



Universidad
de Navarra

**Computer Vision and Deep Learning
based road monitoring towards a
Connected, Cooperative and Automated
Mobility**

DISSERTATION

to obtain the degree of
Doctor of Philosophy in Applied Engineering by

Olatz Iparraguirre Gil

under the supervision of
Dr. Diego Borro Yáguez
Dr. Alfonso Brazález Guerra

Donostia – San Sebastián, November 2022

This thesis has been carried out at the Ceit Technology Center. Part of this research was funded by Basque Government (HAZITEK program) in the frame of project "TRAFIK DATA: On-board module for monitoring road signs and weather conditions with co-operative services communications" under the grant number ZL-2019/00753 and (ELKARTEK program) through the project "AUTOEV@AL: Technological Evolution For Multivehicle Automation And Evaluation Of Highly Automated Driving Functions " under the grant number KK-2021/00123.

*For the women who persist,
let's break the bias.*

Abstract



The future of mobility will be connected, cooperative and autonomous. All vehicles on the road will be connected to each other as well as to the infrastructure. Traffic will be mixed and human-driven vehicles will coexist alongside self-driving vehicles of different levels of automation. This mobility model will bring greater safety and efficiency in driving, as well as more sustainable and inclusive transport.

For this future to be possible, vehicular communications, as well as perception systems, become indispensable. Perception systems are capable of understanding the environment and adapting driving behaviour to it (following the trajectory, adjusting speed, overtaking manoeuvres, lane changes, etc.). However, these autonomous systems have limitations that make their operation not possible in certain circumstances (low visibility, dense traffic, poor infrastructure conditions, etc.). This unexpected event would trigger the system to transfer control to the driver, which could become an important safety weakness. At this point, communication between different elements of the road network becomes important since the impact of these unexpected events can be mitigated or even avoided as long as the vehicle has access to dynamic road information. This information would make it possible to anticipate the disengagement of the automated system and to adapt the driving task or prepare the control transfer less abruptly.

In this thesis, we propose to develop a road monitoring system that, installed in vehicles travelling on the road network, performs automatic auscultation of the status of the infrastructure and can detect critical events for driving. In the context of this research work, the aim is to develop three independent modules: 1) a system for detecting fog and classifying the degree of visibility; 2) a system for recognising traffic signs; 3) a system for detecting defects in road lines. This solution will make it possible to generate cooperative services for the communication of critical road events to other road users. It

Abstract

will also allow the inventory of assets to facilitate the management of maintenance and investment tasks for infrastructure managers. In addition, it also opens the way for autonomous driving by being able to better manage transitions of control in critical situations and by preparing the infrastructure for the reception of self-driving vehicles with high levels of automation.

Resumen



El futuro de la movilidad será conectada, cooperativa y autónoma. Todos los vehículos de la carretera estarán conectados entre sí, así como con la infraestructura. El tráfico será mixto y vehículos tripulados por humanos convivirán junto con vehículos de diferentes niveles de automatización. Este modelo de movilidad traerá consigo una mayor seguridad y eficiencia en la conducción, así como un transporte más sostenible e inclusivo.

Para que este futuro sea posible, las comunicaciones vehiculares, así como los sistemas de percepción, se vuelven imprescindibles. Los sistemas de percepción son capaces de entender el entorno y adaptar la conducción al mismo (seguir la trayectoria, adecuar la velocidad, maniobras de adelantamiento, cambio de carril etc.). Sin embargo, estos sistemas autónomos tienen limitaciones que hacen que en ciertas circunstancias su funcionamiento no sea posible (baja visibilidad, tráfico denso, infraestructura en malas condiciones etc.). Este imprevisto haría que el sistema transfiera el control al conductor, lo que puede convertirse en un problema de seguridad vial. En este punto, la comunicación entre los distintos elementos de la red de carreteras cobra especial importancia, ya que el impacto de estos imprevistos puede mitigarse o incluso evitarse si el vehículo tiene acceso a información dinámica de la carretera. Esta información permitiría anticipar la desconexión del sistema automatizado y adaptar la tarea de conducción o preparar la transferencia de control de forma menos brusca.

En esta tesis, se propone desarrollar un sistema de monitorización de la carretera que, instalado en vehículos que recorran la red viaria, realice una auscultación automática del estado de la infraestructura y pueda detectar a su vez eventos críticos para la conducción. En el contexto de este trabajo de investigación se pretende desarrollar tres módulos independientes: 1) un sistema de detección de niebla y clasificación del grado de visibilidad; 2) un sistema de reconocimiento de señales de tráfico; 3) un sistema de detección

Resumen

de defectos en las líneas de la carretera. Esta solución permitirá generar servicios cooperativos para la comunicación de eventos críticos de la carretera al resto de usuarios. Del mismo modo permitirá realizar el inventariado de activos para facilitar la gestión de tareas de mantenimiento e inversiones a los gestores de la infraestructura. Además, abre camino también a la conducción autónoma pudiendo gestionar mejor las transiciones de control en situaciones críticas y poniendo a punto la infraestructura para la acogida de vehículos con niveles de automatización elevados.

Laburpena



Etorkizuneko mugikortasuna konektatua, kooperatiboa eta autonomia izango da. Errepideko ibilgailu guztiak elkarren artean konektatuta egongo dira, baita azpiegiturarekin ere. Trafikoa mistoa izango da, eta gizakiek gidatutako ibilgailuak hainbat automatizazio-mailatako ibilgailuekin batera biziko dira. Mugikortasun-eredu horrek segurtasun eta eraginkortasun handiagoa ekarriko du gidatzean, bai eta garraio jasangarriagoa eta inklusiboagoa ere.

Etorkizun hori posible izan dadin, ibilgailu-komunikazioak eta pertzepzio-sistemak ezinbestekoak dira. Pertzepzio-sistemak gai dira ingurunea ulertzeko eta gidatzeko modua horretara egokitzeko (ibilbideari jarraitzea, abiadura egokitzea, aurreratzeko maniobrak, errei-aldaketa, etab.). Hala ere, sistema autonomo horiek mugak dituzte, eta, horren ondorioz, zenbait egoeratan ezin dute funtzionatu (ikuspen urria, trafiko handia, baldintza txarreko azpiegitura, etab.). Ezusteko horren ondorioz, sistemak kontrola gidariari transferituko lioke, eta hori bide-segurtasuneko arazo bihur daiteke. Puntu horretan, errepide-sareko elementuen arteko komunikazioa bereziki garrantzitsua da, ezusteko horien eragina arindu edo saihestu egin baitaiteke ibilgailuak errepideari buruzko informazio dinamiko eskura badu. Informazio horri esker, sistema automatizatuaren deskonexioa aurreikusi ahal izango litzateke, eta gidatze-lana egokitu edo kontrol-transferentzia hain zakarra izan gabe prestatu.

Tesi honetan, errepidea monitorizatzeko sistema bat garatzea proposatzen da. Sistema horrek, ibilgailuetan instalatuta, bide-sarea zeharkatzen du, azpiegituraren egoeraren auskultazio automatikoa eginez, eta,aldi berean, gidatzeko kritikoak diren gertaerak hautemanez. Ikerketa-lan honen kontextuan, hiru modulu ezberdin garatu nahi dira: 1) lainoa detektatzeko eta ikuspen-maila sailkatzeko sistema bat; 2) trafiko-seinaleak ezagutzeko sistema bat; 3) errepide-lerroetan akatsak detektatzeko sistema bat. Konponbide horri esker, zerbitzu kooperatiboak sortu ahal izango dira

Laburpena

errepideko gertaera kritikoak gainerako erabiltzaileei jakinarazteko. Era berean, aktiboen inbentarioa egiteko aukera emango du, azpiegituraren kudeatzaileei mantentze-lanen eta inbertsioen kudeaketa errazteko. Gainera, bide ematen dio gidatze autonomoari, egoera kritikoetan kontrol-trantsizioak hobeto kudea dituzan eta automatizazio-maila altuak dituzten ibilgailuak hartzeko azpiegitura prest jar dezan.

Acknowledgements



Hace ya siete años que comencé mi recorrido en Ceit, y no mucho menos que Alfon vió en aquella chica testaruda una futura doctora, incluso antes de tener los estudios que me permitieran siquiera inscribirme. Pero como él dice, a ese “no” sólo le hacían falta pequeños empujoncitos para labrar el camino al sí. Y aunque en muchas ocasiones no lo hubiera hecho, hoy me toca darle las gracias por todos esos empujoncitos que me han traído hasta aquí. Por toda la confianza depositada en mí en todo este recorrido para llevarme siempre más allá de lo posible.

Del mismo modo tengo un enorme gracias reservado para Diego, por aceptar dirigirme la tesis, acogerme en estos años, acompañarme en el camino y tener muchas veces la pregunta o frase acertada. Por abrirme las puertas al mundo de la visión artificial, por las discusiones técnicas, por los correos express sin asunto y por los ¿qué tal? rutinarios de primera hora, muchas veces oportunos.

Extender las gracias también a Ceit por confiar en esta tesis doctoral y preparar el camino para que alce su vuelo en el futuro. Y a Gertek por darme soporte con las imágenes y aterrizar la tesis en el mundo real.

En estos años, incluso en plena época de distanciamiento social, no puedo estar más agradecida de haber tenido la suerte de cruzarme con tantas personas dentro de esta particular “casa”. A los TMS, visionarios/robóticos, diseño, ferrocarril y otros con los que he compartido grupos burbuja y muchas conversaciones. Por aceptar mis idas y venidas de arriba a abajo y hacerme sentir acogida en todas ellas.

Y como no, gracias a mis mayores apoyos que han seguido esta tesis y todos sus altibajos muy de cerca:

A mi ama, asesora particular 24h. A mi aita, piloto y escuchador diario. A los dos por acompañarme en todas las decisiones, sean buenas, malas o regulares.

Acknowledgements

A Borja, por atreverse con la estación intermodal de trenes, coches, y montañas rusas. Por ser revisor, colaborador, público, tribunal, animador, equipo y hogar.

Y a Leire, mi bioquímica de cabecera, mi mitad y compañera de vida desde antes de nacer. Por estar antes que nadie, ante todo y para todo siempre.

Table of contents

Abstract	vii
Resumen	ix
Laburpena	xi
Acknowledgements	xiii
Table of contents	xv
List of figures	xix
List of tables	xxv
Glossary	xxix
Chapter 1 Introduction	1
1.1. Future mobility roadmap	3
1.1.1. Vision 2050.....	3
1.1.2. CCAM impact	4
1.1.3. Main challenges	4
1.1.4. Main enablers	6
1.2. Smart Road Classification.....	10
1.3. Environmental perception.....	12
1.3.1. Why computer vision?.....	13
1.3.2. Computer vision techniques	16
1.3.3. CCAM trends	18
1.4. Motivation.....	20
1.4.1. CCAM impact	20
1.4.2. CCAM limitations	20

Table of contents

1.4.3. Context.....	21
1.5. Objectives.....	24
1.6. Document structure.....	26
Chapter 2 Weather Conditions Monitoring: fog detection	27
2.1. State of the art.....	28
2.1.1. Classical computer vision techniques.....	28
2.1.2. Deep learning techniques.....	31
2.2. Materials and Methods	33
2.2.1. Existing Datasets.....	33
2.2.2. Our datasets	35
2.2.3. Developments	37
2.2.4. Evaluation parameters.....	45
2.3. Results	46
2.3.1. Rule-based method data visualization	46
2.3.2. Models comparison.....	47
2.4. Discussion	52
Chapter 3 Traffic Signs Monitoring: asset inventory.....	53
3.1. State of the art.....	54
3.1.1. Classical computer vision techniques.....	54
3.1.2. Deep learning techniques.....	55
3.2. Materials and methods	58
3.2.1. Existing datasets	58
3.2.2. Our dataset	59
3.2.3. Used datasets	61
3.2.4. Developments	61
3.2.5. Evaluation parameters.....	67
3.3. Results	68
3.3.1. Traffic Sign detection	69
3.3.2. Traffic Sign classification.....	72

3.4. Discussion	77
Chapter 4 Road Damage Monitoring: road lines	79
4.1. State of the art.....	80
4.1.1. Data adquisition systems.....	80
4.1.2. Road condition monitoring.....	81
4.2. Materials and Methods	86
4.2.1. Existing datasets	86
4.2.2. Used datasets	88
4.2.3. Developments	91
4.3. Results	99
4.3.1. Classical computer vision approach.....	99
4.3.2. Deep learning approach	101
4.4. Discussion	108
Chapter 5 Application in Connected, Cooperative and Automated Mobility	109
5.1. Context	110
5.1.1. ITS value chain.....	110
5.1.2. Data exchange via cooperative V2X communication.....	112
5.2. Message handling	115
5.2.1. C-ITS standard	115
5.2.2. DATEX II standard.....	122
5.3. Discussion	130
Chapter 6 Conclusions and Future Research Directions	131
6.1. Conclusions	132
6.2. Future research directions	135
Bibliography	137
Publications.....	155
Appendix	159
A. Appendix – Smart Road Classification framework	160

Table of contents

B. Appendix – Colour spaces.....	161
C. Appendix – TSR confusion matrixes.....	164
D. Appendix – C-ITS messages.....	168

List of figures

Figure 1-1. Machine Learning vs. Deep Learning feature extraction approach.	12
Figure 1-2. Artificial intelligence (AI) data journey or capabilities applied for the self-driving use case.	13
Figure 1-3. Camera view and LiDAR view of the KITTI validation dataset. Image extracted from [9].	16
Figure 1-4. Object detection and multi-class classification example.	17
Figure 1-5. Object segmentation example. Both semantic segmentation and instance segmentation are shown. In the second example, each car is differentiated.	17
Figure 1-6. Car identification and object tracking examples.	18
Figure 1-7. Data acquisition system installed in a road maintenance vehicle of Provincial Council of Bizkaia and managed by Gertek.	23
Figure 2-1. Ceit-Foggy dataset. Sample images showing the annotated four different classes. Three fog levels and the category containing negative images, this is, no fog.	35
Figure 2-2. Class distribution of the Ceit-Foggy dataset.	36
Figure 2-3. The five selected samples of attenuation coefficient per class for the construction of the Foggy Cityscapes DBF – extended dataset.	38
Figure 2-4. Implemented fog detection workflow for the rule-based method classifier.	40

List of figures

Figure 2-5. Studied XYZ features in three different weather scenes. From up to down sunny, cloudy and foggy sample scenes.....	42
Figure 2-6. Model comparison trained on ImageNet ILSVR2012 top-1 Accucary vs. Training Time. Image extracted from [54]	43
Figure 2-7. Transfer learning architecture designed for the new fog classifier	44
Figure 2-8. Image features representation for each ground truth label in the Ceit-Foggy dataset	46
Figure 2-9. Fog detection models' confusion matrix comparison for the Ceit-Foggy and Foggy Cityscapes DBF - extended datasets.	49
Figure 2-10. Fog tone comparison of the two datasets. Synthetic fog presents a colder colour than the real foggy scenes.....	50
Figure 2-11. Off-line video analysis and model comparison for a light-fog scenario with a moderate-fog bank that disappears in the last frames.....	51
Figure 3-1. Sample images for TSR challenges: lighting or weather conditions, artefacts, low-resolution signs, motion blur, rotation, occlusion, damage, inconsistencies and intra-class variation (The vast majority of the samples are extracted from Ceit-TSR dataset).	57
Figure 3-2. Ceit-TSR dataset. Sample images showing some of the challenging conditions: low contrast, fog, reflections, shadows, and heavy rain.	59
Figure 3-3. CVAT tool that shows the labelling task for Ceit-TSR dataset ..	61
Figure 3-4. Implemented Traffic Sign Recognition workflow.	62
Figure 3-5. ACF features. In the first row from left to right: original image, LUV channels, the gradient magnitude and individual representation of HOG features in different angles, of a sample sign.	63
Figure 3-6. Traffic sign samples of the ETSDB dataset. There are 164 different classes grouped into 9 categories: danger, regulatory (priority, prohibitory, mandatory and special regulation) informative (information, direction and additional panels) and others.	64

Figure 3-7. Traffic sign samples of GTSRB dataset. There are 43 different classes.....	65
Figure 3-8. The pre-processing phase before applying the corresponding classifiers. V channel normalization, mean image subtraction and random cropping tasks are presented.....	66
Figure 3-9. Example of false-negative detections due to the small size of the traffic sign. Table 3-4 shows detailed information.	70
Figure 3-10. Traffic sign classes were there is more missed samples than detected for GTSRB dataset.	71
Figure 3-11. Visual analysis of complex situations in Ceit-TSR dataset where the classifier fails. The left side shows the complete scene and the right side shows the cropped detection before and after pre-processing.....	75
Figure 4-1. Sensors and data acquisition platforms schema for Road Condition Monitoring (RCM). Image extracted from [113]	81
Figure 4-2. Sample of Road Damage Dataset 2019 (RDD2019) from (a) to (h) the nine different defects are represented.	88
Figure 4-3. Examples of Ceit Road Damage dataset in different light and weather conditions. Top right sunny with shadows, bottom right rain, top left sunrise and bottom left cloudy.....	89
Figure 4-4. Applied crop correction for keeping the resolution of the input image in Ceit Road Damage dataset.	91
Figure 4-5. Flow chart of the classical computer vision-based approach for road lanes quality assessment.	92
Figure 4-6. Machine learning-based lane condition assessment. Step 1, lane detection.	92
Figure 4-7. Machine learning-based lane condition assessment. Step 2, mask generation.....	93
Figure 4-8. Machine learning-based lane condition assessment. Step 3, day/night scene classifier.	93

List of figures

Figure 4-9. Machine learning-based lane condition assessment. Step 4, analysis of the mask's features.	94
Figure 4-10. Example of the applied six different transformations without combination for the data augmentation.	95
Figure 4-11. Visual results of the ML-based road paint damage detection	100
Figure 4-12. F1-score summary results for the models and datasets considered in this chapter.....	104
Figure 4-13. Comparison of the detection bounding boxes for a mixed and efficientdetv1_d0_spain (V1) pure model in CRDD dataset sample.....	105
Figure 4-14. Difficult labels of CRDD dataset where the detector has missed the defect.	106
Figure 4-15. Samples from the RDD2019 (D44) dataset.....	107
Figure 5-1. ITS value chain representation. It contains all involved actors and domains with its leading standards.	112
Figure 5-2. V2X communication ecosystem diagram plus the communication with the Traffic Control Centre.....	113
Figure 5-3. Table extracted from ISO 14823:2017(E) where general category codes are defined [154].....	117
Figure 5-4. Representation of the traffic sign that is codified in the IVIM message example.....	118
Figure 5-5. DATEX II standard's components schema. Extracted from [155].	123
Figure 5-6. Road marking damage detected in GI-636 exit branch road. .	127
Message 5-7. DATEX II message example for road in poor conditions event.	127
Figure A-1. Representation of (a) RGB and (b) HSV colour spaces.	162
Figure A-2. The CIE XYZ standard observer colour matching functions...	163
Figure A-3. Confusion matrix of ensemble classifier in Ceit-TSR dataset.	164

List of figures

Figure A-4. Confusion matrix of ensemble classifier in GTSRB dataset. ...165

Figure A-5. Confusion matrix of ensemble classifier in GTSDb dataset....166

Figure A-6. Confusion matrix of ensemble classifier in ETSDb dataset. ...167

List of tables

Table 1-1. Summary of the main future mobility challenges in different fields.	5
Table 2-1. Summary of the existing datasets for fog detection and its principal characteristics.	34
Table 2-2. Definition of fog classes following the definition of AFNOR norm NF P99-320. Road fog is differentiated into three different levels, light fog, moderate fog and dense fog	37
Table 2-3. The proposed rule-based method for fog detection and fog level estimation. This method analyses RGB and XYZ colour spaces.	40
Table 2-4. Foggy Cityscapes BDF-extended distribution for training, validation and testing of a deep learning-based classifier.	44
Table 2-5. Fog detection models' evaluation parameters comparison for the Ceit-Foggy and Foggy Cityscapes DBF - extended datasets.	48
Table 3-1. Summary of the publicly available TSR datasets.	60
Table 3-2. Summary of the used datasets for TSR in Chapter 3.	61
Table 3-3. Detection and classification results of the pure and ensemble models tested in the four different datasets used in this chapter.	68
Table 3-4. Ground truth of the example shown in the figure above. A sign is considered small when its width and height are below 15 pixels.	70
Table 3-5. Image size comparison between GTSRB detected traffic signs and missed traffic signs.....	71

List of tables

Table 4-1. Road damage types and definitions proposed by Maeda et al. [122]	83
Table 4-2. Table extracted from [122]. Detection and classification results for each class using the SSD Inception and SSD MobineNet. SIR: SSD Inception V2 Recall, SIP: SSD Inception V2 Precision, SIA: SSD Inception V2 Accucary, SMR: SSD Mobilenet Recall, SMP: SSD Mobilenet Precision, SMA: SSD Mobilenet Accuracy	84
Table 4-3. Existing publicly available datasets for road defects monitoring.	87
Table 4-4. Distribution of used datasets: RDD2019 for D44 defect, Ceit Damage Dataset and its simplified subset.	90
Table 4-5. Resulting used datasets' partition after data augmentation process.	96
Table 4-6. Summary of the properties of the studied different object detection models.....	97
Table 4-7. Results of the machine learning approach for road paint assessment in RDD2019 (D44) dataset and CRDD dataset.	99
Table 4-8. Summary of the hyperparameters used on the different object detection models trained in this section.	101
Table 4-9. F1-score and mAP (mean average precision) for the white line blur detection and each used dataset.	103
Table 5-1. Breakdown of the codes defining the speed limit sign according to ISO 14823:2017 (E) [154].	118
Table 5-2. Fog levels are defined by the DATEX II standard for PoorEnvironmentType events.	124
Table 5-3. Road surface condition definitions of DATEX II standard for Road Damage type events.....	126
Table A-1. Road categories in terms of automation and connectivity support [6].....	160

Table A-2. DENM message's cause and subcause codes summary that covers all possible use cases for Road Hazard Warning (RHW) service..... 168

Glossary



- ACDC** Adverse Conditions Dataset with Correspondences
- ACF** Aggregated Channel Features
- ADAS** Advanced Driving Assistance Systems
- ADS** Autonomous Driving System
- AI** Artificial Intelligence
- ANN** Artificial Neural Network
- AS** Assistedway
- AT** Automatedway
- AU** Autonomouway
- AUC** Area Under the Curve
- BDF** Dual-reference cross-Bilateral Filter
- BiFPN** Bi-directional Feature Pyramid Network
- CA** Content Aggregator
- CATERED** Carla Traffic Sign Recognition Dataset
- CAV** Connected Autonomous Vehicles
- CCAM** Connected Cooperative Autonomous Mobility
- CCTSD** Changsha university of science and technology Chinese Traffic Sign Detection benchmark
- CEDD** Colour and Edge Directivity Descriptor
- CFD** Crack Forest Dataset
- C-ITS** Cooperative Intelligent Transport System
- CNN** Convolutional Neural Network
- CRDD** Ceit Road Damage Dataset
- CRDDC** Crowdsensing-based Road Damage Detection Challenge
- CTSD (Carla)** Carla Traffic Sign Detection Dataset
- CTSD (Chin.)** Chinese Traffic Sign Dataset
- CTSD (Comp.)** Complex Traffic Sign Dataset

Glossary

CVAT	Computer Vision Annotation Tool
CVPR	Computer Vision and Pattern Recognition Conference
DCNN	Deep Convolutional Neural Network
DCT	Discrete Cosine Transform
DENM	Decentralized Environmental Notification
DITS	Dataset of Italian Traffic Signs
DL	Deep Learning
DNN	Deep Neural Network
DT	Digital Twin
EHD	Edge Histogram Descriptor
ERTRAC	European Road Transport Research Advisory Council
ETSD	European Traffic Sign Dataset
FA	Full Automatedway
FCTH	Fuzzy Color and Texture Histogram
FN	False Negative
FOH	Fuzzy Opponent Histogram
FOV	Field Of View
FRIDA	Foggy Road Image Database
FROSI	Foggy ROad Sign Images
GAN	Generative Adversarial Network
GAP	German Asphalt Pavement Distress
GIV	General IVI Container
GLC	General Location Container
GLCM	Gray Level Co-occurrence Matrix
GPU	Graphics Processing Unit
GTSDB	German Traffic Sign Detection Benchmark
GTSRB	German Traffic Sign Recognition Benchmark
HIS	Hue, Saturation and Intensity
HLM CAD	High Level Meeting on Connected and Automated Driving
HMI	Human Machine Interface
HOG	Histogram Oriented Gradients
HSV	Hue, Saturation and Value
HU	Humanway

HW	Hardware
ICF	Integral Channel Feature
IJCNN	International Joint Conference on Neural Networks
IoU	Intersection over Union
ISAD	Infrastructure Support for Automated Driving
IT	Information Technology
IVIM	Infrastructure to Vehicle Information message
IVS	In-Vehicle Signage
JHD	Joint Histogram Descriptor
KTSD	Korean Traffic Sign dataset
LiDAR	Laser Imaging Detection and Ranging
LOSAD	Level Of Service for Automated Driving
ML	Machine Learning
MOR	Meteorological Optical Range
MTSD	Mapillary Traffic Sign Dataset
NAP	National Access Point
NAS	Neural Architecture Search
ODD	Operational Design Domain
OOM	Out Of Memory
ORS	Operational Road Section
OTA	Over The Air
PID	Pavement Image Dataset
RCM	Road Condition Monitoring
RDD	Road Damage Dataset
RF	Random Forest
RGB	Read Green and Blue
RHW	Road Hazard Warning
RSU	Road Side Unit
RTSD	Russian traffic Sign Dataset
SAP	Service Access Point
SCD	Scalable Color Descriptor
SCH	Simple Color Histogram
SDD	Single Shot Detector

Glossary

SIFT	Scale Invariant Feature Transform
SP	Service Provider
SRIA	Strategic Research and Innovation Agenda
STM	Sustainable Transportation and Mobility
SURF	Speed Up Robust Features
SVM	Support Vector Machine
SW	Software
ToC	Transition of Control
TP	True Positive
TSR	Traffic Recognition System

Chapter 1

Introduction



Chapter 1 Introduction

In this introduction, the first section presents a vision of what the mobility of the future could look like and what its impact will be. It then looks at the current main challenges and enablers. This analysis is supported by the “Connected, Cooperative and Automated Mobility Roadmap” recently defined by the European Road Transport Research Advisory Council (ERTRAC) working group [1] and by the “Strategic Research and Innovation Agenda (SRIA)” established by the Connected, Cooperative and Automated Mobility (CCAM) European Partnership [2].

