and Emmanouil Z Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(10):1858–1865, 2008.

[Kim15] Chanho Kim, Fuxin Li, Arridhana Ciptadi, and James M Rehg. Multiple hypothesis tracking revisited. In ICCV, 2015

[ChienYao21] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "Scaled-yolov4: Scaling cross stage partial network." In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, pp. 13029-13038. 2021.

[Tan20] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10781-10790. 2020.

[He17] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. ICCV, pages 2980–2988, 2017.

[Lin17] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, ´ Bharath Hariharan, and Serge Belongie. Focal loss for dense object detection. ICCV, 2017

[Redmon18] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[Long20] Xiang Long, Kaipeng Deng, Guanzhong Wang, Yang Zhang, Qingqing Dang, Yuan Gao, Hui Shen, Jianguo Ren, Shumin Han, Errui Ding, et al. PP-YOLO: An effective and efficient implementation of object detector. arXiv preprint arXiv:2007.12099, 2020.

[Bochkovskiy20] Alexey Bochkovskiy, Chien-Yao Wang, and HongYuan Mark Liao. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.

Appendix A

13

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

C/Bilbao, 72 Edif. A.
08005 - Barcelona
Tel. +34 93 7419 100
info@cidai-catalonia-ai.eu
www.cidai.eu