Since the thesis is especially focused on the perception of the environment employing computer vision techniques, the second section presents the justification of these techniques and selected sensors as well as a summary of the current trends.

Finally, the motivation and objectives of the thesis are presented.

1.1. Future mobility roadmap

Mobility is crossing a new – digital – frontier, allowing vehicles to communicate with each other, with the road infrastructure and with other road users. This will enable the coordination and cooperation between road users, and the management of traffic and mobility at an entirely new level (e.g. warning messages not limited by line-of-sight or congestion management using real-time information).

Current road vehicles already provide Advanced Driving Assistance Systems (ADAS) that can help the driver and take control of some functions in specific situations. Future systems will have a 360° vision of the surrounding environment, significantly reduced reaction times and will be able to control the vehicle for extended periods and, at some point in the future, will no longer rely on human intervention.

CCAM is expected to reshape the way we travel and move, not only in Europe but around the world.

1.1.1. Vision 2050

The community of researchers who are members of ERTRAC propose a long-term goal for CCAM [1] and have drawn a hypothetical scenario where in 2050:

Users and usage are at the centre of development where technology needs are derived from societal goals.

Automation domain costs decrease and mature technology allows working on complex scenarios such as inner city use cases allowing a more efficient last mile transport of people and goods.

Transport modes, road, rail, maritime or air have their role and cooperate to complement their high capacity, to decrease the overall environmental footprint of transportation.

Vehicles have 100% real-time connectivity and all newly registered vehicles will have automation at different levels depending on their nature.

- A vast majority of shuttles, buses and delivery vehicles in cities operate autonomously

- Nearly all vehicles on highways can operate without immediate driver intervention
- All cars and trucks on all roads have sophisticated supporting systems installed such as reaction on traffic lights etc. which contribute significantly to road safety

In 2050, **mixed traffic** exist in some areas where autonomous vehicles coexist with human-driven vehicles.

1.1.2. CCAM impact

The transformational change in mobility will have a huge impact on all road, traffic and driving situations.

The development of the CCAM should benefit all citizens. With the full integration of the CCAM into the transport system, the Partnership expected the following positive impacts on society:

- Safety: reducing road fatalities and accidents caused by human error;
- Environment: reducing transport emissions and congestion by optimising capacity, smoothing traffic flow and avoiding unnecessary journeys;
- Inclusion: ensuring inclusive mobility and access to goods for all.

This impact expected to be generated by the implementation of the CCAM is compromised by the current mobility challenges.

1.1.3. Main challenges

The next decades confront society with the need for fundamental mobility changes, with ambitious goals to meet climate and vision zero objectives.

Although the EU has made enormous progress in improving road safety and halved the number of fatalities on European roads since 2000, progress has stagnated and still 25000 people die on road every year and more than 135000 are seriously injured. EU's Road Safety Policy Framework 2021-2030 aims to reduce road fatalities and serious injuries by 50% by 2030 and make them disappear by 2050 (Vision Zero) [3]. In addition, today, road transport emissions represent around 25% of the EU's total greenhouse gas emissions,

Section 1.1 Future mobility roadmap

and these have increased over recent years. The European Green Deal established the goal of being climate-neutral by 2050 and achieving a 90% reduction in transport-related greenhouse gas emissions by 2050 [4].

To this end, multiple complex challenges need to be addressed and solved at a societal, human, technical, regulatory, economic and operational level (see Table 1-1).

Table 1-1. Summary of the main future mobility challenges in different fields.

Societal challenges	A society ready to accept, adopt and demand CCAM solutions.	Provide shared, on-demand and personalised transportation available to all.	Provide safe and reliable CCAM solutions with few Transition of Control (ToC) events.
Infrastructure challenges	Mature CCAM solutions for a wide market take-up	Handle all driving scenarios. Extend Operational Design Domains (ODDs) by extending Infrastructure Support for Automated Driving (ISAD).	Cooperation between vehicles' stakeholders, infrastructure and connectivity sectors for the pre-deployment developments.
Validation challenges	Innovation-friendly frameworks with effective and efficient ways to validate CCAM solutions.	Update the verification and validation systems to gather realistic and relevant test cases that will be constantly evolving.	Reflect the connectivity context in which CCAM systems will operate as it influences safety-critical functions. The complexity of emerging technologies and their safety criticality requires a huge amount of data for validation testing ¹ .

¹ It has been estimated by RAND corporation [157] that CAVs need to be driven 450 million failure-free kilometers (11,000 times around the world) to assure a similar rate of reliability (ie. 95% confidence) as existing human-driven cars. To gather this data with a fleet of 100 autonomous vehicles driving 24 hours a day, 365 days a year, at an average speed of 40km/h will take 12.5 years. And to demonstrate that their failure rate is 20% better than the human driver failure rate, 18 billion kilometers and 500 years will be necessary.

Chapter 1 Introduction

Policy challenges	Right legal framework at international, EU and national level to prevent patchwork regulations.	Collaborate on the compatibility of safety requirements, liability issues, communication systems and services.
AI challenges	AI-based predictive system state awareness.	Design of advanced levels of AI with reduced energy use and computational efforts.
Data challenges	A joint and acknowledged data exchange framework	Improve the quality of the datasets in order to represent all possible scenarios.

The main challenges facing the mobility of the future are the acceptance of CCAM solutions by society, which requires solving another important question: ensuring that these solutions are mature enough to be safe and reliable in any situation. This also implies other essential points such as the constant evolution and updating of validation systems and databases for all possible scenarios and new ones as they arise. In the same way, this will also require increasingly complex algorithms, but controlling the energy use and computational effort. And last but not least, the regulation and standardisation of these systems, data and communications so that CCAM solutions are global.

The technological challenges of automated driving are reliable sensing of the vehicle's environment using different sensors, localisation solutions so that the self-driving car knows exactly where it is, the development of a safe and secure driving strategy, safe driving strategy, data safety and security, including secure and resilient approaches for over-the-air updates, validation and verification of systems.

1.1.4. Main enablers

In order to address all these challenges, there are some essential enablers to take into account. To achieve global and implementable CCAM solutions that can change the mobility paradigm, there is a clear need to leverage the cooperation models of different types of data feeds in a suitable environment. In the same way, effective, profitable and transparent cooperation among local and regional public authorities and the private sector involving a multitude of highly diverse stakeholders is mandatory. A harmonised European framework is a key enabler for this cooperation, sharing and expanding knowledge on connected and automated driving. So is standardisation to address the many common and safety-critical use cases for Connected and Automated Vehicles (CAVs) and relevant infrastructure, communications, data management and privacy, cybersecurity and vehicle technology. In addition, a new hardware concept for sensors and computing units is key to provide sufficient computing power with low energy consumption, reduced size and affordable integration costs.

However, while all are critical, this thesis focuses on the enablers detailed below.

1.1.4.1. ODD and ISAD

According to the SAE definition ODD are “*Operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental , geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics*”[5]. These ODDs are very important for the automation levels since the limitations of the ODD decrease as the automation level increase reaching level 5 for unlimited ODD functioning. The environmental perception of automated vehicles is limited by the range and capability of onboard sensors, however, the impact of some factors potentially exceeding the ODD can be mitigated and even prevented by infrastructure-related actions.

For this aim, the ISAD classification has been proposed. Assigning ISAD levels to network sections provides the automated vehicles with information about the infrastructure support that can be expected.

ODDs will evolve along with the evolution of the onboard sensors, software and AI technologies. However the roadside infrastructure investment are very costly and hence, ISAD level evolution will likely focus on the digital infrastructure aspects and the physical infrastructure investments in “no-regret” measures.

1.1.4.2. Functional safety

Functional safety aims to protect the correct operating of the autonomous driving system in response to its inputs or failure in a fail-safe manner. For this aim, sensor fusion on the edge and Cooperative Intelligent Transport System (C-ITS) technologies are key.

Road infrastructure can provide and disseminate safety-relevant information about perceived objects, allowing vehicles to manage difficult scenarios without exceeding their ODDs. This would be of particular importance when weather, light or road surface conditions are adverse. External sensors communicate with the CAV, which in turn can process this information and fuse it with its perception thus gaining a better understanding to plan and act accordingly.

1.1.4.3. Data and AI

Data

The ability to generate new knowledge from large amounts of data is a key competence of future automated driving. A sovereign, open data infrastructure that observes security standards will thus become a key enabler for the successful development and deployment of CCAM.

Another key aspect is the collection of data from untypical or critical driving scenarios or corner cases. This requires several types of information like sensors and HD map data that will allow software improvements and updates. For this aim – learning during fleet operations - it is important to gather statistically representative driving data (field data) that will allow to validate, verify and improve driving behaviour.

In addition, the development of basic standards is also a key enabler to guarantee safe and reliable software in vehicles on roads. These basic standards should be developed in three main areas: for the access for stakeholders to collected driving data when critical driving behaviour; for

the definition of minimal implementation and update of safety-relevant software; and for the definition of a data exchange format.

AI

Autonomous driving is based on robust and reliable algorithms for environment recognition using different sensors. Due to its ability to recognise patterns and establish correlations in large amounts of data and learn them, Machine Learning (ML) is indispensable in the area of environmental recognition and driving assistance today.

However, the introduction of Deep Learning (DL) algorithms (a subsection of ML) is a key enabler to enable the vehicle to function correctly and make confident decisions even in the most difficult situations.

Accordingly, accepted metrics for the evaluation of AI algorithms are required. For this aim having relevant sets of real and synthetic scenarios are key to serve the metrics for safety proof and validate its robustness.

Finally, the importance of the AI training data specification, collection, processing and labelling should be stressed: the quality of the training dataset determines the robustness and thus operational reliability of the AI function.

Therefore, this thesis focuses on investigating different perception algorithms for road monitoring to collect safety-critical information to extend ISAD and ODD levels and influence functional safety. In addition, this information can also be used to classify the road and guide automated vehicles to the extent to which they can use their systems.

1.2. Smart Road Classification

There are several road classification systems in the literature each of them supported by different attributes, however, none of them establishes clear thresholds or KPIs of these attributes. In addition, none of the existing systems integrate dynamic conditions of the road – e.g. traffic volume or weather conditions – to determine the capability that a road has to enable automated driving.

Thus, the World Road Association PIARC made a Smart Road Classification (SRC) proposal based on automation and connectivity level indicators [6], this is, Level Of Service for Automated Driving (LOSAD) and Infrastructure Support Levels for Automated Driving (ISAD). LOSAD describes how vehicles' ODDs interact with the road infrastructure and ISAD indicates the connectivity capabilities, and both establish a sound basis to foresee how CAVs are likely to perform along a road network.

This SRC defines five different types of Smart Road segments that can be distinguished with specific characteristics related to CAVs (see Appendix A).

- (1) Humanway (HU): the road is not ready for CAVs. Level 2-3 vehicles would experience many disengagements, prompting their drivers to manually disconnect the system. And Level 4 vehicles may not find clear ODDs and will generally perform in manual mode.
- (2) Assistedway (AS): the road is adequate for Level 2+ vehicles. Level 2-3 vehicles would not induce too many disengagements and will allow drivers to enable their driving automation systems. However, the road segment might be divided into many Operational Road Sections (ORS)² and does not provide a comfortable automated driving experience for Level 4 vehicles.
- (3) Automatedway (AT): the road presents HD maps and can transmit digital information to CAVs, so these can better identify ODD-related factors and ODD terminals. Level 2 would experience fewer

² The zones that are ODD compliant with all vehicles are called Operational Road Sections (ORS).

disengagements than on AS segments and Level 3 vehicles would be able to use the digital information to foresee oncoming disengagements. Longer ORSs would allow longer performance of Level 4 vehicles in automated mode.

- (4) Full Automatedway (FA): the road presents a continuous ORS, so all Level 4 vehicles should be able to operate autonomously along the entire segment. Level 2-3 vehicles experience a much lower number of disengagements compared to AT.
- (5) Autonomousway (AU): the road presents similar physical conditions to FA segments and incorporated exceptional connectivity features that enable cooperative driving. In order to benefit from the best performance and safety levels, only Level 4+ vehicles should operate along these road facilities or dedicated lanes.

1.3. Environmental perception

With recent advances in AI, ML and DL, these techniques have gained prominence in numerous applications.

In a nutshell, AI is a broad term used to describe any system that can perform tasks that usually require human intelligence. This concept includes ML and DL. The principle difference between ML and DL is in the techniques of extracting the features on which the classifier works. ML classification relies on hand-crafted features while DL classifiers, which are considered a subset of ML, learn hidden patterns from data by themselves. Thus, these last ones are known for their ability to build much more efficient decision rules (see Figure 1-1).

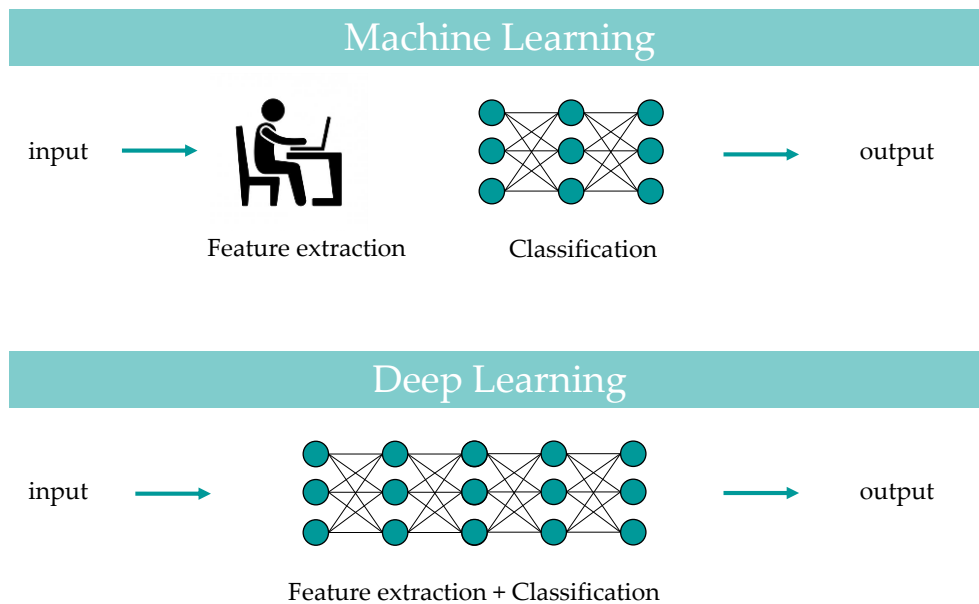


Figure 1-1. Machine Learning vs. Deep Learning feature extraction approach.

In particular, contemporary developments in communication networks and wireless connectivity, the arrival of accurate and robust sensors that continuously miniaturize in size and cost, coupled with AI have been the cornerstone for ADS. The availability of big data related to self-driving vehicles emphasizes the roles of ML and DL as it is infeasible to craft all

possible if-then-else rules that learn all possible situations a self-driving vehicle might encounter in the drive-terrain. These AI systems aim to implement sensory input and intelligent interaction with the environment. These AI capabilities may be divided into the areas of sensing, processing and understanding, decision-making and communication or acting (see Figure 1-2).

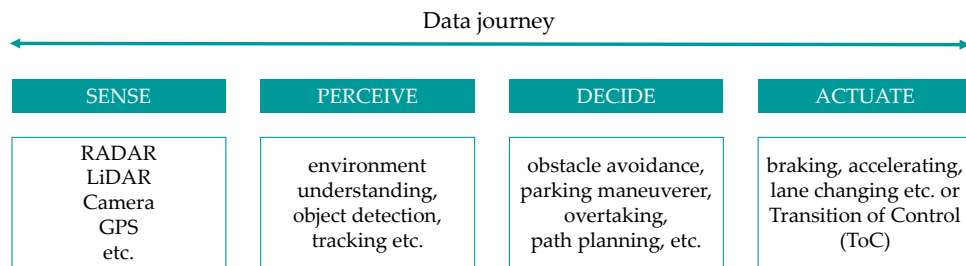


Figure 1-2. Artificial intelligence (AI) data journey or capabilities applied for the self-driving use case.

Sensing is the capability of observing the environment, this will give the capability to perceive an object that will be processed to understand the surroundings of the vehicles. With this data, the AI will be able to make decisions and actuate in consequence for a specific objective.

This thesis will be focused on the first two steps, to sense (using a single RGB camera) and perceive the road based on computer vision techniques.

1.3.1. Why computer vision?

What technology is needed to develop autonomous driving? There is a lot of controversy about this, pure vision-based systems, the fusion of different advanced sensors, etc. However, the renowned Tesla company stated that autonomous driving based solely on cameras is possible [7].

Currently, a manual driving control system is based on two main elements: the brain and the eyes of the driver, this is a biological neural network and vision system, and they are sufficient to understand the design of road networks. Thus, the big challenge for computer-guided autonomous driving

according to Tesla's CEO is to solve real-world AI and vision, this is, silicon neural networks and cameras ³.

This artificial intelligence is the basis for autonomous driving for road analysis, lane following, and detection of signs, pedestrians, cars or other objects. These AI systems require big data and the main sources of raw data in self-driving cars are the automotive sensors. Whereas Laser Imaging Detection and Ranging (LiDAR) is the most powerful camera, it is expensive and, in certain conditions, images captured using RGB cameras should be sufficient for self-driving applications.

1.3.1.1. LiDAR vs. RGB Camera

Cameras are optical sensors that offer functionality analogous to the human eye. The main purpose of CAVs is to mimic the human eye's ability to visually sense the environment. A single camera operating alone is called a mono-vision system, which results in a 2D image. This system can satisfactorily fulfil some CAV sensing requirements such as traffic sign detection. Nevertheless, when information about the distance to an object is necessary, a 3D stereo-vision system is used which is composed of two or more cameras to achieve depth perception. On the other hand, LiDAR is an active sensor that can generate high-resolution 3D maps of the vehicle's surroundings by emitting laser beams in all directions.

One of the main strengths of RGB cameras over LiDAR sensors is their relatively low cost. In the context of CAV hardware the camera's price range from multiple hundreds of euros (e.g. Blackfly) to even less than one hundred (e.g. Logitech webcam). However, research is underway to manufacture low-cost LiDAR but currently, their price is about thousand euros, and the cheapest one is between five hundred to one thousand euros [8].

Another characteristic of RGB cameras is that they can also sense colour information which is important for some road elements such as traffic lights. If the camera has a good resolution and the lens are kept in good condition,

³Testimony made by Elon Musk in TED | Tesla Texas Gigafactory interview.

cameras with colour detection capabilities can outperform LiDAR sensors in visual recognition since the last one needs much more data processing to create images and identify objects.

In terms of installability, the camera is generally easier and more versatile to embed in a vehicle while LiDAR requires more space and makes installation bulkier.

Although cameras have many strengths they have also limitations. They present some issues when dealing with environmental light variations such as shadows, sharp glare from the sun or dark views at night. Even though some of these problems could be solved by the vehicle's headlights, other environmental phenomena such as bad weather conditions - rain, snow or fog - are difficulties that cameras have to deal with in this field of application. In these cases, LiDAR can fill the gaps of camera-based systems since they have been hailed for being able to see objects even in bad weather conditions. However, it is also affected by wavelength stability and detector sensitivity and the laser can suffer some variations. In addition, to control the environment through which the vehicle circulates it is necessary to create a map previously with LiDAR, to insert in it all the characteristics of the road, lanes, intersections, signs, traffic lights etc. Generating and maintaining these high-definition maps with a precise location is a very complex task and doing it at a global level is really difficult (see Figure 1-3).

Thus, in summary, for public safety critical cases currently the sensor fusion of these two sensors together with others can offer many benefits. However, if only one of the two mentioned above could be chosen, for its versatility, ease of installation, lower computational complexity and price, the one chosen for the development of this thesis is the RGB camera. Moreover, cameras still present a gap for further research to achieve more reliable and robust systems that can perform well in any circumstances. And this will be one of the points to be addressed in this thesis.

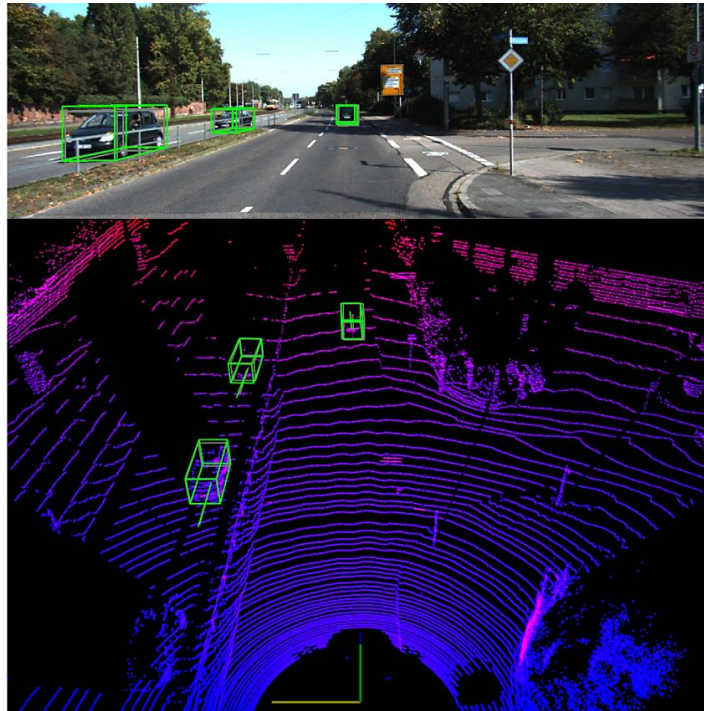


Figure 1-3. Camera view and LiDAR view of the KITTI validation dataset. Image extracted from [9].

1.3.2. Computer vision techniques

Classification, localization, detection, segmentation, tracking and identification are frequently used terms when talking about computer vision and intelligent applications.

- **Classification:** Assigning a category to a whole image or bounding box. This classification can be a single-class classification, where the output determines whether there is a car or not in the image or, a multi-class classification, where the output predicts the class between a sort of categories e.g. bicycle, truck, airplane, boat or a car (see Figure 1-4).
- **Localization and detection:** Identifying the location of a specific object in the image. Object detection finds all the objects and draws a rectangle around them (see Figure 1-4).

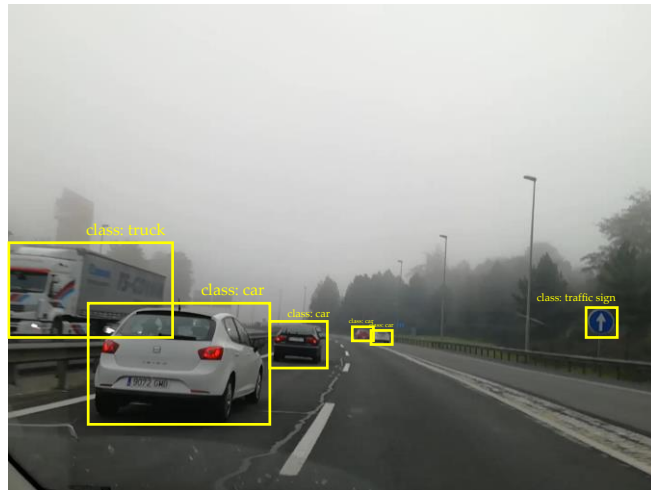


Figure 1-4. Object detection and multi-class classification example.

- Segmentation: Partitioning an image into a set of pixels that share certain characteristics and assigning them a label. In this case, object detection creates a pixel-wise mask and adds shape information of the object. This segmentation can treat multiple objects within a single category as one entity which is called semantic segmentation (Figure 1-5 a). Whereas instance segmentation, on the other hand, identifies individual objects within these categories (Figure 1-5 b).



(a) Semantic segmentation



(b) Instance segmentation

Figure 1-5. Object segmentation example. Both semantic segmentation and instance segmentation are shown. In the second example, each car is differentiated.

- Identification: Recognising an individual instance of an object. Examples include the identification of a specific person due to its face characteristics, identification of handwritten digits or the identification of a specific vehicle (see Figure 1-6 a).
- Tracking: taking an object or a set of object detections in an initial frame and re-identify them in the next frame (see Figure 1-6 b).

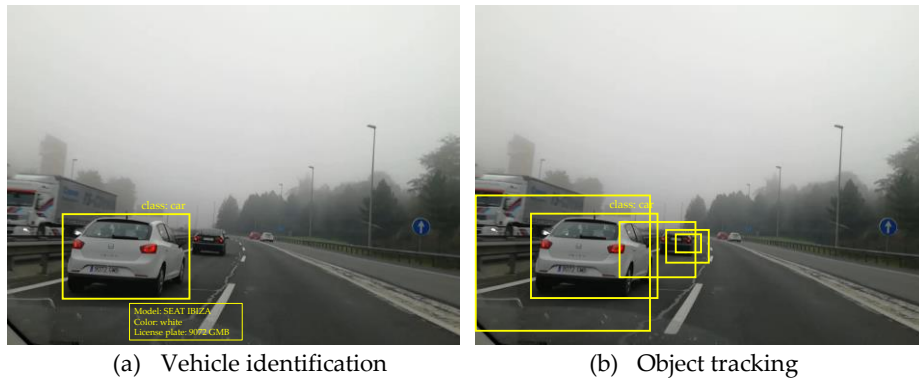


Figure 1-6. Car identification and object tracking examples.

1.3.3. CCAM trends

Recent trends and developments in environment perception of CAVs revealed that convolutional neural networks (CNN) are the most applied technique for object detection due to their remarkable ability to function as feature extractors. With GPU and cloud-based fast computation, DL could process captured information in real-time and communicate it to the nearby cloud and other vehicles in the meaningful vicinity. In order to improve the performance of these CNN-based models, transfer learning is frequently used by researchers.

However, in the context of self-driving cars, CNN-dependent strategies still need to be fine-tuned to achieve the precision level of the human eye and there is a huge scope for additional advancements in object detection and scene understanding. It is yet to be investigated when and under what conditions CNNs cease to perform well and can pose a threat to human life in self-driving scenarios. Much of the earlier tests were conducted on open

roads and in good weather, but more recent tests include adverse weather conditions such as fog.

On the other hand, artificial driving intelligence is still incapable to annotate and categorize the driving environment automatically without human assistance.

Last publications on the CAV topic conclude that self-driving cars are no longer a question of if but more of when and how[10]. The penetration rate of these self-driving cars into human society depends on their ability to drive safely. This puts forth a critical need for reliable computer vision techniques, mathematical models and simulations to mimic reality and arrive at the best parameters and configurations that can adapt to changes in surroundings.

Tools such as big-data, DL and CNNs give the possibility to achieve high levels of accuracy to solve perception problems of CAVs. These tools provide researchers with the ability to break complex problems into easier ones and previously impossible problems into solvable but slightly expensive ones such as capturing and annotating data to create the necessary ground truth.

1.4. Motivation

This section will introduce first the general impact of a road monitoring system on the road network as a CCAM solution. Next, some specific CCAM limitations will be highlighted which will be addressed during the thesis.

1.4.1. CCAM impact

This CCAM reality is still years away and it is not clear when it will come. Despite the great effort of the automotive industry during the last decade, the most advanced systems nowadays are SAE level 2 (3 at the most). Existing autonomous vehicles consist of diverse ADAS that allow an automated driving experience under specific circumstances. Therefore, how the human and the vehicle share the driving becomes very important.

To enable and promote a quick and reliable take-up of automated vehicles, user confidence in self-driving capabilities is key. A driver should not voluntarily decide whether to connect the assistance, but rather by having objective information about its operation along the segment to be travelled. This information must be provided by combining the characteristics of the driving automation system and the infrastructure (both static and dynamic).

An integral road classification system would also **enable efficient planning of public investments in physical infrastructure (I1)**, by improving the operability of driving automation, and in digital infrastructure, by increasing the benefits of connectivity (V2X). Information-sharing of all elements in the intelligent transportation environment – vehicles, road users, infrastructure, traffic, weather data and so on – will lead to a more accurate knowledge of traffic situation across the network, and consequently, end-users will be informed about the level of automation they can enable through each road segment. As a result, **a safer, sustainable and more comfortable road network (I2)** is expected.

1.4.2. CCAM limitations

Although vehicle technology is constantly enhancing and evolving, the sensors and algorithms included in CAVs cannot cope with specific challenging situations [6].

- Most Traffic Sign Recognition (TSR) systems recognize only posted speed limits and priority of way signs such as stop or yield. The performance of these systems depends on the position and orientation of the traffic sign and they are also influenced by maintenance status or lighting conditions.
- Despite the few studies quantifying the influence of pavement conditions on CAVs' performance, it is clear that the status of the road surface plays a critical role in automated driving. A good pavement condition is needed to achieve the highest levels of automation.
- Environmental factors such as weather or lighting also affect CAVs performance. Unfavourable weather conditions – heavy rain, fog, etc.– make road marking and traffic sign recognition very difficult as these tasks are performed with vision cameras that are very sensitive to visibility.
- Speed is also a critical factor in road marking or traffic sign recognition because information processing must be faster as the vehicle speed increases. Thus, real-time functioning systems are needed.
- Connectivity allows automated vehicles to monitor everything even beyond the range of their sensors. Thus, rules and regulations are key to ensuring the quality and reliability of the transmitted information.

All of these factors - infrastructure, environment, traffic conditions, vehicle speed - many of which are variable, influence the ODD of automated systems and generate road segments that are not suitable for automated vehicles. However, the goal to reach a high level of automation is to have all road ODD compliant with all vehicles. Therefore, this thesis addresses the above limitations in order to increase the ODD constraints of the vehicles as well as to provide valuable information on the areas that are not compliant with the road administrations and operators so that they can work actively towards improving them and adapting new sections.

1.4.3. Context

The development of this thesis has been carried out in collaboration with the Sustainable Transportation and Mobility (STM) group and the Intelligent

Systems for Industry 4.0 group at Ceit. Two of the centre's strategic lines of action converge here: Cooperative Services for Intelligent Transport Systems and Computer Vision for Smart Monitoring Systems. Those lines are perfectly aligned with the thesis' objective for the development of technologies for CCAM. In fact, cooperative services are defined for the correct integration of automated vehicles into the traffic environment. In addition, the STM group has experience in standard vehicle communications and has led several pilots regarding cooperative services in C-Mobile and C-Roads projects. Ceit also participates in the European CCAM initiative ensuring that developments are aligned with the European strategy on cooperative, connected and automated vehicles.

On the other hand, from the analysis of the challenges and enablers of future mobility, it becomes clear that continuous monitoring of the road to collect dynamic data and detect where critical situations occur is key towards CCAM.

In this context, Ceit has worked on several projects aimed at improving the quality of road information through the development of specific perception systems. These projects are based on the collection of real-time data employing an onboard acquisition system installed in administrations' dedicated vehicles – maintenance, police, cleaning, and other services - that are on the road during all working hours (see Figure 1-7).

The projects that have allowed the development of the thesis are:

“TRAFIK DATA: On-board module for monitoring road signs and weather conditions with co-operative services communications” industrial project in which Ceit worked together with the company Gertek, involved in urban and interurban traffic management and mobility tool provider. In this project, the SW and HW architecture of the embedded system was designed and the preliminary algorithms for the inventory of traffic signals and fog detection modules were developed. Additionally, a proprietary communication protocol was defined for direct communication with the traffic control centre.



Figure 1-7. Data acquisition system installed in a road maintenance vehicle of Provincial Council of Bizkaia and managed by Gertek.

“AUTOEV@L: Technological Evolution For Multivehicle Automation And Evaluation Of Highly Automated Driving Functions ” public funding project which, with the generated knowledge of the first project, allowed to improve previous developments with enriched data and new algorithms. A new module for road line marking condition monitoring was also developed within this project.

Both projects pursue the same ultimate motivation of improving road safety by taking steps toward cooperative driving where real-time information from the environment can help drivers make better decisions as well as anticipate and plan autonomous vehicles’ behaviour. Furthermore, the autonomous inspection tool developed in this thesis will also help to better manage maintenance tasks by prioritising resources and reducing costs.

1.5. Objectives

The future of mobility will be connected, cooperative and automated where perception and vehicle-to-everything (V2X) communication will have a key role to understand and interact with the environment. The reality of the road is complex and its complexity will increase as new transport systems emerge and vehicles become more automated. There have been great advances in the core technology that solve autonomous driving, however many perception algorithms still fail to work in critical driving conditions, this is when the ODDs are exceeded. The impact of this weakness can be mitigated and even prevented by providing a real-time twin of the physical and digital infrastructure, but sensorizing the whole road network is not affordable.

To this end, the hypothesis of this thesis is to study if it is possible to **monitor road's critical events and conditions by using specialized vehicles as sensors on the move**. These vehicles will have an in-vehicle image acquisition system for applying computer vision and deep learning techniques. Real-time monitoring will allow the collection of useful road information and the creation of C-ITS services that help road users to drive more safely and efficiently. In addition, this information will also be helpful for infrastructure managers who will have a real inventory that will enable them to manage tasks and investments. To achieve this, the following sub-objectives have been identified:

- Research different computer vision techniques for fog bank detection and visibility level classification to inform the driver or the autonomous vehicle control system and generate situational awareness for adverse weather conditions.
- Research different computer vision techniques for the recognition of vertical traffic signs to generate an asset inventory that can assist in the management of infrastructure maintenance.
- Research different computer vision techniques for road line marking damage detection to advise maintenance tasks to the competent authorities or to activate the ToC in case the automated vehicle is using a lane-keeping based system on a section of road that does not meet the necessary conditions to operate this function.

- Study the functional requirements needed for a future application of the solution as an embedded system that fulfils the following characteristics:
 - Modular, to facilitate the integration of new functionalities;
 - Compact, to facilitate the installation in different vehicles;
 - Low-cost, to facilitate accessibility and user acceptancy;
 - Real-time, to generate valid and useful information in time for other road users;
 - Common language, to facilitate communication between all stakeholders;
 - Standardized, to ease its development and implementation;
 - Interoperable, to facilitate data exchange;
 - Universal, to be applicable to roads worldwide and to all types of roads;
 - Useful, to facilitate the application by road administrations or road operators.

1.6. Document structure

The report of the thesis entitled “Computer Vision and Deep Learning based road monitoring towards a Connected, Cooperative and Automated Mobility” is composed of six chapters:

- Introduction
- Weather conditions Monitoring: fog detection
- Traffic Signs Monitoring: asset inventory
- Road Damage Monitoring: road lines
- Application in CCAM
- Conclusions and Future Research Directions

In this first chapter, an introduction has been given to the road mobility of the future, its impact, challenges and enablers as well as to the different alternatives for sensing and building perception systems.

In Chapter 2 the development of a fog bank detection system is presented. For this purpose, the state of the art for fog modelling and detection is first studied. This is followed by the development and validation of the results of the two algorithms developed for the resolution of this problem (rule-based and DL-based) and a comparison is carried out.

In Chapter 3 the development of a system for traffic sign recognition is presented. First, the state of the art is studied and the developed solution is detailed, which is divided into the detection and classification stages.


In Chapter 4 the development of a system for the detection of road marking damage is presented. The state of the art is first studied and then the developments are presented. Firstly, the initial algorithm based on classical computer vision techniques is presented, followed by the development of the final solution based on DL.

In Chapter 5 the importance of V2X communications in CCAM is highlighted. In addition, messages in DATEX II and C-ITS format are presented for the communication of fog bank events, traffic signal information as well as road marking damage warnings.

Finally, Chapter 6 presents the conclusions of this work and proposes future lines of research.

Chapter 2

Weather Conditions Monitoring: fog detection



Part of this chapter has been published in:

Iparraguirre, O.; Amundarain, A.; Brazalez, A.; Borro, D. "Sensors on the Move: Onboard Camera-Based Real-Time Traffic Alerts Paving the Way for Cooperative Roads". Sensors, Special Issue on Sensors for Road Vehicles of the Future, Vol.21, No. 4, 1254. February 2021. doi:10.3390/s21041254.

2.1. State of the art

The weather has a great influence on road conditions and therefore on driving. Poor conditions in which visibility is reduced hinder the driver's ability to drive cautiously and affect directly driving safety. Likewise, it also significantly affects camera-based intelligent driving systems as it has been shown that their performance in foggy scenarios is much more challenging than in clear weather scenes [11–13]. Therefore, current technologies focus on fog detection or visibility estimation to further work on image enhancement or even image restoration for fog removal.

Fog visibility estimation algorithms can be broadly classified into classical computer vision and deep learning-based approaches.

2.1.1. Classical computer vision techniques

The first classical computer vision methods for fog detection rely on the analysis and measurement of visibility distance by using image processing techniques such as edge, vanishing point or horizon line detection or region growing with a special focus on contrast and brightness study. Koschmieder was one of the first researchers that treated this phenomenon and proposed a simplified atmospheric scattering model that relies on the attenuation of brightness contrast by the atmosphere [14].

$$L = L_o e^{-\beta d} + L_f (1 - e^{-\beta d})$$

Where L is the total luminance reaching the observer from an object at distance d through a diffusing media such as fog with an extinction coefficient β (m^{-1}) and L_o is the luminance of the object at close range and L_f the luminance at the horizon. Thus, the first term is the light from the object itself while the second term is the diffuse light from the environment which has been scattered into the vision of the observer. In this way, the greater the distance to the object, the less light the subject will see from the object and more from the environment light, which will cause a decrease in the contrast between the object and the background.

This definition allows W. Middleton to formulate an attenuation law of atmospheric contrasts [15]:

$$C = \frac{L - L_f}{L_f} = C_o e^{-\beta d}$$

Where C is the contrast between the object and the background at distance d and C_o is the contrast at close range. To define standardized visibility, a black object against a white background is considered, which will have a C_o of 1. Furthermore, a minimum threshold contrast of 0.05 is assumed to be distinguishable for the human eye. Therefore, by solving the equation above an objective measure of the visibility distance or Meteorological Optical Range (MOR) is obtained, which is one of the most popular methods in automatic visibility measurement for ADS.

$$MOR = -\frac{\ln 0.05}{\beta} \approx \frac{2.996}{\beta}$$

This law fostered many other works among which the work carried out by Hautière *et al.* [16–21] should be highlighted. They do various refinements implementing dynamically Koschmieder’s law to extract the road and sky within the image, compute the weather conditions and restore its contrast, which can be considered the first approaches dedicated to transportation systems. They also created one of the most popular datasets called FRIDA (Foggy Road Image Database). Based on the same principles Negru *et al.* [22–24] also published various works to present an image dehazing method from a moving vehicle and advise drivers with the adequate speed for the detected fog density. The method considers an exponential decay in the foggy image and applies a median filter which increases the clarity of the reconstructed image.

Other leading research works on dehazing techniques were developed by Tan [25] that removes haze by maximizing the local contrast of the restored image and Fattal [26] who estimates the albedo of the scene and the medium transmission under the assumption that the transmission and the surface shading are locally uncorrelated. However, this may not be physically valid or fail in some cases. Later, He *et al.* [27] proposed a novel prior for single image haze removal, the dark channel prior, which is based on the statistics of outdoor haze-free images. They found that some local regions which do not cover the sky and have very low intensity (called dark pixels), can directly provide an accurate estimation of the haze transmission. With this haze thickness estimation, a high-quality haze-free image can be restored by the atmospheric scattering model. Yeh *et al.* [28,29] based on the same idea

improved the last method by estimating the atmospheric light using haze density analysis and using a bilateral filter to calculate and refine the transmission map. That results in better colour information and lower computational time compared to He's method. Huang *et al.* [30] also analysed this dark channel prior method and developed a new one to fix the problems related to halo effects, colour distortions and insufficient transmission map. This method based on depth estimation, colour analysis and visibility restoration outperformed other methods.

However, all these physical models lie on strong priors. Methods based on Koschmieder's law depend on several assumptions such as static and uniform atmosphere and flat and diffuse ground surface. In addition, they are valid only for daytime scenarios, making them unusable for real-life automotive applications where the system can work 24h/day. Methods based on dark channel prior are also based on the assumption that at least one of the colour channels has a very low intensity at some pixels.

There are also several methods in the literature based on the extraction of image characteristics for fog detection and classification. Some of these methods analyse these features to discern a final result while others use classical vision as a pre-processing step to subsequently compute a classifier (ML). For example, Pavlic *et al.* [31] analyse the power spectrum (squared magnitude of Fourier transform) of the image without considering any special information. This way, fog scenes that contain frequency components near-zero can be differentiated from much more high-frequency components that present the non-fog scenes. Later on, the author extended the method for night conditions [32], however, it still fails to detect clear weather when high contrast elements appear in the image (eg. oncoming vehicles, overtaking trucks or bridges) in a foggy scene. Spinneker *et al.* [33] also focused their work on frequency characteristics and analysed the power spectrum slope around the vanishing point to obtain a visibility range. Other researchers studied the fog looking for its colour characteristics. Asery *et al.* [34] analyse the image Gray Level Co-occurrence Matrix (GLCM) features and considered its contrast, correlation and homogeneity to construct a Support Vector Machine (SVM) classifier that will differentiate foggy and non-foggy images, but the background of the image seems to have great influence when applied to natural images. In [35] Alami *et al.* proposed a fog detection algorithm based on the analysis of correlation and saturation

characteristics in RGB colour space. This analysis was focused on the vanishing point which is calculated using an edge-based algorithm and Hough Transform. Li *et al.* [36] presented a fog level detection method based on grayscale features where they measure the average grey value of each row in the image and classify them into non-foggy, little-foggy or dense-fog analysing the slope of the calculated curves. And Liu *et al.* [37] addressed the fog level detection method based on image information characteristics like H (hue), S(saturation) and V(value) which reduces the number of judgment threshold value demand and thus detection complexity. Finally authors in [38] made a comparison between seven histogram-based methods: Color and Edge Directivity Descriptor (CEDD), Edge Histogram Descriptor (EHD), Fuzzy Color and Texture Histogram (FCTH), Fuzzy Opponent Histogram (FOH), Joint Histogram Descriptor (JHD), Scalable Color Descriptor (SCD), and Simple Color Histogram (SCH). They used SVM classifier for no fog, light fog and heavy fog scenarios, and observed that JHD and FCTH had the best performances.

2.1.2. Deep learning techniques

On the other hand, Neural Networks are also used for fog detection. This Machine Learning method has been on the rise in recent years due to its ability to solve complex non-linear functions. In the image classification area, Deep Neural Networks (DNN) are mostly used which analyse the global features of the image. Chaabani *et al.* [39] proposed a three-layer neural network with a global feature descriptor based on Fourier transform that captures the power spectrum of the image to learn six different visibility ranges between 60-250 meters. The achieved classification rate was 90.2% tested on synthetic images from Foggy Road Sign Images (FROSI) dataset [40]. Later on, Palvanov *et al.* proposed VisNet, an approach based on deep integrated CNNs for the estimation of visibility distances from camera imagery. The implemented networks use three streams of DCNN connected in parallel and pre-processed the input image by applying one filter in the frequency domain and another spectral filter for the extraction of low-contrast regions. This approach achieved 94% on FROSI dataset images. Later on, Vaibhav *et al.* [41] develop a CNN with two image inputs, one original and the second input a block-wise discrete cosine transform (DCT) and Shannon entropy features. This model classifies three visibility ranges

Chapter 2 Road Damage Monitoring: road lines

less than 50 m, 50-150 m and 150 and above and obtains a 94.51 % accuracy for the FROSI dataset. Both Palvanov and Vaibhav tested their algorithms also on real scene images, which decreased the performance by 4.5% and 7.06% points respectively, however, none of these datasets is publicly available. In summary, the performance of the current state-of-the-art algorithms in real driving scenarios is still inconclusive.

2.2. Materials and Methods

In this section an analysis of the existing databases is made and the ones chosen or created for the development of this thesis are presented. Subsequently, the different methods studied for fog detection are explained.

2.2.1. Existing Datasets

Although there are large-scale road datasets such as KITTI [42], Cityscapes [43], Mapillary Vistas [44], ApolloScape [45] and BDD100k [46] the availability of useful image datasets for the foggy road scenes evaluation is very low. Most of the existing datasets contain few or even no foggy scenes due to the difficulty of collecting and annotating them. For example, the Mapillary Vistas database contains 10 images out of 25000 misty images (not dense fog). Thus, some of the existing foggy datasets are generated synthetic images or real-world images post-processed with synthetic fog. See Table 2-1 for a summary.

Foggy Road Image Dataset (FRIDA) is the most popular one, which was created by Hautière *et al.* with synthetic images [20] and was later extended [21]. This last image set presents 66 diverse road scenes that have associated 4 different fog types (no fog, uniform fog, heterogeneous fog, cloudy fog, cloudy heterogeneous fog) which comprises a total of 330 synthetic images.

The Foggy Road Sign Images (FROSI) dataset was introduced by Belaroussi *et al.* [40] and contains a set of 504 base synthetic images 1400x600 with different road signs placed at different distances on the image. For each image 7 types of uniform synthetic fog are applied with visibility ranging from 50 m to 400 m. Therefore FROSI set is made of a total of 3528 images.

A more recent dataset derivated from the Cityscapes dataset [43] was generated by Sakaridis *et al.* which is called Foggy Cityscapes [47]. It constitutes a collection of 25000 images from the original dataset that are processed and automatically annotated into 3 foggy levels using a fog simulator ranging visibility of 600, 300 and 150 m. Later, an improved version of this dataset called Foggy Cityscapes –DBF (Dual-reference cross-Bilateral Filter) [47] was published which additionally uses both colour and semantics as reference for the transmittance map refinement and comprises 3475 synthetic foggy images with better adherence to semantic boundaries

Chapter 2 Road Damage Monitoring: road lines

in the scene than the latter dataset. The same authors also generated two new datasets with real-world foggy road scenes called Foggy Driving and Foggy Zurich. The Foggy Driving dataset contains 101 light fog images captured with a cell phone camera at different points of Zurich and also with images collected from the web. Later on, they extended this dataset with collected video frames of the same city and its suburbs improving the resolution and having much variety of scenes and different fog levels, this last dataset contains 3808 images and is named Foggy Zurich [48]. Finally, they published the Adverse Conditions Dataset with Correspondences (ACDC) [49] with the aim of training and testing semantic segmentation methods on adverse visual conditions. ACDC consists of a large set of 4006 annotated images, containing fog, nighttime, rain and snow scenarios equally distributed.

Table 2-1. Summary of the existing datasets for fog detection and its principal characteristics.

	Number of images	Classes	Fog levels	Resolution	Country	Synthetic scenarios	Publication year
FRIDA1	90	--	4+1	640×480	--	yes	2010
FRIDA2	330	--	4+1	640×480	--	yes	2012
FROSI	3528	--	7	1400×600	--	yes	2014
Foggy Cityscapes	20000	8	3+1	2040×1016	Germany and Switzerland	synthetic fog	2018
Foggy Cityscapes DBF	550	8	3+1	2040×1016	Germany and Switzerland	synthetic fog	2018
Foggy Driving	101	19	--	960×1280	Zurich	no	2018
Foggy Zurich	3808 (40 annotated)	19	--	1920×1080	Zurich	no	2019
Seeing Through Fog	1429060	--	--	1920×1024	Germany, Sweden, Denmark, and Finland	no	2020
Ceit-Foggy	4480	--	3+2	several	Spain	no	2020
Foggy Cityscapes DBF extended	11000	--	4+1	2040×1016	Germany and Switzerland	synthetic fog	2021

Finally, the SeeingThroughFog dataset [50] was developed in the context of DENSE project. It records 10,000 km of driving in Northern Europe under different weather and illumination conditions. It contains 12000 samples in real-world driving scenes and 1500 samples in controlled weather conditions within a fog chamber. The resulting annotations contain 5,5k clear weather frames, 1k captures in dense fog, 1k captures in light fog, and 4k captures in snow/rain.

2.2.2. Our datasets

2.2.2.1. Ceit-Foggy

Ceit-Foggy dataset consists of a set of 41 videos corresponding to approximately 300 km of driving through the Basque Country and Segovia. These videos were recorded in different weather conditions as shown in Figure 2-1.

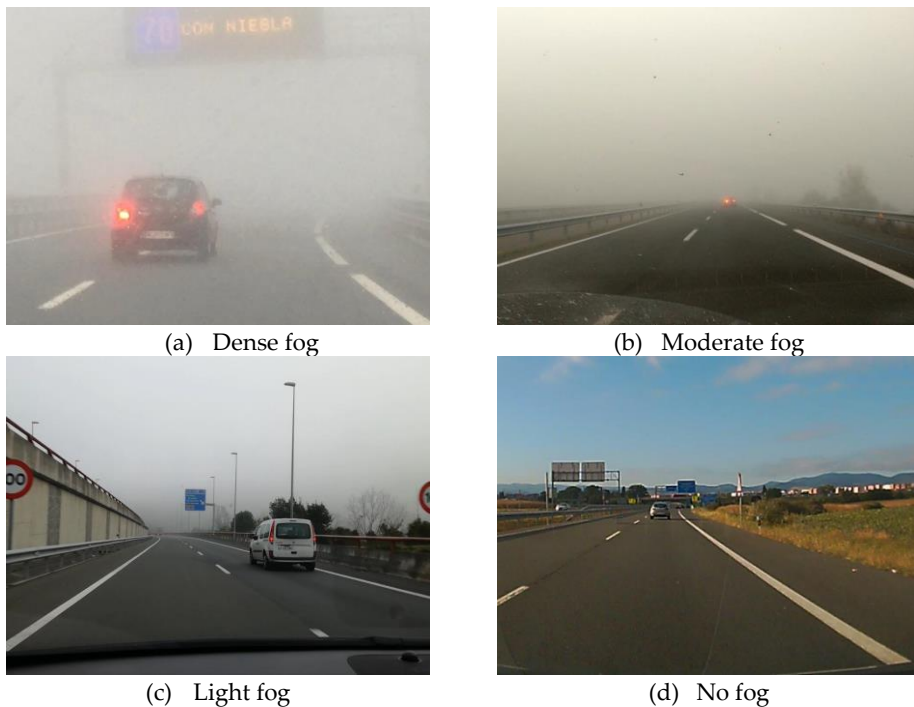


Figure 2-1. Ceit-Foggy dataset. Sample images showing the annotated four different classes. Three fog levels and the category containing negative images, this is, no fog.

Around 4000 frames were extracted and of these 1681 images were labelled in 4 categories: no fog, light fog, moderate fog and dense fog. These images were recorded using different mobile phones and onboard cameras situated on the dashboard. The images were subsequently cropped to avoid showing the interior of the car. Therefore, the resolution of the images may vary from 549×411 to 1441×1080.

Due to the difficulty of finding this meteorological phenomenon the distribution of the classes is not balanced as it is shown in Figure 2-2. Although this imbalance is not appropriate for training purposes, it could be used for model evaluation as it is representative of the samples that would occur in natural driving.

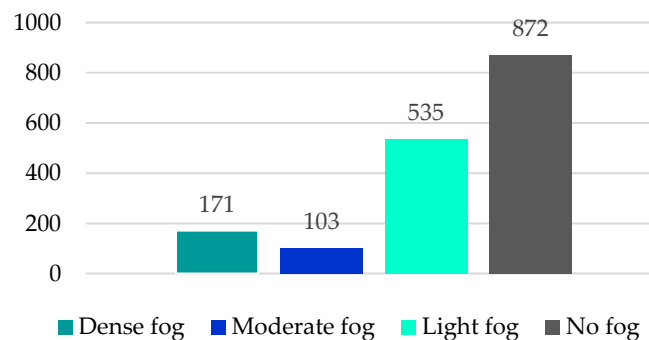


Figure 2-2. Class distribution of the Ceit-Foggy dataset.

2.2.2.2. Foggy Cityscapes DBF – extended

This dataset was generated from the Foggy Cityscapes DBF dataset to extend the number of samples associated with different visibility ranges. This refined version has 550 original real-world images and applies a synthetic fog based on the standard optical model of Koschmieder [47][51]. The dataset characterises three fog levels with an attenuation coefficient β of 0.005, 0.01 and 0.02 m^{-1} corresponding to visibility ranges of 600, 300 and 150 m.

However, following the work done in DENSE project, it was found more convenient to follow the visibility ranges defined by NF P 99-320 norm for road fog [52]. It is worth noting that due to vehicle speed and depending on the availability of contrasted objects along the road the distance to the furthest visible object can be different to the general visibility distance. Thus, this norm specifies that road fog has a lower threshold of visibility than

meteorological fog. These ranges are specified in Table 2-2. This thesis is focused on the analysis of road fog and considers the meteorological fog as no foggy scene. A slight adaptation is applied by merging the last two road fog visibility levels to reduce road fog levels to three for comparative reasons with the Ceit-Foggy dataset. The attenuation coefficient for each class is inferred from Koschmieder's law (see section 2.1.1) as in the paper of the original dataset.

Table 2-2. Definition of fog classes following the definition of AFNOR norm NF P99-320. Road fog is differentiated into three different levels, light fog, moderate fog and dense fog

	Road visibility level	Visibility distance (m)	Attenuation coefficient (m^{-1})	Number of images
	Meteorological fog	< 1000	[0.007 - 0.0003]	2975
Road fog	light fog	200 to 400 100 to 200	[0.03 - 0.007]	2750
	moderate fog	50 to 100	[0.06 - 0.03]	2750
	dense fog	< 50	$[\infty - 0.06]$	2750

To construct a balanced dataset, 2750 samples were generated for the four different visibility ranges: light fog, moderate fog, dense fog and no fog (see Figure 2-3). Thus, finally, the dataset contains 11,225 images in total of size 2048×1024 .

2.2.3. Developments

As it was explained in the state-of-the-art revision there are two main approaches for the detection of foggy scenes based on vision techniques: (1) measurement of the visibility range and (2) extraction of image characteristics. The first approach was discarded for this application because there is no direct relation to the fog's physical properties since several factors affect it such as background light, road curvature, presence of contrasted objects etc. [53]. The high complexity of this problem could lead to the study of a solution with neural networks. However, this option could be

problematic if the onboard system has hardware limitations. Thus, in this thesis, a classic computer vision approach is explored in Section 2.2.3.1 which can be less demanding in terms of computation, although, later a deep learning-based algorithm is also addressed in Section 2.2.3.2.

Category	MOR (attenuation coefficient)				
	5m ($\beta=0.599$)	15m ($\beta=0.200$)	25m ($\beta=0.120$)	35m ($\beta=0.086$)	45m ($\beta=0.067$)
DENSE					
	55m ($\beta=0.054$)	65m ($\beta=0.046$)	75m ($\beta=0.040$)	85m ($\beta=0.035$)	95m ($\beta=0.032$)
MODERATE					
	145m ($\beta=0.021$)	195m ($\beta=0.015$)	245m ($\beta=0.012$)	295m ($\beta=0.010$)	345m ($\beta=0.009$)
LIGHT					
	475m ($\beta=0.006$)	605m ($\beta=0.005$)	735m ($\beta=0.004$)	865m ($\beta=0.0035$)	995m ($\beta=0.0030$)
METEO. FOG					

Figure 2-3. The five selected samples of attenuation coefficient per class for the construction of the Foggy Cityscapes DBF – extended dataset.

It is worth mentioning that the hardware used in these developments is a Windows 10 PC with an Intel Core i7 processor, NVIDIA GeForce RTX 3080 10 GB GPU and a total RAM of 32 GB. The software used for the classical computer vision approach was Matlab 2020a (Image Processing Toolbox V11.1 & Computer Vision Toolbox v9.2) while for the deep learning approach Python 3.7.13, Tensorflow V2.8.2 and keras.applications were employed.

2.2.3.1. Classical computer vision approach

In this section, a classical computer vision approach is applied where the colour thresholding technique is used to model and detect fog characteristics. Although, most of the previous works have analysed grayscale images; our work aims to study whether other colour spaces could help to get more information on the image and help to improve the results.

After several experimental studies in the RGB and HSV colour spaces, it was concluded that this information was not sufficient to properly differentiate between cloudy and foggy scenes. The thresholding was highly complex and failed to generalize the different cases presented in the reference images. Therefore, it appealed to a not-so-popular colour space, the XYZ which defines quantitative links between distributions of wavelengths in the electromagnetic visible spectrum (see Appendix B). This new colour space allows us to define a rule-based method from scratch that can classify sunny, cloudy and foggy scenes by using XYZ features. Afterwards, the designed algorithm estimates the fog level of the foggy images by using RGB features (see Figure 2-4).

Thus, our work extracts the specific features of the images and establishes several rules to classify scenes as sunny, cloudy and foggy. These rules are summarised in Table 2-3 and they will be presented in detail in the following lines.

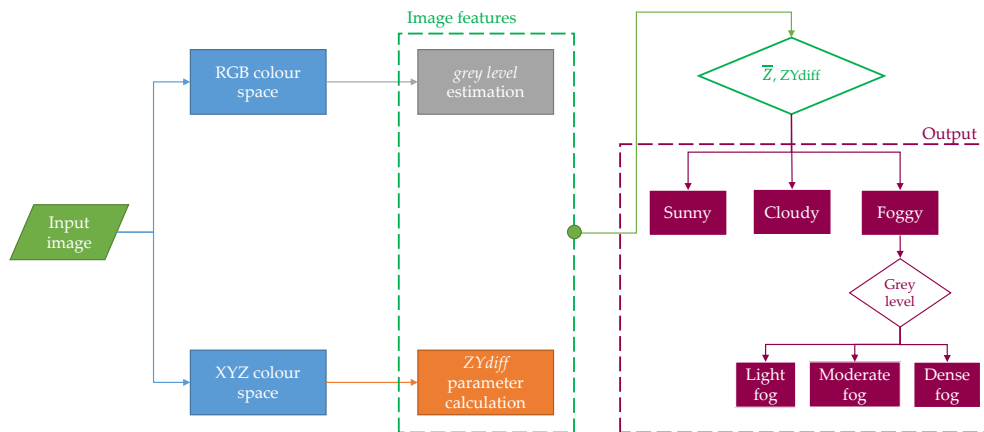


Figure 2-4. Implemented fog detection workflow for the rule-based method classifier.

Table 2-3. The proposed rule-based method for fog detection and fog level estimation. This method analyses RGB and XYZ colour spaces.

Sunny	$Z > 0.35 \ \&\& \ ZYdiff > 0.1$		
Cloudy	$Z < 0.35$		
		Light	<i>greylevel</i> [10-30]
Foggy	$Z > 0.35 \ \&\& \ ZYdiff < 0.1$	Moderate	<i>greylevel</i> [30-60]
		Dense	<i>greylevel</i> [60-100]

First, the XYZ colour space was analysed, here the Z channel will find cloudy scenes and the parameter defined as *ZYdiff* will differentiate between foggy and sunny scenes. Second, once the foggy scene is detected, our algorithm will classify the foggy scenes into light fog, moderate fog and dense fog by using the RGB colour space-based features. One of the main characteristics of fog is that it blocks visibility from a certain distance. This causes a decrease in the contrast between the object and its background so that the scene takes on a white/grey colour. The calculated grey level will provide an estimation of how dense the fog is.

It is worth mentioning that for these analyses, just pixels from the upper half image will be considered. In this portion of the image, it will be mostly the sky after having previously calibrated the camera position.

Z and ZYdiff calculation

In the XYZ colour space, the Z average level of the pixels located on the upper half of the image (referred to as \bar{Z}) was analysed. It is observed that this channel makes a difference between cloudy scenes and the rest since the \bar{Z} value is lower in cloudy scenes ($\bar{Z} < 0.35$).

However, this characteristic presents similar values both for sunny and foggy scenes. Therefore, a further feature is calculated by the $\frac{|Z-Y|}{Y}$ formula, which represents the difference of the Z and Y channels averages with respect to luminance (Y); hereinafter referred to as *ZYdiff*. This value is not relevant for cloudy situations but leads us to differentiate between sunny and foggy scenes as can be seen in Figure 2-5. *ZYdiff* is higher in sunny scenes ($ZYdiff > 0.1$) than in foggy ones ($ZYdiff < 0.1$).

Grey level estimation

The grey level allows approximating how much contrast has been lost in the image due to fog. This feature has been extracted by establishing several rules for the RGB channels. Firstly, it was considered as grey pixels those RGB values enclosed in the (140-255) range. This range represents bright pixel values. Additionally, a limitation was established for the difference between each channel to 20, this rule will ensure that the saturation of the pixel is low.

Thus, this grey level would be the percentage of pixels that meet these conditions compared to the total number of pixels analysed on the upper half of the image.

The calculated *greylevel* seems to be a good representative of the fog level. Thus, based on experimental tests three thresholds that will conform to the three different fog levels were defined. This way, the light fog scene is expected to have 10-30% of grey pixels, moderate fog conditions will oscillate between 30-60% and an image with more than 60% grey pixels will be considered a dense fog scenario

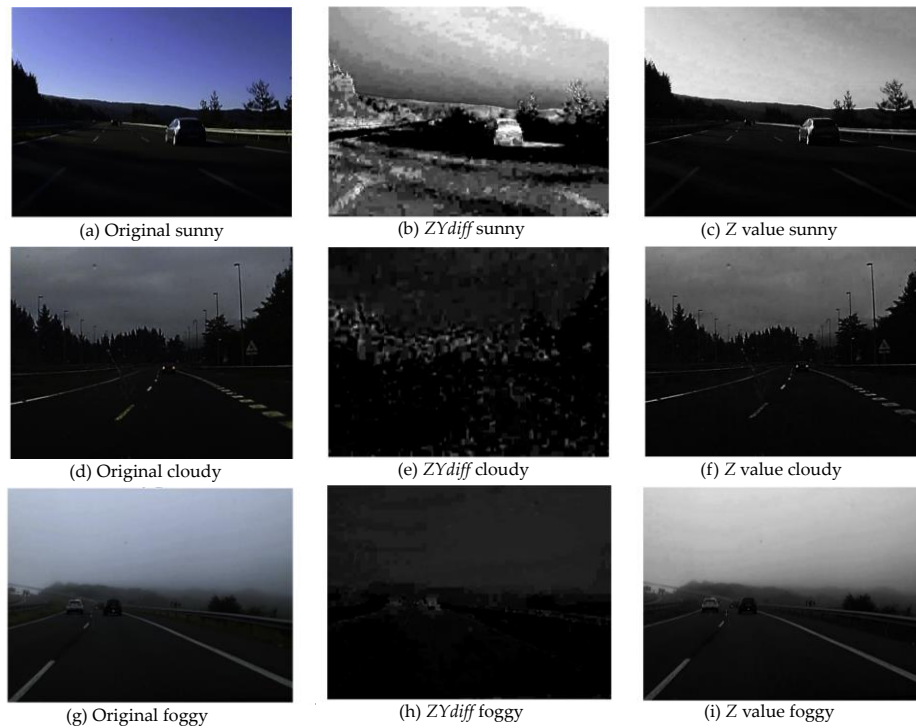


Figure 2-5. Studied XYZ features in three different weather scenes. From up to down sunny, cloudy and foggy sample scenes.

In this manner, the fog detection matrix presented in Table 2-3 is finally constructed.

2.2.3.2. Deep Learning-based approach

Due to the high complexity of the problem, currently, many works are focused on solving it by applying DL techniques because of their ability to solve complex non-linear functions. Therefore, in this section, the approach to construct a neural network-based model is presented. This model can deal with different scenarios and classify three different levels of road fog depending on the visibility range.

Network architecture

For this approach EfficientNetV2 [54] is introduced, a new family of CNN that presents a faster training speed and better parameter efficiency than previous models while being up to 6.8x smaller (see Figure 2-6). This model

increases both training speed and parameter efficiency by using training-aware Neural Architecture Search (NAS) and scaling the image size, however, this technique often causes a drop in accuracy. To solve this inconvenience, and achieve both fast training as well as good accuracy, EfficientNetV2 proposes to adaptively adjust regularization along with the image size by a dropout or data augmentation. This technique is called progressive learning which jointly increases image size and regularization during training.

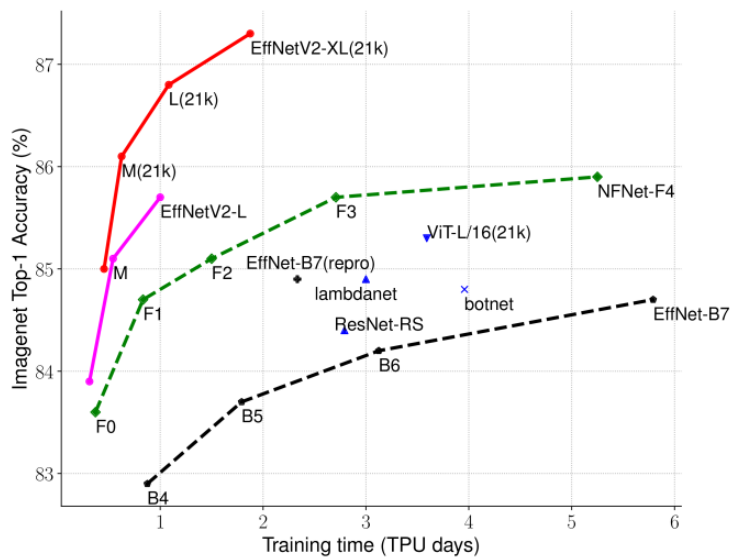


Figure 2-6. Model comparison trained on ImageNet ILSVR2012 top-1 Accuracy vs. Training Time. Image extracted from [54]

Model setup

The selected architecture is EfficientNetV2 B0 since it is the most lightweight one, initialize the model with the pre-trained ImageNet weights and fine-tune it to apply transfer learning to our target domain. The dataset used for this training is the Foggy Cityscapes BDF-extended that has been partitioned in 80% train 10% test and 10% validation (see Table 2-4). The experiments in this Section were carried out using Keras API. To use this model the original images were re-scaled to the shape $224 \times 224 \times 3$ and stored as TFRecords (binary records) so that they can be read efficiently.

Chapter 2 Road Damage Monitoring: road lines

Table 2-4. Foggy Cityscapes BDF-extended distribution for training, validation and testing of a deep learning-based classifier.

Foggy Cityscapes BDF-extended	
Train (80%)	9353
Valid (10%)	936
Test (10%)	936
11225	

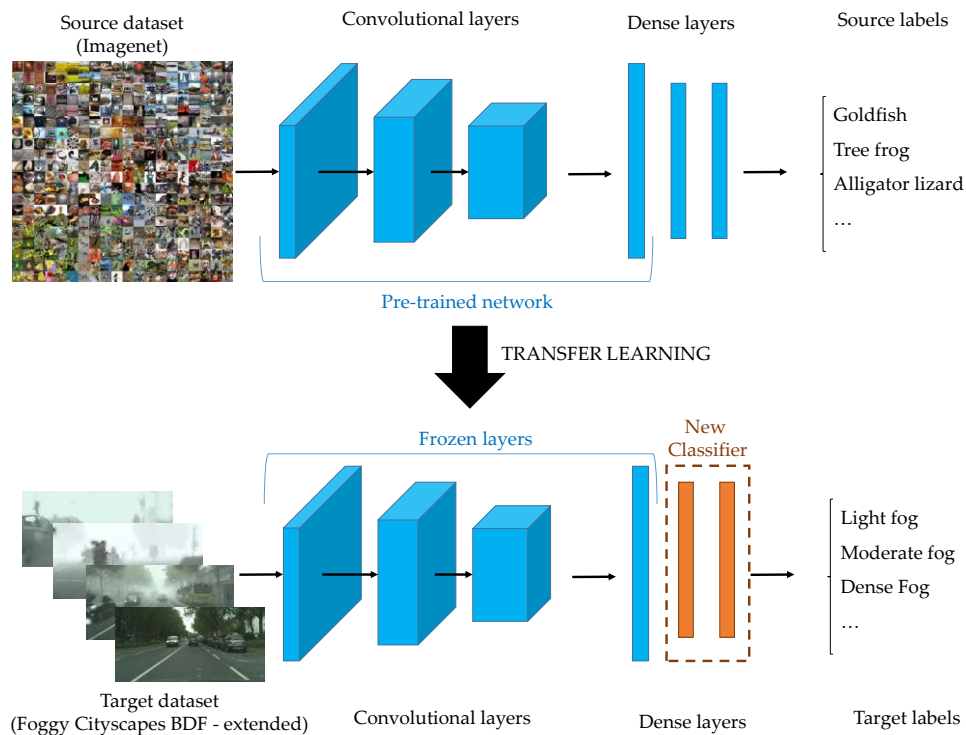


Figure 2-7. Transfer learning architecture designed for the new fog classifier

The first step to transfer learning was freezing all layers so as to avoid destroying any of the information the pre-trained model contains during future rounds. Then add some new trainable layers that will learn to turn the old features into predictions on the new dataset (see Figure 2-7). For this step, it is selected a relatively large learning rate of 0.001 and set 100 epochs with

an early stopping callback to monitor the *val_loss*. The next step was to unfreeze part of the model (the top 155 layers) and reduce the learning rate to $1e^{-6}$ for another 60 epochs followed by another 40 epochs with a learning rate of $1e^{-8}$. This fine-tuning phase where part of the model is re-trained can potentially achieve meaningful improvements by incrementally adapting the pertained features to the new data.

2.2.4. Evaluation parameters

For the evaluation of the multiclass classification model's performance, accuracy metric was employed that calculates how often predictions match the one-hot label. Thus, the accuracy is defined as:

$$Accuracy = \frac{Matched\ predictions}{Total\ predictions}$$

Inference times have also been measured to assess their real-time performance

Additionally, the confusion matrix is also generated to illustrate the results of the classifier for each class, which allows to analyse what are the most difficult scenarios.

For the comparison of the two different models presented above, it is used the Foggy Cityscapes DBF-extended validation partition that contains 936 images and the whole Ceit-Foggy dataset with 1681 images. Both of them have 4 categories: light fog, moderate fog, dense fog and no fog.

2.3. Results

In this section, it is first illustrated an offline analysis of the rule-based method to illustrate how the defined thresholds fit the datasets. Then the results of the two models are presented and discussed.

2.3.1. Rule-based method data visualization

This validation was executed for the Ceit-Foggy dataset in order to check whether the defined thresholds for the rule-based method were representative. Thus, the algorithm was executed and all parameters were recorded and drawn in the following figure (see Figure 2-8).

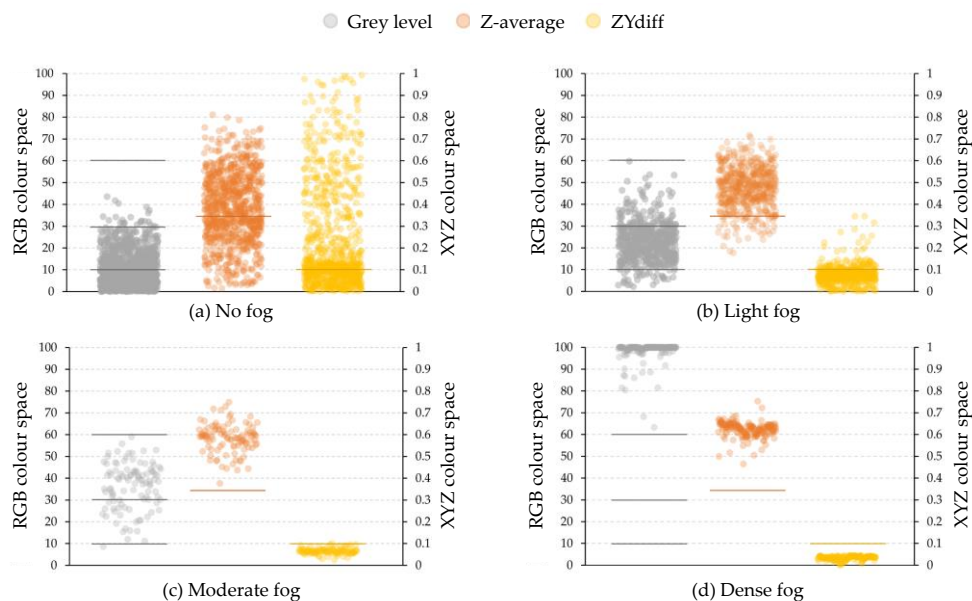


Figure 2-8. Image features representation for each ground truth label in the Ceit-Foggy dataset

Figure 2-8 shows the representation of grey level, \bar{Z} and $ZYdiff$ parameters obtained by the analysis of RGB and XYZ colour spaces.

All the parameters analysed range from [1,100]:

Grey level: which has a direct link with the fog level when the scenario is classified as foggy. It is represented in grey in the figures above.

\bar{Z} value: which differentiates cloudy from foggy scenarios when the sky is not clear. It is represented (multiplied by a factor of 10) in orange in the figures above.

ZYdiff: which differentiates foggy from sunny scenes. It is represented (multiplied by a factor of 10) in yellow in the figures above.

Represented in grey in the first column data, the grey level shows a clear difference between the four scenarios. With a special focus on the foggy scenarios, it is observed that most of the points could be clustered from [10-30] for light fog, [30-60] for moderate fog and [60 -100] for dense fog. On the other hand, *ZYdiff* is mostly under 10 and \bar{Z} above 35 for foggy scenes. However, for no-fog scenarios where both sunny and cloudy situations are included the range of this values is much wider.

2.3.2. Models comparison

For the comparison of the two models developed in this chapter, the two datasets are presented in Section 0. were used and executed both algorithms.

Table 2-5 shows the summarized results of this comparison where the accuracy and the processing time per image were analysed

The best performance result achieved for the deep learning model is 95.83% while for the rule-based method is 80.07%. This can be explained by the high complexity of modelling a complex phenomenon such as fog in different light conditions and scenarios with only four parameters. However, the neural network-based model learns by itself the characteristics of the training images and can collect thousands of different parameters, thus presenting much more adaptability to the different scenes that can be presented. Regarding the processing time of the algorithms, it is observed that the processing in the rule-based method is always slower. However, both models have an adequate computation time (below 300 ms) to run the analysis in real-time with no problem as long as the same current computational capabilities can be commanded.

Chapter 2 Road Damage Monitoring: road lines

Table 2-5. Fog detection models' evaluation parameters comparison for the Ceit-Foggy and Foggy Cityscapes DBF - extended datasets.

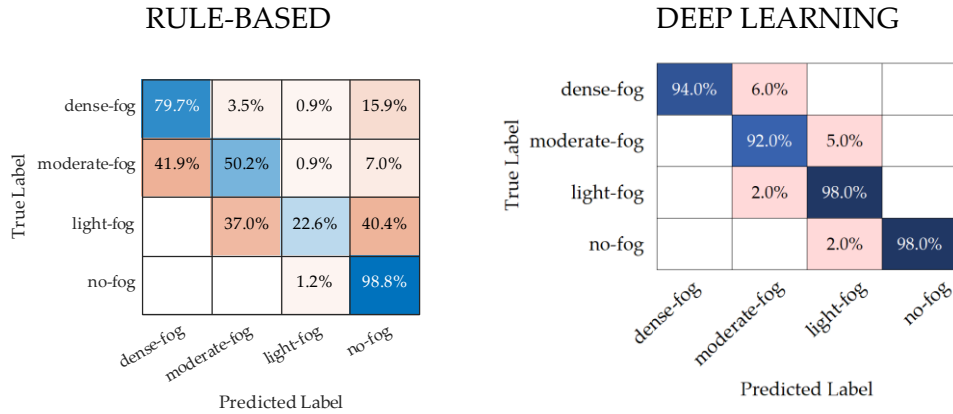
		RULE-BASED		DEEP LEARNING	
DATASET	Number of images	Accuracy (%)	time per image (ms)	Accuracy (%)	time per image (ms)
Foggy CityScapes BDF-extended	936	0.6368	296.27	0.9583	252.6
Ceit-Foggy	1681	0.8007	164.03	0.7055	74.4

For a deeper analysis of the performance of the two models, the confusion matrix was calculated for each test that allows studying in which scenarios each algorithm fails or succeeds.

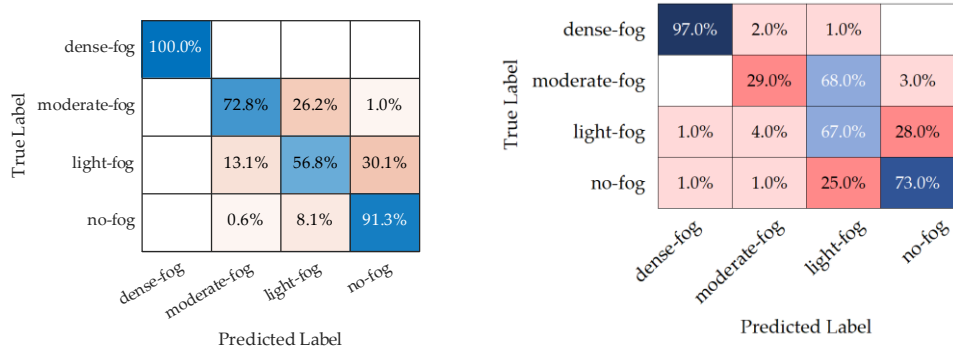
Rule-based method

In Figure 2-9 it is shown that the rule-based method confuses light fog and non-fog classes on several occasions for both datasets (40.4% and 30.1%), which is also true for the human eye in many of these cases. Though, a detailed analysis has shown that many of these cases have a limiting grey level. Failures in this scenario are considered to be of low importance since the alert level is lower and no false alarm is generated for the driver.

However, in Figure 2-9 (a) the confusion that occurs between the moderate fog and light fog classes for the synthetic dataset is considered more severe since the alert level is higher than the ground truth. The prediction in this classes is of a higher fog level in 41.9% and 37% of cases respectively. In the detailed analysis of the images, it is concluded that this behaviour may be explained due to the colour of the synthetic fog of the Foggy CityScapes BDF-extended dataset. The rule-based method has used real fog images as a reference to define the boundaries, in these pictures the fog takes a warmer colour than synthetic fog which is represented with a much pure and cold white tone which results in a foggier sensation.



(a) Foggy CityScapes BDF-extended dataset



(b) Ceit-Foggy dataset

Figure 2-9. Fog detection models' confusion matrix comparison for the Ceit-Foggy and Foggy Cityscapes DBF - extended datasets.

Deep learning-based method

The reverse occurs for the deep learning-based model when tested on real fog images from the Ceit-Foggy dataset. Both in the case of light and moderate fog 28% and 68% of the cases respectively are predicted to have a lower fog level. This use case is also related to the previous fog tone phenomenon: since the deep learning model was trained with synthetic fog images the real fog does not have enough white level to be classified at the level of the ground truth (see Figure 2-10).

It is also worth mentioning the good results obtained for the deep learning model when testing with synthetic fog images. Since the fog tone of the

Chapter 2 Road Damage Monitoring: road lines

training and test images is the same, the model fits very well with the new test images.



(a) Synthetic fog

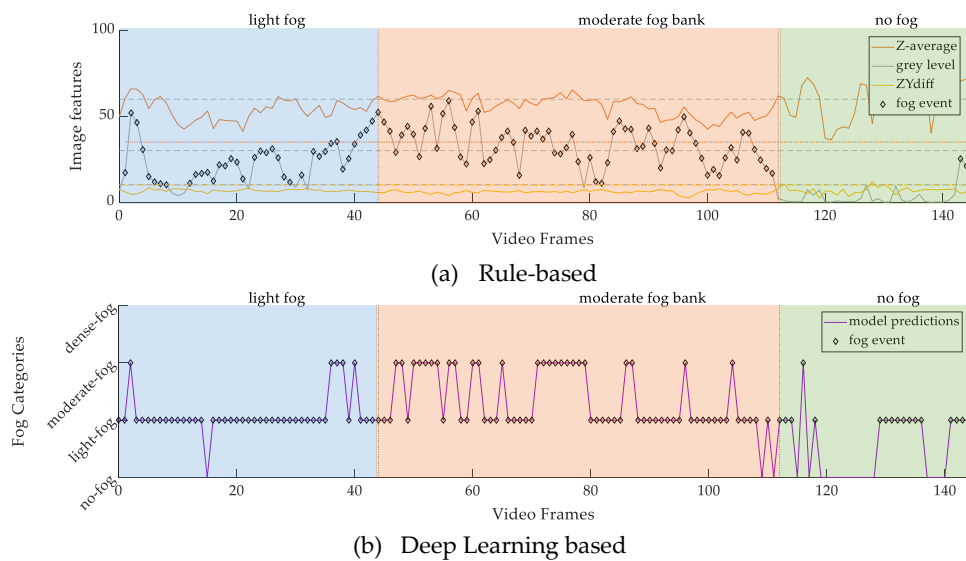


(b) Real fog

Figure 2-10. Fog tone comparison of the two datasets. Synthetic fog presents a colder colour than the real foggy scenes.

Video analysis for both models

Finally, Figure 2-11 shows a comparison of the two models with one of the original videos used to generate the Ceit-Foggy dataset. This video was recorded in a light-fog scenario, but during the route, the vehicle enter a thicker fog bank which then fades away at the end of the video.





(c) Example frames

Figure 2-11. Off-line video analysis and model comparison for a light-fog scenario with a moderate-fog bank that disappears in the last frames

Figure 2-11 (a) shows the result of the rule-based method where the evolution of all the parameters was illustrated. The diamond markers present the fog event alarm and judging de grey level it is shown how the light-fog [10-30] evolves to the moderate-fog level [30-60] and finally decreases to a no-fog scenario.

The same tendency is reflected in Figure 2-11 (b). This graphic presents the behaviour of the deep-learning model which has a discrete output of four different levels. It is also observed the change from light to moderate when entering the fog bank and to no fog when exiting the fog bank.

Lastly, Figure 2-11 (c) shows three samples of the video with the three different scenarios commented above.

2.4. Discussion

Fog is one of the most dreaded weather phenomena on the road as it can significantly reduce visibility abruptly in a matter of seconds and can last from a few metres to kilometres. It poses a safety hazard to both human-driven and automated vehicles. Nowadays, information regarding fog can be displayed on variable messaging boards, however, this method is not very effective as its location is imprecise.

In this chapter, a new dataset was generated containing 1681 foggy images that were labelled in three different fog levels. In addition, the synthetic Foggy Cityscapes DBF dataset was extended for 4 different visibility levels. The development of two fog detectors has been also presented, the first one addressed classic computer vision techniques while the second one uses the most recent deep learning techniques.

Both models have shown good results in tests with different datasets and videos. Accuracy levels of more than 80% and processing times of less than 300ms have been reported. Therefore, it could be said that both are valid detectors to integrate into an on-board system that can monitor in real-time this meteorological phenomenon. This sensorisation would provide better and more accurate information on the location of fog banks, making it possible to generate ad-hoc warnings to help drivers and automated vehicles to be prepared and make better decisions while driving and this enhance road safety.

It is worth mentioning that a preliminary version of this system is already integrated on an onboard barebone industrial mini PC and is being validated on-site generating real-time traffic alerts (Linux i7-6600U CPU & 8 GB RAM).

Chapter 3

Traffic Signs Monitoring: asset inventory



Part of this chapter has been published in:

Iparraguirre, O.; Amundarain, A.; Brazalez, A.; Borro, D. "Sensors on the Move: Onboard Camera-Based Real-Time Traffic Alerts Paving the Way for Cooperative Roads". Sensors, Special Issue on Sensors for Road Vehicles of the Future, Vol.21, No. 4, 1254. February 2021. doi:10.3390/s21041254.

3.1. State of the art

Vertical signalling is essential for the safe coexistence of all road users and road management. For this reason, there are already many studies on traffic sign recognition in the literature. These systems are of great interest for ADAS, for the decision-making of autonomous vehicles as well as for road network maintenance (asset inventory) [55]. TSR can be categorized into two subtasks, detection which aims on locating the traffic sign in an image and classification which will later predict the category of the detected sign.

Computer-vision methods have been widely spread to address this task. Those have evolved over the years, first, classical vision techniques were used most of them based on shape or colour modelling, ML techniques were also employed, however, these methods present weakness when dealing with different scenarios with illumination change, occlusions etc. Thus, deep learning techniques arise to solve these problems, however, they require large annotated datasets. Next, a brief review of these methods will be presented.

3.1.1. Classical computer vision techniques

3.1.1.1. Colour and shape-based methods

These methods are the oldest in the state of the art and were mainly used for traffic sign detection.

Colour-based methods take advantage of the traffic sign design which means to be easily distinguishable from the background. Segmentation techniques by colour thresholding of different colour spaces are the most used among the researchers. RGB is the most intuitive colour space, however, it is also very sensitive to light conditions, weather conditions, reflections etc. Thus, choosing the colour space is very important. Other works developed colour-based detection methods using hue, saturation and value (HSV), hue, saturation and intensity (HIS) as well as other various colour spaces [56].

Regarding the shape-based methods, the most used one is the Hough Transform which consists of voting of each pixel edge image for the object centre at the object boundary [57,58]. Others apply template matching techniques [59] or similarity detection by analysing symmetry [60,61]. Some

works use Distance Transforms [62] capturing object shape template hierarchy, Edge Detection features [63] and Haar-like features [64].

Despite TSR requiring colour and shape information, the problems of illumination changes or colour fading of traffic signs, as well as the deformation and occlusions are still unresolved [65].

3.1.1.2. Machine learning methods

Thus, conventional computer vision methods are employed to extract and learn new specified visual features. The most popular features are Haar-like features, SIFT (Scale Invariant Feature Transform) features, HOG (Histogram Oriented Gradients) features and SURF (Speed Up Robust Features) features, but there are also others such as ICF (Integral Channel Feature), ACF (Aggregated Channel Features) etc. They were further applied both for detection and classification tasks.

The increase in complexity of the features to learn requires using more powerful algorithms. The most popular ones are the cascaded detectors as Viola-Jones [66] strategy, which performs classifier training based on AdaBoost [67]. The benefit of this classifier is that detection runs fast and its accuracy is fair [68]. Other popular methods are also used for TSR, such as SVM [69] that contract an N-dimensional hyperplane that optimally separates the data into two categories; Random Forests (RF) [70] which operates by constructing multiple decision trees during the training time and outputting the class of individual trees, genetic algorithms [71] based on natural selection processes; artificial neuronal networks (ANN) [72] or CNNs [73,74]. These last two methods are increasingly gaining popularity in recent years due to the advances in graphics processing units.

However, these methods also still present limitations, especially when the number of features to be learned increases, and the speed of the algorithms also increases, making them ineffective for real-time analysis. In addition, some of these features still present difficulties when there are abrupt changes in the background.

3.1.2. Deep learning techniques

In recent years, DNNs have received great attention in computer vision research and have been widely used in both object detection and recognition.

The main difference and also its strongest point is that deep learning uses multilayer neural networks to automatically extract and learn the features of visual objects.

The most popular approach for TSR is the end-to-end CNN-based models. For these models, there are two kinds of algorithms, the two-stage detectors which are based on the generation of region proposals by selective search to predict the candidate bounding boxes followed by the classification, and in contrast the one-stage detectors who do the object classification and bounding-box regression directly in a single-step without using pre-generated region proposals. The first ones are normally more accurate; however, they are also slower than single-stage algorithms.

The TSR systems have evolved as the algorithms themselves have advanced. Zhu et al. [75] proposed a six layers Fast-RCNN model to simultaneously classify and locate traffic signs. Zuo et al. [76] used Faster R-CNN-based model. Shao et al. proposed a simplification of the Gabor wavelet to improve Faster R-CNN for traffic sign detection [77,78]. Zhang et al. [79], Yang et al. [80] and Yuan et al. [81] added an attention module to the CNN to improve the detection of small traffic signs or under complex backgrounds. With the emergence of single-stage detectors, new works were developed. Zhang et al. [82] apply CNN inspired by YOLOv2 while Wang et al. [83] used YOLOv4. Whereas Shan et al. [84] and Jin et al. [85] proposed different improvements for Single Shot Detector (SSD) CNN-based algorithms.

There are countless examples in this field and currently, in the state of the art, many of them achieve results that exceed human performance (98.84%) [86]. However, these deep learning methods are highly dependent on the quality of the datasets they use. Although technically the level is very high, many non-technical challenges can be overcome to jeopardise the performance of these algorithms, especially when dealing with complex scenarios where the resolution of the images is low, there are different lighting or weather conditions, fading and blurring, occlusions or other artefacts, multiple appearances of signs etc. [87] (see Figure 3-1).



Figure 3-1. Sample images for TSR challenges: lighting or weather conditions, artefacts, low-resolution signs, motion blur, rotation, occlusion, damage, inconsistencies and intra-class variation (The vast majority of the samples are extracted from Ceit-TSR dataset).

3.2. Materials and methods

3.2.1. Existing datasets

The traffic sign recognition research field has increased its attention in the last years and therefore since 2011 many new large datasets have been publicly available. This has allowed many comparative studies that have helped to improve existing algorithms.

The German Traffic Sign Recognition Benchmark (GTSRB) is one of the first large and public datasets collected in various locations in Germany that was created to evaluate the classification branch of the problem with the International Joint Conference on Neural Networks (IJCNN) 2011 competition [88]. The authors of this dataset added a new benchmark for detection purposes (GTSDB) and organised also the IJCNN 2013 [89]. These were the boom for the creation of many other new public datasets for TSR in different countries.

Belgium Traffic Sign Dataset [90], STS Dataset recorded in Sweden [91], RUG from Netherlands [92], Stereopolis dataset from France [93] and MASTIF dataset from Croatia [94] which together with the GTSDB led to the creation of the European Traffic Sign Dataset (ETSD) [95] which includes all these datasets and also extend the annotations of some of them to label all possible classes. Some datasets contain images from China such as Tsinghua-Tencent Dataset (TT100K) [96], Chinese Traffic Sign Dataset (CTSD) [97], Changsha University of Science and Technology Chinese traffic sign detection benchmark (CCTSD) [82], and Complex Traffic Sign Dataset CTSD [98]. Russian traffic Sign Dataset (RTSD)[99], LISA dataset with American signs [100], Dataset of Italian Traffic Signs (DITS) [99], Korean Traffic Sign dataset (KTSD) [101], DFG from Slovenia [102] and Cure TSD of Belgium [103].

Finally, there is a recent dataset called Mapillary Traffic Sign Dataset (MTSD) [104] for detection and classification on the Global Scale created by Facebook. This dataset aims to cover the diversity of countries all over the world in urban and rural areas, images of different quality and captured under varying conditions. The last paper that reported on TSR datasets proposed two new synthetic datasets [105] which consist of Carla Traffic Sign

Detection (CTSD) and Carla Traffic Sign Recognition Dataset (CATERED) both created through the Carla simulator.

All the characteristics of these publicly available datasets are summarized in Table 3-1.

3.2.2. Our dataset

Ceit-TSR consists of 264 colour images captured from 40 different videos of driving tracks within the Basque Country (Spain). They were recorded using different mobile phones and onboard cameras located on the dashboard. The images of this dataset were specifically selected so that in addition to the different weather and light conditions that are covered in other existing datasets, they would also include other complex conditions. Those incorporate images with motion-blur, low-resolution signs, distant signs, low contrast and windshield artefacts (reflections, raindrops, dirtiness etc.) as it is shown in Figure 3-2. All images were manually annotated using the *Computer Vision Annotation Tool* (CVAT) [106] resulting in 583 bounding boxes that were also classified in 49 different classes (see Figure 3-3).



Figure 3-2. Ceit-TSR dataset. Sample images showing some of the challenging conditions: low contrast, fog, reflections, shadows, and heavy rain.

Table 3-1. Summary of the publicly available TSR datasets.

Purpose	Number of images	Number of bounding	Classes	Categories	Resolution	Videos	Challenging conditions	Country	Publication year
RUG	Detec./Classif. 84	--	3	NA	360x270	no	no	Netherlands	2003
Stereopolis	Detec./Classif. 847	251	NA	10	960 x 1080	no	no	France	2010
GTSD	Detection 900	1213	43	4	1360x1024	no	no	Germany	2011
BTSD	Detection 25634	13444	NA	NA	1628x1236	yes (4)	no	Belgium	2011
BTSC	Classification 7125	NA	62	NA	1628x1237	yes	no	Belgium	2011
STS	Detec./Classif. 20000	3488	7	NA	1280x960	no	no	Sweden	2011
MASTIF	Detec./Classif. 4875	13036	NA	5	720x576	yes	no	Croatia	2011
LISA	Detec./Classif. 6610	7855	49		640x480 - 1024x522	yes (17)	no	USA	2012
GTSRB	Detec./Classif. 51840	1728	43	4	1360x1025	no	no	Germany	2013
TT100K	Detec./Classif. 100000	30000	45	NA	2048x2048	no	no	China	2016
RSTD	Detec./Classif. 179138	15630	156	6	1280x720 - 1920x1080	no	no	Russia	2016
CTSD (Chin.)	Detec./Classif. 1110	1574	48	NA	1024x768 & 1270x800	no	no	China	2016
DITS (1)	Detection 1887	NA	NA	3	1280x720	no	yes	Italy	2016
DITS (2)	Classification 9254	NA	58		1280x721	no	yes	Italy	2016
CCTSD	Detec./Classif. 10000	13361	NA	3	several	no	no	China	2017
KTSD	Detection 498	832	NA	3	several	no	yes	Korea	2017
ETSDB	Detec./Classif. 82476	NA	164	4	several	no	yes	⁶ European countries	2018
CTSD (Comp.)	Detec./Classif. 2205	3755	153	3	several	no	yes	China	2018
DFG	Detec./Classif. 6957	4359	200	NA	1920x1080 & 720x576	no	*augmented	Slovenia	2019
Cure TSD	Detec./Classif. 896700	648186	14	NA	1628x1236	yes (2989)	yes	Belgium	2020
MTSD	Detec./Classif. 105830	354154	400	NA	several	no	yes	Global	2020
CTSD (Carla)	Detection 55323	--	43	NA	1920x1080	no	yes	Synthetic	2021
CATERED	Classification 94478	--	43	NA	1920x1080	no	yes	Synthetic	2021



Figure 3-3. CVAT tool that shows the labelling task for Ceit-TSR dataset

3.2.3. Used datasets

In addition to the Ceit-TSR dataset and to be able to evaluate the generalisability of the developed algorithms and contrast them with the works reported in the state of the art, the datasets GTSDB, GTSRB and ETSD have also been used (see Table 3-2).

Table 3-2. Summary of the used datasets for TSR in Chapter 3.

	Number images	Labels	Classes	Country
GTSDB (test)	300	361	43	Germany
GTSRB (test)	12630	12630	43	Germany
ETSD	18550	18550	164	Belgium, Croatia, France, Germany, Netherlands, Sweden
Ceit-TSR	264	583	49	Spain

3.2.4. Developments

The implemented traffic sign recognition system is composed of two modules: detection and classification (see Figure 3-4). As will be explained in the following sections, both algorithms were fine-tuned, detector and

classifier, except for the neural networks used for the classification which were pre-trained models.

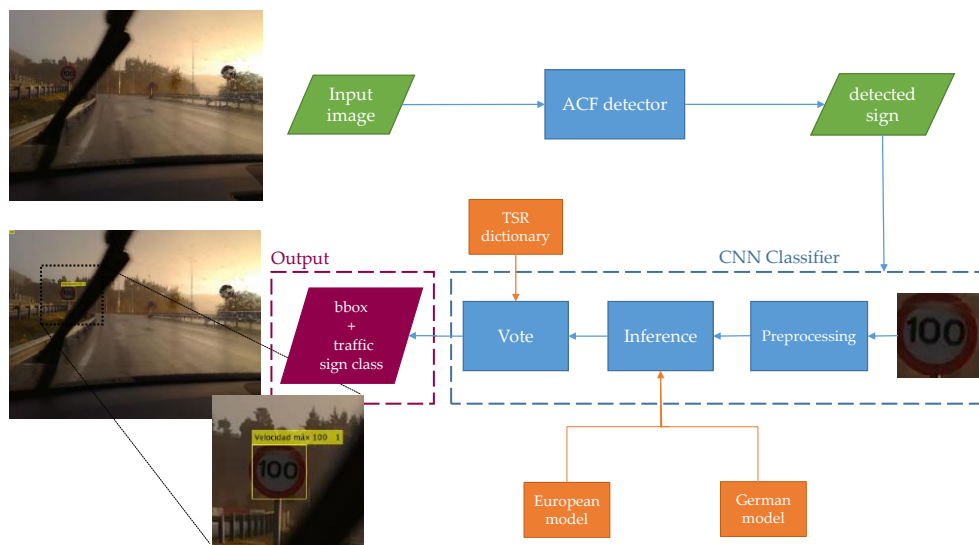


Figure 3-4. Implemented Traffic Sign Recognition workflow.

The hardware used in these developments is a Windows 10 PC with an Intel Core i7 processor, NVIDIA GeForce RTX 3080 10 GB GPU and a total RAM of 32 GB. As for the software, the authors employed Matlab 2020a (Image Processing Toolbox V11.1, Computer Vision Toolbox v9.2, Piotr's Matlab Toolbox v3.5, Deep Learning Toolbox v14.0, Deep Learning Toolbox Importer for Caffe Models v20.1.0 and Deep Learning Toolbox Importer for Tensorflow-Keras Models v 20.1.0).

3.2.4.1. Traffic Sign Detection

After the revision of the state-of-the-art of traffic sign detection methods, it can be concluded that there is no clear framework that achieves the best results. Thus, it was decided to implement and compare some of the most popular methods. Firstly, classical features were used such as colour and shape for modelling a Viola-Jones cascade detector. However, these alternatives were finally discarded as they were difficult to adjust and not very flexible with changing light conditions.

Usually, existing detectors could be improved in two ways: using more complex features or implementing more powerful learning algorithms. Since

the combination of boosting and cascading is proven to be very efficient for object detection [107], the key is to find representative characteristics at a low computational cost. In this context, a combination of record-breaking characteristics has emerged for pedestrian detection [108] and in this thesis, this method was applied for traffic sign detection. The ACF detection framework uses an AdaBoost classifier trained with ACF features to classify image patches. The entire image is searched by using a multiscale sliding window approach. These ACF features consist of ten different channels: three from the LUV colour space, the gradient magnitude, and the six oriented gradient maps (see Figure 3-5). Afterwards, the sum of every block pixel of these channels is computed using fast features pyramids and downscaled. Features are single-pixel lookups in the aggregated channels. Boosting is used to train and combine decision trees over these features (pixels) to locate accurately the object [108]. The channel extension offers a rich representation capacity, while the simplicity of the features allows a low computational cost.

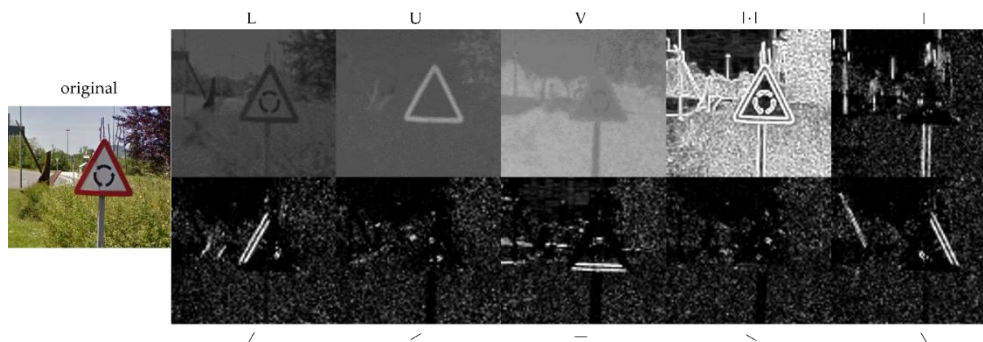


Figure 3-5. ACF features. In the first row from left to right: original image, LUV channels, the gradient magnitude and individual representation of HOG features in different angles, of a sample sign.

3.2.4.2. Traffic Sign Classification

In this phase, two different pre-trained classifiers were used for inference and then a voting system was employed to give a final prediction.

- European Classifier: this is an 8-layers model with VGG architecture modified adding 1) L2 regularization of $1e-4$ value on each convolutional and fully connected layer and 2) Batch Normalization

after each convolutional layer and before the ReLu activations. Adam optimizer was used and the learning rate was set to 1e-3. This model was trained by Serna et al. [95] on the ETSDB dataset. The input shape is $48 \times 48 \times 3$ whereas the output is an array of 164 different classes (see Figure 3-6).



Figure 3-6. Traffic sign samples of the ETSDB dataset. There are 164 different classes grouped into 9 categories: danger, regulatory (priority, prohibitory, mandatory and special regulation) informative (information, direction and additional panels) and others.

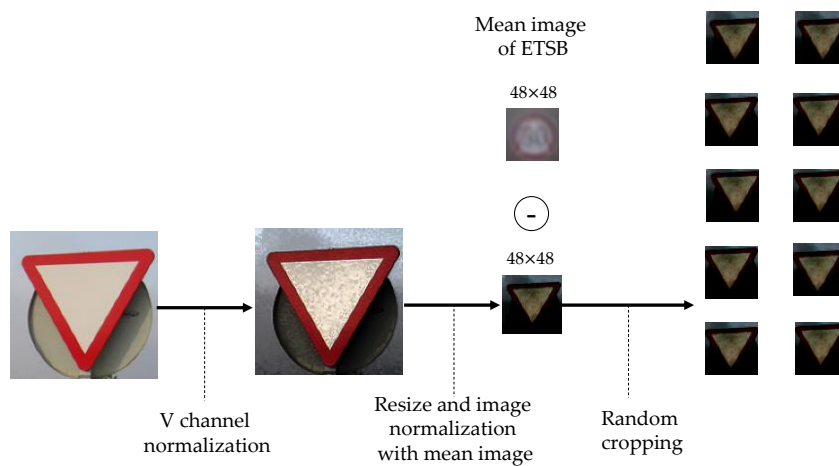
- German Classifier: this is an 8-layers model with AlexNet architecture. However, the authors do not provide details of the training. This model was trained by people of the Center for Digital Technology and Management [109] on the GTSDDB dataset. The input shape is $227 \times 227 \times 3$, whereas the output is an array of 43 different classes (see Figure 3-7).



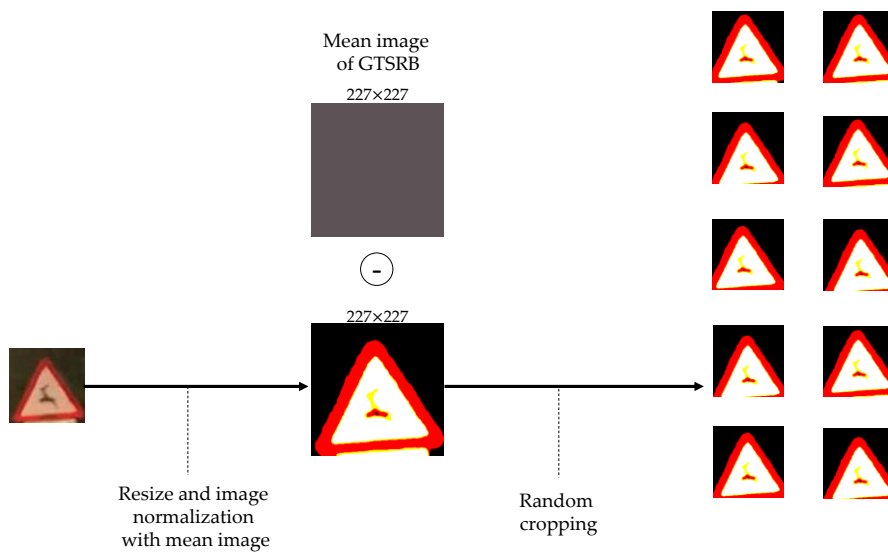
Figure 3-7. Traffic sign samples of GTSRB dataset. There are 43 different classes.

Image pre-processing

Several studies had probed that pre-processing methods that normalize the image and give better contrast can improve the performance of the existing pre-trained CNN models [95,110–112]. Thus, after studying different combinations a specific pre-processing routine was finally defined for each of the classifiers that are used in this chapter. (1) The V channel normalization is firstly applied for the input images of the European classifier to smooth the pixels' luminance distribution. (2) Then, both models require the mean image subtraction to make the network less sensitive to the changing background and lighting conditions. For this step both images must be of the same size, thus they are resized to the CNN input shape before doing the subtraction. (3) Next, the input image will go through a random cropping loop to generate ten different images. All of them will be passed to the classifier and the one with the highest confidence score will be selected (see Figure 3-8).



(a) For European classifier



(b) For German classifier

Figure 3-8. The pre-processing phase before applying the corresponding classifiers. V channel normalization, mean image subtraction and random cropping tasks are presented.

Voting system

Voting is an ensemble method that combines the performance of multiple models to make predictions. In this case, the predictions of the german and the European classifiers are combined to obtain the final output. The aim of incorporating this voting is to mitigate the risk of one model making an inaccurate prediction by having other models that can make the correct prediction. For the case of these two classifiers, the European classifier covers a wider variety of traffic sign classes whereas the german model output is more reduced but presents better performance for some of the classes.

Therefore, a soft voting system was implemented where every individual classifier provides a probability of the output class and the target label with the greatest sum of probabilities wins the vote. Therefore, since there are only two votes, in this case, the winning class will be the one in which the two votes coincide or, if they differ, the class with the higher confidence score.

3.2.5. Evaluation parameters

For the analysis of the detector's performance precision and recall metrics were employed, where precision measures the proportion of the total number of signals detected that are correct, while recall measures the proportion of the total number of signals to be detected that are correctly detected.

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

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

For the evaluation of the multiclass classification model's performance, accuracy metric was employed that calculates how often predictions match the one-hot label. Thus, the accuracy is defined as:

$$Accuracy = \frac{Matched\ predictions}{Total\ predictions}$$

Inference times have also been measured to assess their real-time performance.

3.3. Results

In this section, the results of both the detector and the ensemble classifier are evaluated in four different datasets. GTSDB and Ceit-TSR contain complete road scenes where detection followed by classification is necessary. GTSRB and ETSDB are composed of cropped images containing just a traffic sign, so just the second phase should be needed. However, GTSRB detection is also applied since the original dataset includes a border around the actual sign of 10 per cent of the sign size, at least 5 pixels.

Table 3-3 shows the detection and classification results of the pure and ensemble models presented in this chapter for the detection and classification phase. In addition, the computational time per frame is measured.

Table 3-3. Detection and classification results of the pure and ensemble models tested in the four different datasets used in this chapter.

		Model		
		European	German	Ensemble
GTSDB (300 imgs)	detection prec./recall	1.00/0.76	1.00/0.76	1.00/0.76
	classification accuracy	0.967	0.971	0.985
	processing time	0.069	0.121	0.138
European (18145 imgs)	detection prec./recall	--	--	--
	classification accuracy	0.889	0.868	0.924
	processing time	0.009	0.062	0.071
GTSRB (12630 imgs)	detection prec./recall	1.00/0.52	1.00/0.52	1.00/0.52
	classification accuracy	0.915	0.934	0.964
	processing time	0.011	0.065	0.072
Ceit-TSR (264 imgs)	detection prec./recall	0.87/0.57	0.87/0.57	0.87/0.57
	classification accuracy	0.691	0.552	0.713
	processing time	0.090	0.146	0.138

3.3.1. Traffic Sign detection

Since the learning process of an ACF detector is very similar to that of the cascade detector, this function also needs images with and without traffic signs. It uses as positive images all those that are passed as an argument while the negative ones are automatically generated.

The design and selection of parameters for the detector are crucial to achieving optimal implementation of it. Thus, several fine-tuning experiments were done showing that:

- The inference time was always about 200ms per image.
- Up to 5000 learners or decision trees do not improve the results.
- Object training size should be a maximum of 30×30 pixels, bigger dimensions may increase the false negatives.
- The confidence threshold is key to compensate false positives and negatives.

The final detector was trained with an object training size of 30×30 pixels. It used 5000 weak learners in 30 stages.

As it can be observed in Table 3-3 the precision of this detector is pretty good (1.00 - 0.87), this is, most of the predicted bounding boxes are correct. However, the recall is lower (0.76 - 0.52). This means that there are some traffic signs that the detector will miss. Nevertheless, this factor is not considered to be that relevant for the actual application of this thesis, since it is a model that will be in an on-board vehicle that is continuously driving on the same roads and therefore will make several passes and will have the opportunity to detect the traffic sign that it has previously missed. Even so, a detailed analysis has been carried out to try to identify the scenarios where the detector fails. Figure 3-9 shows an example of Ceit-TSR dataset where there is one True Positive (TP) or detected sign and three False Negatives (FN) or missed signs. It is shown that, when the traffic sign to detect is small, this is, less than 15×15 pixels (see Table 3-4) the detector fails.



Figure 3-9. Example of false-negative detections due to the small size of the traffic sign. Table 3-4 shows detailed information.

Table 3-4. Ground truth of the example shown in the figure above. A sign is considered small when its width and height are below 15 pixels.

Image	Bounding Box	Class	Width	Height
IMG (5).jpg	[454,335,467,348]	80	13	13
IMG (5).jpg	[848,174,977,303]	24	129	129
IMG (5).jpg	[225,339,239,353]	41	14	14
IMG (5).jpg	[347,337,358,348]	87	11	11

For the case of the GTSRB dataset in which the image has almost no background, the size of the images where the detector failed versus those where it succeeded to predict the bounding box has been analysed. After doing a Wilcoxon Rank Sum test it is observed that there is a significant

difference in the image size between both groups of samples ($p < 0.0001$) (see Table 3-5).

Table 3-5. Image size comparison between GTSRB detected traffic signs and missed traffic signs.

	Mean size	Standard deviation
Detected signs	60.33 × 61.71	24.71 × 26.59
Missed signs	39.71 × 38.72	16.6 × 15.86

Analysing the traffic sign classes that the detector is missing there is no clear pattern. In Figure 3-10 are illustrated the traffic signs that present more difficulties when analysing the GTSRB dataset, this is, the ones that have more missed samples than detected ones.

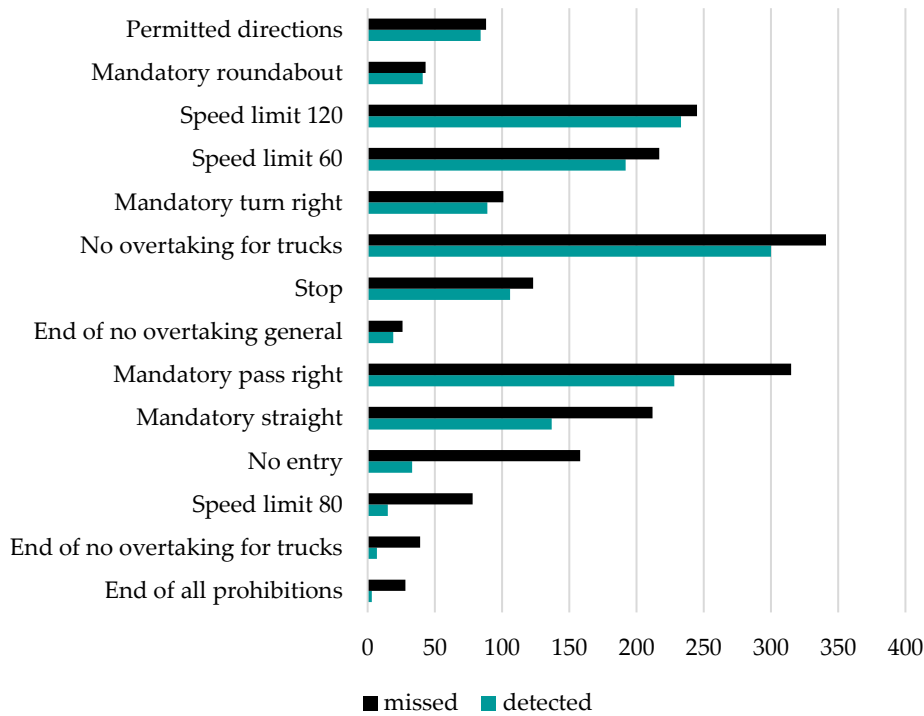


Figure 3-10. Traffic sign classes where there are more missed samples than detected for the GTSRB dataset.

3.3.2. Traffic Sign classification

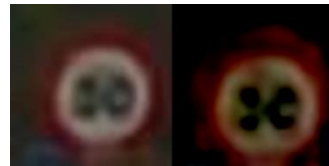
Concerning the classification of the detected signs, as was explained in section 3.2.4.2 two models were used that were already for TSR, thus they did not require any hyperparameter fine-tuning. Nevertheless, an additional pre-processing was added to the input images for improving their performance and the final voting system to ensemble both outputs.

Table 3-3 shows the classification results of the pure and ensemble models for the four different datasets used in this chapter. In overall conclusion, it can be observed that the ensemble model performs in all cases better than the pure models. Regarding the datasets, the best result is obtained with GTSDDB achieving a 98.5% classification accuracy followed by GTSRB and ETSDDB which achieve also accuracies higher than 92.4%. The worst performance is for Ceit-TSR where the classification accuracy drops to 71.3%. This is probably due to the complex conditions presented in this dataset. Next, Figure 3-11 shows some examples where the classifier has failed in different conditions. The left side shows the complete scene and on the right side the cropped problematic detection. These complicated conditions are listed below:

- The classifier (and detector) can recognise traffic signs that the authors did not label since they were not able to differentiate them. See Figure 3-11 (a).
- The low resolution of the image gives confusion to some of the symbols although the overall category of the sign is detected correctly. See Figure 3-11 (b).
- The pre-processing harmed the image in some specific scenarios and the classifier receives an input image that is not clear. See Figure 3-11 (c).
- The model detects objects with similar characteristics to traffic signs but which are not traffic signs. See Figure 3-11 (d).

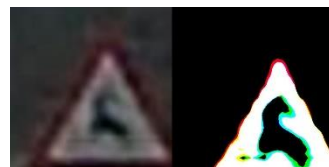


Pred.: Closed all directions



Pred.: Speed limit 80

(a) Traffic signs not labelled



Pred.: Attention bottleneck
right



Pred.: No overtaking trucks

(b) Low-resolution traffic sign

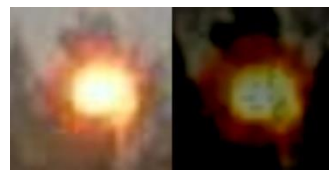


Pred.: No entry trucks



Pred.: Speed limit 60

(c) Pre-processing harmed the input image



Pred.: Closed all directions



Pred.: Standing and parking prohibited



Pred.: Mandatory pass right.

(d) Detected other objects

Figure 3-11. Visual analysis of complex situations in Ceit-TSR dataset where the classifier fails. The left side shows the complete scene and the right side shows the cropped detection before and after pre-processing.

Chapter 3 Road Damage Monitoring: road lines

In addition, the classification results were analysed through the confusion matrix of each test. These matrixes can be seen in Appendix C. After studying them it is concluded that there is no specific pattern for the failures. In general, the bad classifications confuse the symbol but the overall category (warning, prohibition, regulatory etc.) is generally correct, this is especially noticeable for speed limit signs (see Figure A-3).

3.4. Discussion

Vertical traffic signal recognition technology has come a long way in recent years and is already on the market today. However, in order to achieve a future of autonomous driving, it is imperative to further improve the robustness of this technology so that it can function in all circumstances and expand the number of recognised signals. This chapter has presented the TSR system developed to be embedded in a maintenance vehicle that can continuously monitor the signalling to create an asset inventory. This system will allow to study the status of the signals and manage the necessary maintenance tasks.

Nevertheless, in order to improve the robustness of these systems, many new images are needed to cover all types of scenarios. Therefore, in this thesis, a small dataset containing 264 annotated images has been generated to detect and classify up to 49 different signals. This dataset provides new images from another country that was not covered by the current datasets and presents complicated scenarios with low-resolution images, dirt on the glass and different weather conditions. The recognition was addressed in two stages, one model for signal detection and one for signal classification. These algorithms have been tested on some of the best-known public datasets and have produced results suitable for the application.

It is worth mentioning that this system is already integrated on an onboard barebone industrial mini PC and is being validated on-site (Linux i7-6600U CPU & 8 GB RAM). Currently, it is working as an assistant system for the road management authorities. However, this solution could also help to update in-vehicle maps and traffic signage or assist with the dynamic speed limit system.

Chapter 4

Road Damage Monitoring: road lines



The final development presented in this chapter has been published in:

Iparraguirre, O., Iturbe-Olleta, N., Brazalez, A., Borro, D., "Road marking damage detection based on deep learning for infrastructure evaluation in emerging autonomous driving". IEEE Transactions on Intelligent Transportation Systems July 2022. doi: 10.1109/TITS.2022.3192916.

4.1. State of the art

Increasing traffic, harsh weather conditions, ageing, poor construction quality and lack of proper maintenance cause the road infrastructure to deteriorate. This deterioration in turn leads to a loss of driving quality, passenger comfort and road safety. In addition, poor road infrastructure, and especially poor signage, could lead to the malfunctioning of autonomous vehicles that would fail to detect the environment. Therefore, maintaining the roadway infrastructure in an optimal state is of vital importance.

Currently, these tasks of observing and detecting infrastructure failures in order to make maintenance decisions are entirely manual, which is tedious, time-consuming and costly work. Therefore, in recent years, more and more research is prioritising safety and the reduction of inspection costs to improve the efficiency of infrastructure maintenance by developing automatic road condition monitoring systems to drive a new type of intelligent maintenance of road infrastructure. The emerging cost-effective Road Condition Monitoring (RCM) systems allow to rationalize periodic inspections and thus minimize the costs associated with failing pavement structures and warrant long-standing structural integrity and safety levels.

4.1.1. Data acquisition systems

RCM systems rely on data acquisition systems, which are a combination of non-intrusive sensors and their platforms for the collection of 1D data, 2D visual data or 3D depth data (see Figure 4-1). The convenient deployment of these sensors can be done in different data acquisition platforms: Unmanned Aerial Vehicles (UAVs), smartphones and ground vehicles or robots.

Low-cost sensors such as accelerometers, gyroscopes, magnetometers and GPS are employed to measure one-dimensional parameters such as motion, rotation, velocity, orientation and location for vibration-based RCM. This kind of sensor cannot be used for real-time applications and a drawback is detection is limited only along the wheel path.

The most commonly used sensors are the ones acquiring 2D imaging which allows studying multiscale low-level and high-level feature extractions. These sensors are also economical and they can be used for real-time

applications depending on the processor’s capability. However, as a drawback, they are sensitive to illuminance levels.

The depth consideration (3D) of these bidimensional images can be acquired by using thermal imaging sensors, LiDAR, laser sensors or radars. Those sensors are not sensitive to light effects and facilitate the examination of intrinsic characteristics, however, these systems increase the cost significantly.

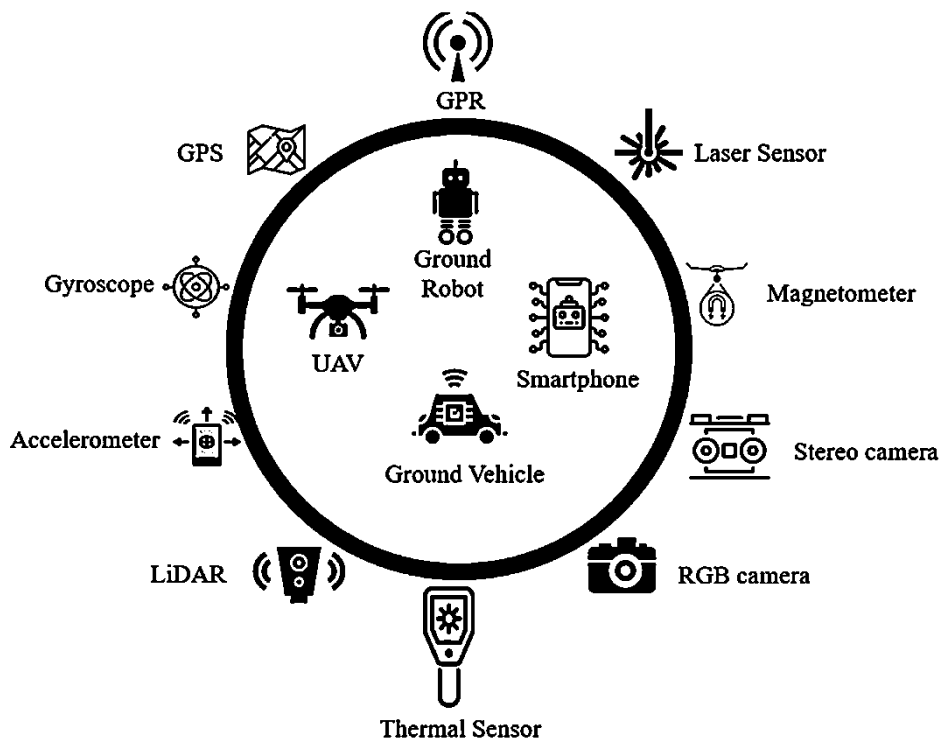


Figure 4-1. Sensors and data acquisition platforms schema for Road Condition Monitoring (RCM). Image extracted from [113]

4.1.2. Road condition monitoring

4.1.2.1. Vibration-based RCM

Inertial sensors such as gyroscopes or accelerometers, including those embedded in smartphones, are used to measure accelerations and estimate the IRI of the road in many scientific works [114,115]. The most novel

attempts, consider the dynamic characteristics of the vehicle (ie. suspension) to improve the measurement accuracy [116] or reconstruct the pavement profile [117]. This road condition measurement method is memory friendly and suitable for real-time detection, but they are vulnerable to errors due to noise or other road obstacles and the monitored area is limited.

4.1.2.2. Vision-based RCM

On the other hand, vision-based object detection, classification and segmentation have largely contributed to road distress detection, monitoring and analysis since these systems are economical solutions that are capable of monitoring the entire area covered by an image.

Machine Learning techniques

Some studies explore ML methods in pavement engineering to detect, classify and analyse anomalies. Generally, the algorithms used are SVMs, ANNs, RF or Canny edge detection combined with Otsu Thresholding. [118–120].

Deep Learning techniques

However, DL methods have become the most extensively used computational approach in the field of civil engineering and ITS. In comparison with the conventional feature extraction techniques, the DL-based techniques learn multi-level image features in detail, which are more descriptive than the handcrafted ones. Thus the DL models are better than ML computer vision approaches in terms of performance [121]. Overall, all these DL models are CNN-based models and they cover three different pattern recognition tasks: object detection, classification and segmentation [113]. Classification identifies the category of the defect, while object detection apart from classifying also locates where the object is at a bounding box and finally segmentation predicts the categories of each pixel and distinguishes the object instance.

Most of the work done on the classification and segmentation of road defects focuses exclusively on potholes and cracks. Some specific neural network architectures have even been proposed, such as CrackGAN or CrackU-net. In the field of damage detection, a more extensive study has been carried out with a greater variety of defects to be detected: potholes, joints, manholes, longitudinal lateral and alligator cracks, patches, fatigue etc. In these

researches, some images were used but in most of the cases, they were proprietary. The large-scale Road Damage Dataset (RDD) published by the University of Tokyo was the breakthrough that provides a platform to the scientific community for the comparison and evaluation of state-of-the-art DL models. This dataset proposed eight different damage types (see Table 4-1). and [122] and two different Road Damage Detection Challenges were organised in 2018 and 2020. In this last challenge, they addressed a multiclass detector trained just four out of the eight defined defects (potholes, longitudinal cracks, alligator cracks and lateral cracks) and the best team reported an F1 score of 0.6748.

Table 4-1. Road damage types and definitions proposed by Maeda et al. [122]

Damage type		Detail	Class name
Crack	Linear	Longitudinal	Wheel-marked part D00
			Construction join part D01
	Crack		Equal interval D10
		Lateral	Construction join part D11
	Alligator Crack	Partial pavement, overall pavement	D20
Other damage		Pothole	D40
		Cross walk blur	D43
		White line blur	D44

Thesis use case: Road marking damage

The novelty of this RDD2020 dataset is that it introduces annotations of new defects such as crosswalk blur or white line blur that very few works have addressed to date.

Vokhidov proposed a CNN-based method to recognize arrow-road markings in different light and damage conditions [123]. Kawano applied the YOLO model to detect road markings blur in colour and white lines and marks as well as in crosswalks [124]. Xu *et al.* proposed a hybrid feature detector and threshold-based method for line-making identification, classification and worn percentage calculation at pixel level [125]. On the other hand, Wei *et al.* built a road marking inspection system based on semantic segmentation to estimate the damage ratio by comparing the marking's damage part vs. the marking region [126]. However, one of the most outstanding works in this field is the one done by Maeda *et al.*, authors of RDD datasets, who proposed a multi-class classifier for eight different

defects including white line and crosswalk blur [122]. This study tries several CNN architectures and obtained the best performance with SDD Inception V2 and SDD Mobilenet achieving an accuracy of 0.83 for the specific white line blur damage (D44) (see Table 4-2).

Table 4-2. Table extracted from [122]. Detection and classification results for each class using the SSD Inception and SSD MobineNet. SIR: SSD Inception V2 Recall, SIP: SSD Inception V2 Precision, SIA: SSD Inception V2 Accucary, SMR: SSD Mobilenet Recall, SMP: SSD Mobilenet Precision, SMA: SSD Mobilenet Accuracy

Class	D00	D01	D10	D11	D20	D40	D43	D44
SIR	0.22	0.60	0.10	0.05	0.68	0.03	0.81	0.62
SIP	0.73	0.84	0.99	0.95	0.73	0.67	0.77	0.81
SIA	0.78	0.80	0.94	0.92	0.85	0.95	0.95	0.83
SMR	0.40	0.89	0.20	0.05	0.68	0.02	0.71	0.85
SMP	0.73	0.64	0.99	0.95	0.68	0.99	0.85	0.66
SMA	0.81	0.77	0.92	0.94	0.83	0.95	0.95	0.81

The authors of this thesis, therefore, see the detection of defects in road markings as a field in which progress can be made and the next sections will focus on this topic.

Road Monitoring systems' interoperability

Concerning the interoperability of the RCM systems, D. Arya et al. presented extensive work to study the usability of a single-source model in other countries and proposed models capable of detecting and classifying road damages in more than one country [127]. They conclude that the performance of a model is significantly degraded when applied to road images from another country and recommend the mixed-modelling strategy.

Moreover, these same authors have recently announced a new challenge oriented to address road damage detection in multiple countries. This challenge is named as "Crowdsensing-based Road Damage Detection Challenge" (CRDDC2022)[128] and allows participants to develop/propose their own datasets. This way, after a suitability analysis the selected datasets have been added officially to the RDD2022 dataset [129], a continuation of the original RDD2018 dataset. This new dataset contains 47,420 road images from six different countries: India Japan Czech Republic, Norway, the United States, and China. The images are annotated with four types of road

damage: Longitudinal Cracks (D00), Transverse Cracks (D10), Alligator Cracks (D20) and Potholes (D40).

4.2. Materials and Methods

4.2.1. Existing datasets

Nowadays, a few public datasets are available, for road damage detection. The lack of standardisation in the classification of road defects and in the systems for acquiring them is one of the major problems for the development of novel algorithms in this area.

Currently, there are several open-source datasets, the oldest one was generated in 2015, which indicates how recent research in this field is.

Most of the studies tested their methods on their own datasets [130–132]. After a detailed review, eight meaningful open-source datasets [133–141] were found and summarized in Table 4-3. As it is observed the most encountered defect type are cracks of different topologies followed by potholes. However, these categories do not have always the same characteristics since there is no standard, thus they are handcrafted. For the rest of the categories, there is a huge imbalance as is the case of the road marking blurring. In addition, most of these datasets were recorded with a specific data acquisition hardware which difficult a lot the comparison with new labelled images.

Therefore this thesis aims to follow in the footsteps of the Japanese RDD2018-2019 dataset that includes 8 different defects [141] and continues in the RDD2020 dataset [140] and RDD2022 dataset [129] with newly collected data from India and the Czech Republic as well as Norway, the United States and China respectively. These last two datasets contain only 4 specific damage types focused on cracks and potholes. This family of datasets are currently the ones that include most samples and its acquisition method seems to be the most easily replicable and thus it is selected to be the reference dataset in this thesis. Those opted for using the camera mounted on the front part of the vehicle with a wide view angle that records the road and other elements of the infrastructure. .

Dataset	Number of images	Image type	Data acquisition system (camera angle)	Classes	Distress types	Annotation type	Year	Country
German Asphalt Pavement Distress (GAPs)	1969	grayscale	Top-down (Mobile mapping vehicle)	6	Cracks, potholes, inlaid patches, applied patches, open joints, bleedings	bounding box	2015	Germany
Road Damage Dataset 2019 (RDD2019)	13135	RGB	Wide view (Dashboard mounted smartphone)	9	Long, wheel mark crack (D00), Lateral equal interval crack (D10), Lateral joint crack (D11), Alligator crack (D20), Rutting/bump/pothole/separation (D40), Crosswalk blur (D43), White line blur (D44)	bounding box	2019	Japan
Crack Forest Dataset (CFD)	118	grayscale	Top-down (Smartphone)	1	Cracks	bounding box	2016	China
CrackDataset (AigleRN, ESAR, LCMS, LRIS,	269	grayscale	Top-down (5 different sensors)	1	Cracks	bounding box	2016	France
Pavement Image Dataset (PID)	7237	RGB	Wide view (Google API)	9	Reflective, lane longitudinal, longitudinal crack, sealed longitudinal crack, block cracks, potholes, transverse cracks, alligator cracks, sealed reflective crack	bounding box	2020	U.S.A.
UAV Asphalt Distress dataset (UAPD)	3151 (2401 open-source)	RGB	Wide view (UAV)	6	Transverse crack, longitudinal crack, oblique crack, alligator crack, repair, pothole	bounding box	2022	China
EdmCrack600 dataset	600	RGB	Wide view (rear-mounted GoPro camera)	1	Cracks	pixel level	2020	Canada
Road Surface Damage dataset	2200	RGB	Wide view (Dashboard mounted smartphone)	4	Longitudinal and lateral cracks (D00, D10, D11) and potholes (D40)	bounding box	2020	Mexico
Road Damage Dataset 2020 (RDD2020)	26336	RGB	Wide view (Dashboard mounted smartphone)	4	Longitudinal cracks (D00), Transverse cracks (D10), Alligator cracks (D20) and potholes (D40)	bounding box	2020	Japan, India and Czech Republic
Ceit Road Damage (our proposed dataset)	971	RGB	Wide view (Smartphone)	1	White line blur (D44)	bounding box	2021	Spain
Road Damage Dataset (RDD2022)	47420	RGB	Wide view (Smartphones, High-resolution Cameras, Google Street View)	4	Longitudinal cracks (D00), Transverse cracks (D10), Alligator cracks (D20) and potholes (D40)	bounding box	2022	6 countries

Table 4-3. Existing publicly available datasets for road defects monitoring.

4.2.2. Used datasets

Following the analysis of the state of the art, this chapter will be focused on monitoring the while line blur defect. For this aim, the public reference dataset RDD2019 will be used so that it is comparable to the latest state-of-the-art work that uses the already defined category D44. In addition, a proprietary dataset is generated with Spanish images to study the interoperability and generalization capability of the developed models.

4.2.2.1. Road Damage Dataset 2019 (D44)

Road Damage Dataset 2019 (RDD2019) is a continuation of RDD2018 created by the researchers of the University of Tokyo Maeda *et al.* It contains 13,135 images and 30,989 annotations for nine different damage categories, the ones listed in Table 4-1 and the category D50 which corresponds to utility hole (see Figure 4-2). These annotations are in Pascal VOC format.

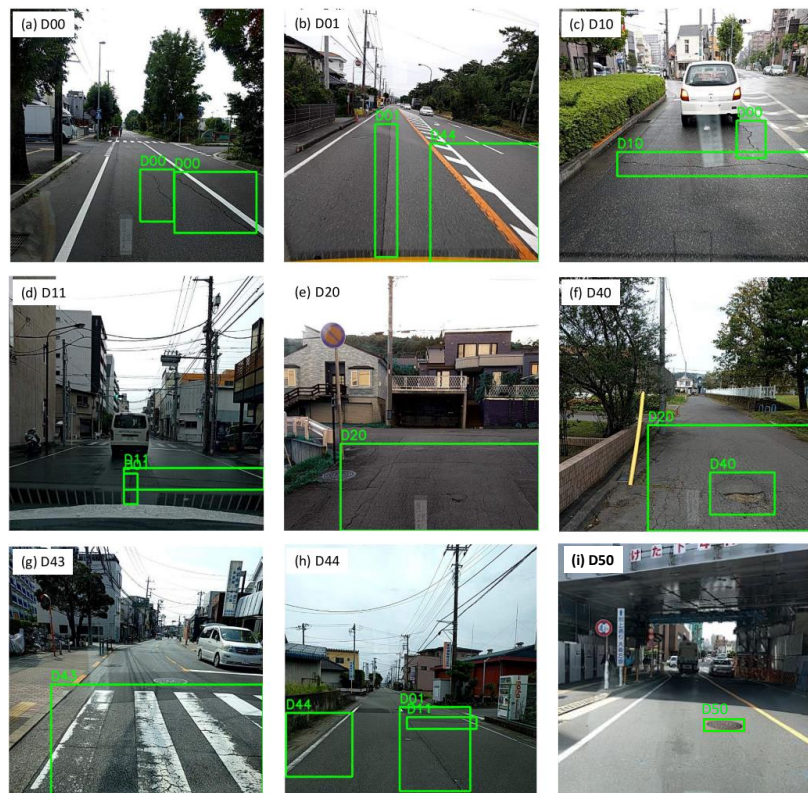


Figure 4-2. Sample of Road Damage Dataset 2019 (RDD2019) from (a) to (i) the nine different defects are represented.

Since this work is exclusively focused on the blurred line defect the RDD2019 dataset [140] was truncated to use only the images containing D44 damage type. Those are 3,290 images and 4,104 labels collected in Japan with a resolution of 600×600.

4.2.2.2. Our dataset: Ceit Road Damage Dataset (CRDD)

On the other hand, new images were collected for this study on Spanish roads using an onboard RGB fish-eye camera. These images were recorded between February 2020 and September 2021 in the territory of Bizkaia, Basque Country (Spain), saving some months in which the maintenance vehicle was not operational. Most of them show highway scenes. Their resolution is 1280×720 and was captured at an average speed between 80 and 120 km/h. This dataset provides also new complex situations, there are daylight and night images considering a wide variety of light and weather conditions. In addition, these images have the added difficulty that the camera is installed on the roof of the vehicle and therefore, as there is no windscreen, droplets or other artefacts may appear (see Figure 4-3).



Figure 4-3. Examples of Ceit Road Damage dataset in different light and weather conditions. Top right sunny with shadows, bottom right rain, top left sunrise and bottom left cloudy.

Firstly, a preliminary cleaning was applied, where invalid images were discarded, either because the vehicle is on the breakdown lane, because the eyepiece was excessively dirty or because they were overexposed. Next, the remaining 6101 images were manually labelled using the *LabelImg* tool [142] and saved in PASCAL VOC format. After all, the set of images proposed in this study has 971 images and 1,262 labels of defect D44.

It is necessary to clarify that as a result of different experiments carried out for this study, it was decided to select a subset of the Spanish image set by discarding some of the more complex images, mainly those containing night scenes. Additionally, in order to maintain the resolution of the input images for the neural networks and not distort the defect area, a square crop of 720×720 was also applied taking into account that in most of the samples the road defects are located in the central part of the image (see Figure 4-4). After these changes, the subset used for this work has 879 images and 1,132 labels as it is shown in Table 4-4.

Table 4-4. Distribution of used datasets: RDD2019 for D44 defect, Ceit Damage Dataset and its simplified subset.

Dataset	Country	Number of images	Labels	Resolution	Night scenes
RDD2019 (D44)	Japan	3290	4104	600×600	-
CRDD	Spain	971	1262	1280×720	✓
CRDD simplified	Spain	879	1132	720×720	-

It should be stressed that after the tests carried out throughout this work, this arrangement to maintain the resolution of the images, as well as a good selection of the transformations applied in the dataset augmentation phase, is vital for the good performance of the models.

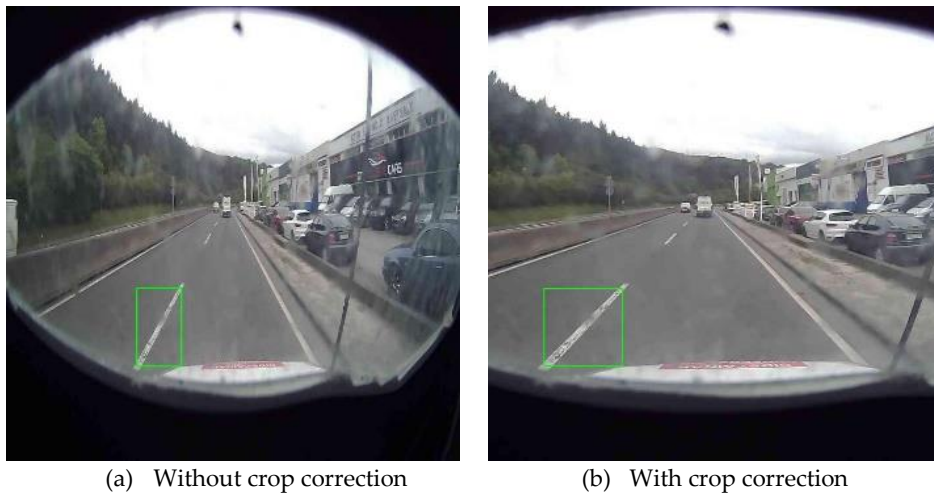


Figure 4-4. Applied crop correction for keeping the resolution of the input image in Ceit Road Damage dataset.

Lastly, it is worth mentioning that this study will also sum up both datasets and create a new mixed one to evaluate the influence of multiple-source models.

4.2.3. Developments

First, the problem was tackled using classical computer vision techniques. However, the results were not good and it was switched to an end-to-end deep learning approach.

The hardware used in these developments is a Windows 10 PC with an Intel Core i7 processor, NVIDIA GeForce RTX 3080 10 GB GPU and a total RAM of 32 GB. The software used for the classical computer vision approach was Matlab 2020a (Image Processing Toolbox V11.1 & Computer Vision Toolbox v9.2) and Python 3.7.9. Whereas for the deep learning approach Python 3.8.13, Tensorflow v2.8.0 together with the Tensorflow 2 Object Detection API were employed.

4.2.3.1. Classical computer vision-based

This is a hybrid algorithm which uses first a pre-trained CNN model for road lanes detection and applies classical machine vision techniques for paint

condition assessment. Figure 4-5, shows the defined steps to detect and evaluate road lane conditions.

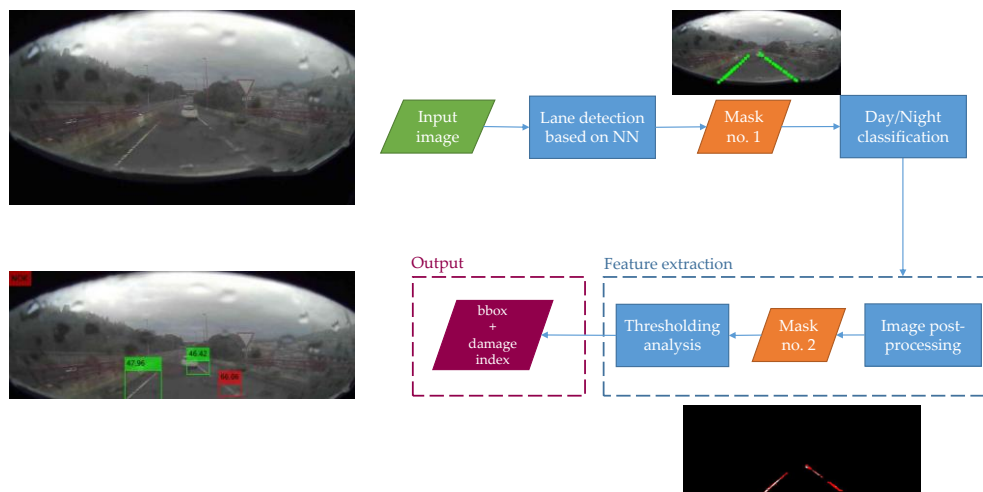


Figure 4-5. Flow chart of the classical computer vision-based approach for road lanes quality assessment.

Step 1: Detection of lanes in the scene. For this purpose, a pre-trained CNN-based model is used called PINet_new, which is based on PINet model [143] and has been optimized. This network was trained with the CULane dataset which includes different urban roads, rural roads, and highways with different light and weather conditions including night images or crowded roads. The output of the CNN gives segmentation of the road lanes given by a dotted green line (see Figure 4-6).



Figure 4-6. Machine learning-based lane condition assessment. Step 1, lane detection.

Step 2: Lane mask generation. This step applies classical vision techniques such as morphological operations to adapt the output of the CNN model and obtain a mask suitable for the study of lane quality (see Figure 4-7).

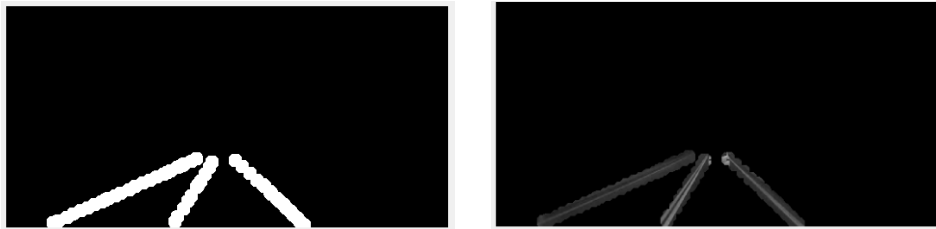


Figure 4-7. Machine learning-based lane condition assessment. Step 2, mask generation.

Step 3: Day/night image classification. The classifier consists of an HSV colour space analysis of the top half of the images. Several experiments have shown that the V value can differentiate between these two cases, with night scenes being those with an average V value of less than 100. This step is necessary to separate the night and day images since each scene contains different characteristics. The night images are pre-processed to improve the light conditions so that the subsequent stage of the algorithm works correctly for both cases (pre-processed day/night) (see Figure 4-8).

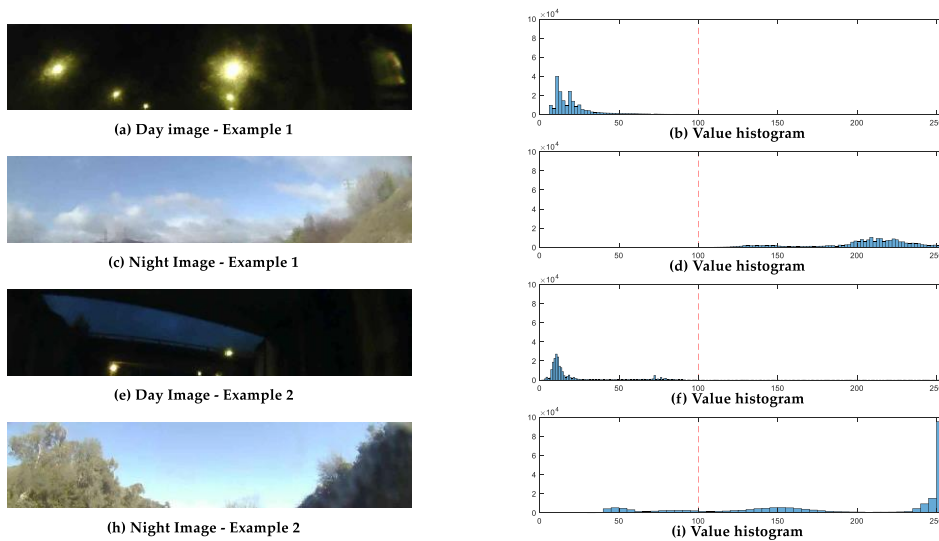


Figure 4-8. Machine learning-based lane condition assessment. Step 3, day/night scene classifier.

Step 4: Analysis of the mask's features and detection of deteriorated areas. In this last step, the condition of the paint of the detected lanes is classified using thresholding techniques. For this aim, the image is segmented into 3 levels using Otsu's method to next evaluate each area independently. In this analysis, the pixels that meet the conditions to be categorized as "bad" are counted. However, since the segmentation always takes part of the asphalt as well as paint, different pixel weights are set depending on whether the pixels are from inside or outside of the segmented area. Finally, each detected lane is classified as OK/NOK according to the proportion of pixels defined as "bad" (see Figure 4-9).

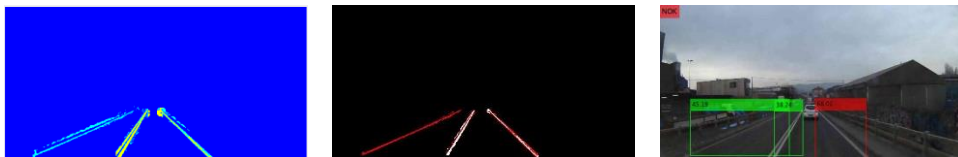


Figure 4-9. Machine learning-based lane condition assessment.
Step 4, analysis of the mask's features.

Evaluation parameters

The results of this algorithm will depend on the good performance of the first block, the detection of lines using the PINet model. No metric has been used to evaluate the segments detected in this first step, the analysis of the results is explained in detail in section 4.3.1.

4.2.3.2. Deep Learning-based

Data augmentation

Since the Ceit-TSR dataset did not contain as many images with defects as expected, and given that the benefits of enhancing this type of image have been demonstrated [144] [145], it was decided to apply this pre-processing to generate "synthetically" novel training data from both datasets. Given that our dataset did not contain as many images with defects as expected, and given that the benefits of enhancing this type of image have been demonstrated [133,144], it was decided to apply some pre-processing to generate synthetically novel training data from both datasets.

For this aim, the *imgaug* library [146] was used, which supports a wide range of augmentation techniques. Six types of image transformation methods were combined paying special attention to ensure that the images will not break by changing them too much. The operations used are listed below and shown in Figure 4-10: horizontal flip, linear contrast, multiply, additive brightness, additive hue and saturation and additive Gaussian noise. The augmenter has been configured in such a way that applies a combination of none to six transformations on the original image random in number and type of operation.

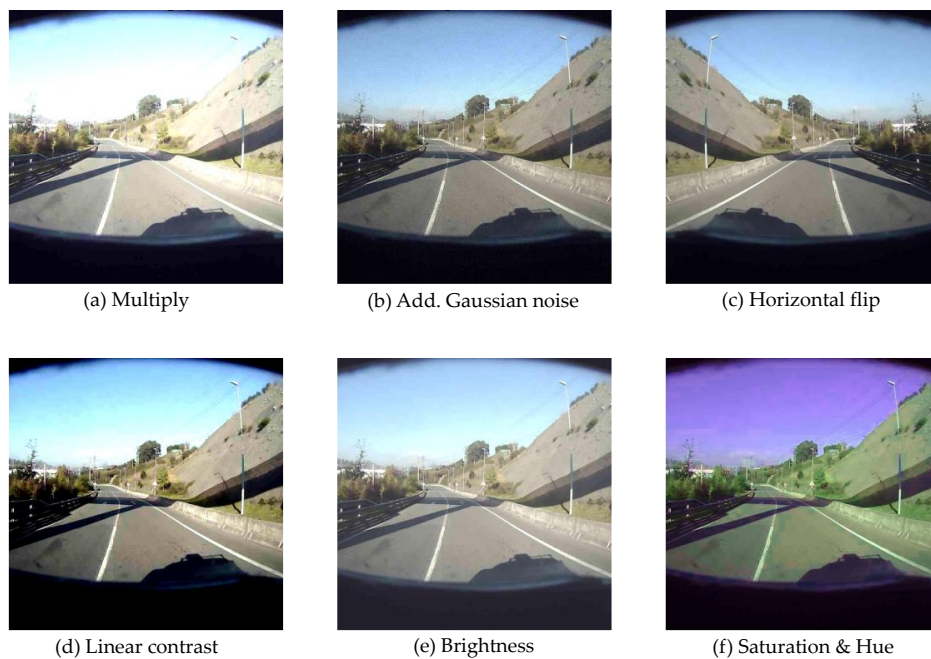


Figure 4-10. Example of the applied six different transformations without combination for the data augmentation.

The objective of this pre-processing is to reach in both training datasets the 5,000 labelled images recommended by [141,147] for each class for an image processing-based classification task to provide satisfactorily accurate results. Thus, the times that an original image has been augmented are different for each dataset. Table 4-5 shows the composition of the new augmented datasets. It is worth noting that before doing the augmentation all datasets were split into 80% training and 20% for test and validation partitions (10%-10%).

Chapter 4 Road Damage Monitoring: road lines

Table 4-5. Resulting used datasets' partition after data augmentation process.

	RDD2019 (D44)		CRDD		Mixed	
	Images	Labels	Images	Labels	Images	Labels
Original Dataset						
Train (80%)	2632	3289	703	903	3335	4213
Test (10%)	329	413	88	112	417	515
Valid (10%)	329	402	88	117	417	508
Augmented Dataset						
Train	5264	6578	4911	6311	6668	8824

Model setup

Given the recent advancements in image content analysis using CNNs, there are several proposed methods for generic object detectors based on deep learning architectures that demonstrate to have a very good performance. Therefore, in this study transfer learning is applied to train and optimize different algorithms to fit our application.

The experiments have been carried out using the open-source Tensorflow environment and its TF2 Object Detection API. This platform offers a collection of detection models pre-trained on the COCO 2017 dataset to facilitate initializing models when training on novel datasets. For this study, the comparative study made by Huang *et al.* [148] was used as a guideline to meet the balance between speed and accuracy of the convolutional object detector use. Finally, the base model Faster R-CNN with Inception Resnet V2 feature extractor was considered. This two-stage detector attained the best possible accuracy achieving the state-of-the-art single model performance at that moment.

However, in the last years, the TF2 Model Zoo has been updated with new single-stage detectors which seems to improve the performance of the above ones. Thus, it is also selected for this study the Single Shot Detector (SSD) with Mobilenet V2 which is one of the most used models in the literature and the EfficientDet D0 as one of the most recent architectures. The overall mAP numbers for these models are shown in Table 4-6.

Table 4-6. Summary of the properties of the studied different object detection models.

Model Name	Speed (ms)	COCO Mean Average Precision (mAP)	Outputs
Faster R-CNN Inception Resnet V2 640×640	203	37.7	Boxes
SSD Mobilenet FPNLite 640×640	39	28.2	Boxes
EfficientDet V1 D0 512×512	39	33.6	Boxes

The major difference between the selected architectures lies in the age of the architectures and the optimisation improvements that have been applied in subsequent iterations. Two-stage detectors such as Faster R-CNN [149] achieve very good accuracy levels, however, they are usually the slowest. In the first stage, the network proposes regions where the object can be found and in the second step it predicts the class of the object. However, single-step detectors (such as SSD [150]) get rid of the first step and explore the network's ability to predict the presence and class of the object. Thus, one-stage detectors have gained in popularity because of their potential to be faster and simpler but they tend to lag behind two-stage detectors in accuracy. However, later on, a new family of object detectors was born called EfficientDet [151], which based on the single-stage detectors, goes deeper into the network architecture design to improve efficiency and find the right balance. They propose a weighted bi-directional feature pyramid network (BiFPN), which allows easy and fast multiscale feature fusion; as well as a compound scaling method that uniformly scales the resolution, depth, and width for all backbone, feature network, and box/class prediction networks at the same time

Object Detection Model

This study carries out the training of different models based on some of the state-of-the-art architectures. These training can be differentiated into two:

- Pure models: those who use images only collected in one country.
- Mixed models: those who merge the database of different sources and countries to seek greater generalisation of the models.

These models will be tested on data collected from the target country as well as on images from different countries.

Evaluation parameters

For the evaluation of the model's performance, the PASCAL VOC 2012 object detection competition evaluation metric [152] was employed. Thus, a correct prediction is the one in which the corresponding predicted bounding box has over 50% Intersection over Union (IoU) with the ground truth bounding box.

With this aim, a code based on the work developed by Padilla R. et al. was used [153]. The current metrics calculated are the Precision-Recall curve and Average Precision. An object detector of a particular class is considered good if its precision stays high as recall increases, this is, when varying the confidence threshold, the precision and recall will still be high. Another way to compare the performance of object detectors is to calculate the area under the curve (AUC) of the Precision-Recall curve. This value is the precision averaged across all recall values between 0 and 1.

In addition, the F1-score metric is also calculated. The F1-score measures accuracy using the statistics of precision and recall.

Finally, the inference time of the trained models has also been measured for comparative reasons.

4.3. Results

4.3.1. Classical computer vision approach

The success of this algorithm depends on the good performance of the initial phase, this is the detection of road lanes. If no lanes are detected, the following analysis will not be possible.

Table 4-7. Results of the machine learning approach for road paint assessment in RDD2019 (D44) dataset and CRDD dataset.

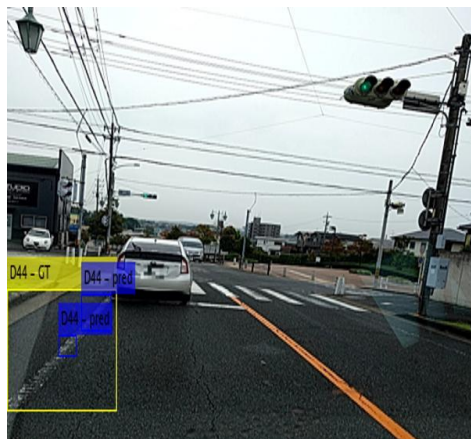
	RDD2019 (D44) dataset	CRDD dataset
Total images	329	88
Overall result		
no-lane	181 (55%)	33 (38%)
NOK	18 (12%)	23 (49%)
OK	130	32
Bounding box result		
TP	4 (1.2%)	3(3.3%)
TN	0	0
FP	43	36
FN	290	53
Processing time (s)		
Lane Detection	136.72	36.73
Lane condition analysis	17.08	6.04
Total proc. time / frame	0.47	0.49

As is shown in Table 4-7 about 55% and 38% of the images in datasets RDD2019 (D44) and CRDD were discarded in this first phase because PINet was not able to detect any lane. This may be because the lanes are too degraded to be detected or because of the camera's perspective with which the algorithm has not been trained.

In the next phase, the detection of lanes in poor condition has been assessed in two ways. On the one hand, the result of the global scene was analysed, this is, whether there is an alert in the image that the paint needs to be repaired (taking into account that all the images in the dataset should contain at least one). And on the other hand, a more exhaustive evaluation in which the bounding boxes of the ground truth have been contrasted with those calculated by the algorithm.

Chapter 4 Road Damage Monitoring: road lines

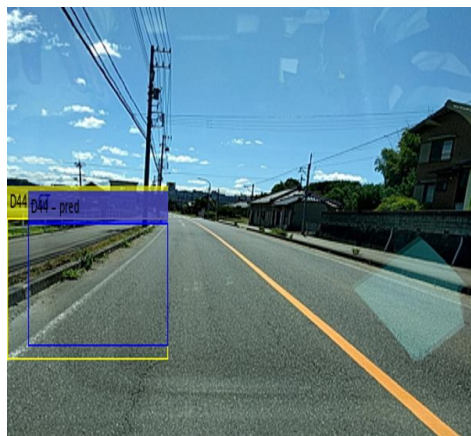
In the first analysis, 12% and 39% of the scenes contain a warning for RDD2019 (D44) and CRDD datasets respectively. And when the analysis focuses on the bounding boxes the result worsens. Just 4 (1.2%) and 3 (3.3%) bounding boxes for RDD2019 (D44) and CRDD datasets respectively were detected correctly. This may be due to the difficulty of modelling pixels in need of repair for any kind of environmental conditions. There are also situations where a crack or joint has been detected as a lane. In addition, the size of the bounding boxes calculated by the ML algorithm is usually smaller than those of the ground truth so the IoU does not meet sufficient conditions to consider the result valid (see Figure 4-11).



(a) Small detected bounding boxes



(b) Pedestrian pavement detected



(c) Good result of RDD2019 (D44) dataset



(d) Good result of CRDD dataset

Figure 4-11. Visual results of the ML-based road paint damage detection

Due to the bad behaviour of this algorithm, this thesis proposes another way of addressing the problem using deep learning techniques that are explained in the following section.

4.3.2. Deep learning approach

4.3.2.1. Object detection models

As it was advanced in the previous section this work studied two different neural network model architectures presented in Table 4-6.

Regarding the fine-tuning of these networks, several hyperparameters were explored to optimize the pre-trained model that will be used. Finally, the following configuration was selected and defined on the different pipelines. Since the images available of the different datasets were originally in different sizes all input images were re-scaled to 512×512 size. All models were trained for 25,000 steps. The SGD optimizer was used with a momentum of 0.9 and the L2 regularized or weight decay was fixed to 3.9E-05. However, both learning rate and batch size were tuned for each model and dataset as shown in Table 4-8.

It is worth noting that architectures like Faster RCNN Tensorflow Object Detection API encounters memory consumption problems and for the resources of our machine, the batch size should be reduced considerably to avoid OOM (Out Of Memory) errors, however, for SSD-like architectures, a little increase on this number was permitted.

Table 4-8. Summary of the hyperparameters used on the different object detection models trained in this section.

Network architecture	Model version	Hyperparameters	
Faster RCNN Inception Resnet V2	V1	input size	512×512
		batch size	2
		num steps	25000
		optimizer	SGD
		momentum	0.9
		regularizer L2	3.90E-05
		learning rate	Japan 0.07
		(cosine decay)	Spain 0.0003
			Mixed 0.07

Chapter 4 Road Damage Monitoring: road lines

SSD Mobilenet V2	V1	input size	512×512
		batch size	18
		num steps	25000
		optimizer	SGD
		momentum	0.9
		regularizer L2	3.90E-05
		learning rate	Japan 0.07
		(<i>cosine decay</i>)	Spain 0.0003
			Mixed 0.07
EfficientDet V1 D0	V1	input size	512×512
		batch size	8
		num steps	25000
		optimizer	SGD
		momentum	0.9
		regularizer L2	3.90E-05
		learning rate	Japan 0.07
		(<i>cosine decay</i>)	Spain 0.0003
			Mixed 0.07
EfficientDet V1 D0	V5	input size	512×512
		batch size	8
		num steps	25000
		optimizer	SGD
		momentum	0.9
		regularizer L2	3.90E-05
		learning rate	Japan 0.07
		(<i>exponential decay</i>)	Spain 0.0003
			Mixed 0.07

4.3.2.2. Performance of the models for different countries

Table 4-9 presents the F1-score and mAP values obtained for the detection of D44 damage type for all the experiments conducted in this work. A total of eight pure models and four mixed models were trained. Marked in bold, the best results obtained for each country and model type based on the F1-score value are shown.

In general, it is observed that models work better for the Japanese dataset. Comparing the best two pure models it is shown that the best Japanese pure model achieves 75% of mAP while the best Spanish pure model's mAP is less than 30%, both tested with their target country. This may be because the Spanish images are more complex and although they were augmented, there were originally fewer samples.

Nevertheless, it should be noted that training with the mixed dataset clearly improves the model performance for the two countries, although the difference for the Spanish images is much more noticeable (an improvement from 27.73% to 83.79%). It is therefore clear that building a dataset containing images from different sources greatly enriches the dataset and helps the generalisation of the trained models.

Table 4-9. F1-score and mAP (mean average precision) for the white line blur detection and each used dataset.

Model name	test set	F1	mAP	inference time (ms)
frcnn_inception-resnetv2_japan	Japan	0.743	65.61%	191.582
	Spain	0.197	6.50%	189.726
frcnn_inception-resnetv2_spain	Japan	0.270	17.75%	193.120
	Spain	0.390	25.05%	190.478
frcnn_inception-resnetv2_mixed	Japan	0.860	86.85%	193.519
	Spain	0.600	58.27%	189.524
ssd_mobilenetv2_japan	Japan	0.755	71.07%	25.171
	Spain	0.220	8.00%	21.217
ssd_mobilenetv2_spain	Japan	0.120	3.20%	25.102
	Spain	0.335	18.62%	20.897
ssd_mobilenetv2_mixed	Japan	0.920	92.48%	24.690
	Spain	0.848	77.94%	21.565
efficientdetv1_d0_japan (V1)	Japan	0.830	75.00%	47.049
	Spain	0.194	6.91%	44.464
efficientdetv1_d0_spain (V1)	Japan	0.235	11.90%	49.159
	Spain	0.422	27.73%	44.281
efficientdetv1_d0_mixed (V1)	Japan	0.790	92.34%	45.400
	Spain	0.934	83.79%	43.901
efficientdetv1_d0_japan (V5)	Japan	0.785	76.07%	47.816
	Spain	0.238	8.39%	43.542
efficientdetv1_d0_spain (V5)	Japan	0.233	8.46%	47.267
	Spain	0.418	30.69%	43.878
efficientdetv1_d0_mixed (V5)	Japan	0.929	91.74%	47.361
	Spain	0.828	78.97%	44.363

It is worth mentioning that the F1-score results obtained in this paper with the mixed models exceed the results reported in the state of the art for both Spanish and Japanese images.

4.3.2.3. Empirical analysis

Figure 4-12 shows visually the F1 scores in the previous table.

It can be seen that the tendency of the results is very similar for each model. This is, the mixed model obtains better results than the pure one and models trained with Japan dataset perform better than models trained with the Spanish dataset. In addition, for pure models, the performance with target images is much better than the performance with images from a different country.

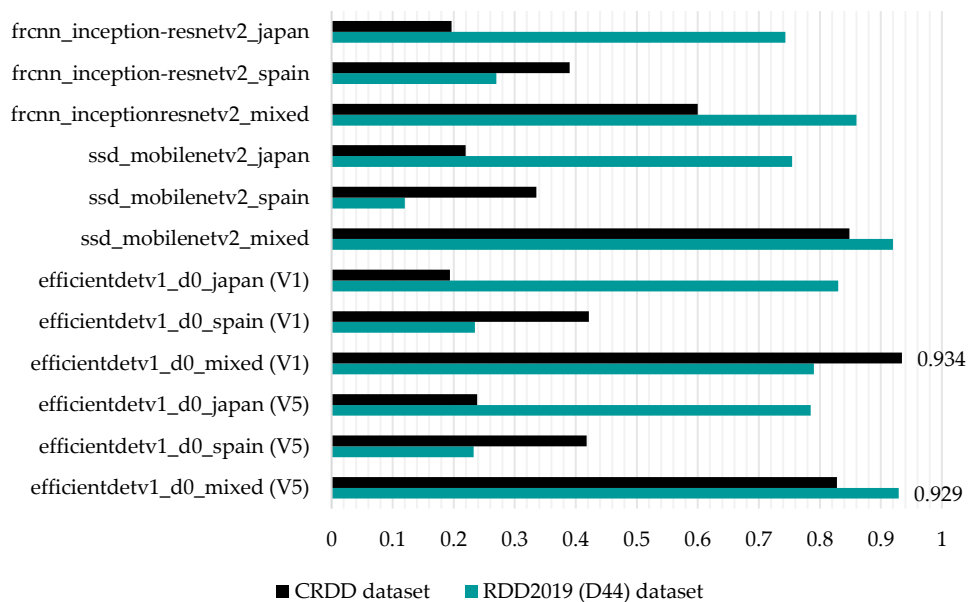


Figure 4-12. F1-score summary results for the models and datasets considered in this chapter

Regarding the comparison between the different models, it can be noted that the ones based on EfficientDet perform somewhat better than the other architectures. Besides that, concerning inference time there is a clear difference between single-stage detectors and two-stage detectors, being the last ones the slowest. However, if the system were time-critical, it should also be noted that within the single-stage models, the ones based on SSD Mobilenetv2 could predict in half the time of those based on EfficientDetV1 D0 (20 milliseconds/image faster).

4.3.2.4. Visual analysis

The results presented in the last section are promising. However, in this section, the results of the predictions will be analysed visually in order to better understand the differences between the two datasets and the failures of the detectors. Note that in the next images the red box corresponds to the ground truth label with the green box illustrating the prediction done by the detector.

For the visual analysis, some images of the inference of four models marked in bold in Table 4-9 were extracted.

To predict the Japanese dataset's defects:

- Pure model: `efficientdetv1_d0_japan (V1)`
- Mixed model: `efficientdetv1_d0_mixed (V5)`

To predict the Spanish dataset's defects:

- Pure model: `efficientdetv1_d0_spain (V1)`
- Mixed model: `efficientdetv1_d0_mixed (V1)`

Generally, mixed models present higher confidence score detections than pure models. In addition, these models detect more defects and their bounding boxes fit better the area of interest (see Figure 4-13).

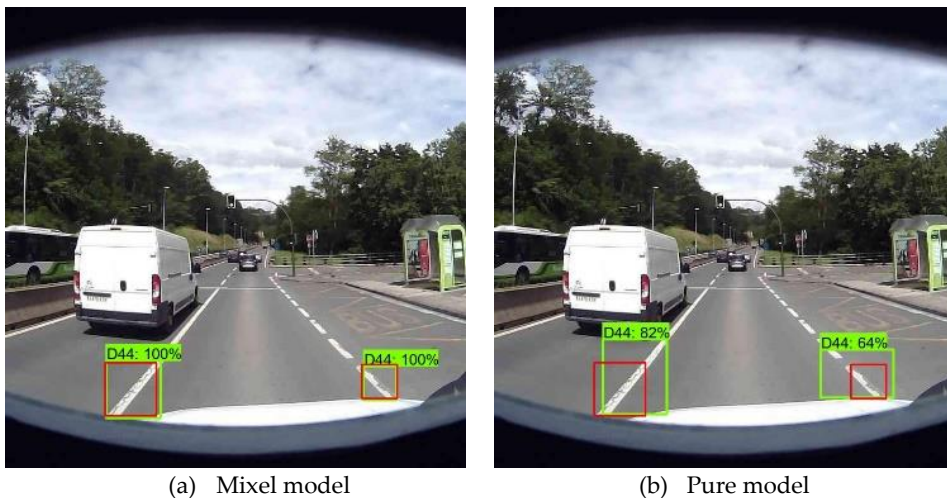


Figure 4-13. Comparison of the detection bounding boxes for a mixed and `efficientdetv1_d0_spain (V1)` pure model in CRDD dataset sample.

Chapter 4 Road Damage Monitoring: road lines

On the other hand, the comparison of the two databases shows that the images from Spain dataset (see Figure 4-14) present more complex situations than Japan dataset (see Figure 4-15).

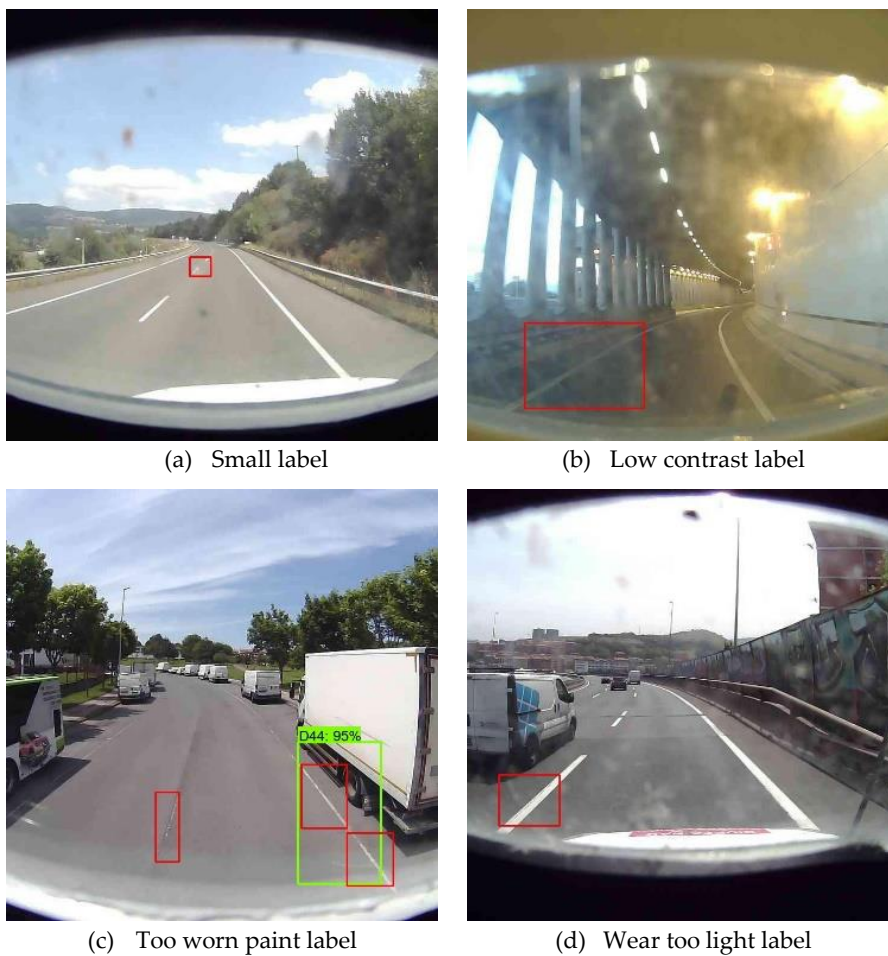


Figure 4-14. Difficult labels of CRDD dataset where the detector has missed the defect.

Figure 4-14 shows some of the difficulties identified in CRDD datasets, such as:

- smaller labels;
- dirtiness of the glass used in the on-board system;
- lower contrast of the images;

- a wider range of labelled paint wear. Samples with paint too worn to be detected or samples where the wear is too light to be detected;
- the camera setup is not focused on the asphalt zone. Large focal length.



Figure 4-15. Samples from the RDD2019 (D44) dataset.

These factors plus the fact that this dataset was originally smaller may influence the performance of the detectors for these images to be somewhat lower than for the Japanese dataset.


4.4. Discussion

Good quality of the road infrastructure is key for road safety but also of vital importance for the future's autonomous driving. Potholes, cracks etc. affect driving comfort, safety and make our driving much less efficient. But poorly maintained road signals or markings difficult to interpret the road and the road rules that govern the coexistence of all road users and seriously jeopardise our safety whether the vehicle is autonomous or human-guided. This chapter has discussed road line monitoring as a key element for many of the ITS that is based on lane-keeping on the intelligent or autonomous vehicles

However, road marking damage detection is a task that has hardly been addressed in the literature due to its complexity and the lack of sufficient images. Therefore, this thesis provides a set of 971 images labelled for this specific defect and considers high complexity scenarios to be used in the future. In addition, this dataset provides the scientific community with new images from other countries that was not covered by the current datasets. Several experiments were carried out for three different architectures, obtaining a promising performance of an F1-score value higher than 0.92, which exceeds the results reported in the state of the art (F1-score 0.743) by 25%.

It should be noted that for this task it is vital to have both quantity and quality of data. It is very important to have enough images that represent all possible scenarios (or countries) as well as the labelling task to select the data to be trained with criteria so that they do not confuse the neural network. It is also crucial the pre-process the images and do a good data augmentation that does not distort the original image (for this particular defect, maintaining the resolution of the image was indispensable).

Chapter 5
Application in Connected,
Cooperative and
Automated Mobility

A teal-colored triangle pointing to the left, positioned to the right of the chapter title.

5.1. Context

Until a few years ago, the vehicle was considered an individual element on the road network. It was governed solely by the sensors and technologies installed in it. And to this was added the perception of the user who was driving it, which was limited to the visual horizon of the road area in which it was located. However, vehicular communications have changed this paradigm, creating a much more cooperative and enriched environment. These communications allow the exchange of information in real-time between different elements and road users. In this way, the perception of the environment is enhanced as the vehicle can receive information about the manoeuvres of the users around it even before they are in its field of vision. In addition, it allows the vehicle to cooperate with the infrastructure (V2I) and other road users (V2V, V2P) and to anticipate situations further ahead on its route such as road works, accidents, fog banks, etc. Vehicular communications in road transport have involved new systems and services that help improve road safety and driving efficiency, moreover, they will be an essential part of autonomous driving.

5.1.1. ITS value chain

In the same way, vehicular communications have also evolved in the last 25 years. Originally, the vehicle as an individual element was connected directly to the traffic control centre. However, in the meantime, other actors have appeared in between (see Figure 5-1).

- Service Provider (SP): this actor came in due to the first digitalization step of road traffic information. The SP exchange the information in real-time between the traffic centre and the vehicle, an electronic format is needed to automate these services.
- Content Aggregator (CA): more and more data sources emerged and thus SP needed a CA who combines traffic information with other kinds of relevant information for road users. Here, the need for standardization of the electronic processable data became obvious and need to be used by all the actors in the ITS value chain.
- National Access Points (NAP): From this need to standardize information, the NAPs arise from the delegated acts on ITS. This

actor gathers several types of traffic information and provides traffic data according to European standards. These EU regulations aim to accelerate the development of interoperable services to end-users throughout Europe.

- Service Access Point (SAP): this actor emerged in the domain of advanced traffic information and individual navigation services intending to manage the information provided to the road user and vehicle from different service providers efficiently and effectively. Therefore, the devices used by cars and travellers do not need a direct relationship with the service providers anymore.
- Road Side Unit (RSU): this last actor came in to connect the vehicle with the traffic centre (and close the circle), this is direct digital communication between road operators and vehicles. The C-ITS technology enables this and RSUs are developed to support this.

Technological innovations make that all actors in this ITS value chain can communicate with each other. However, consistency and reliability of information throughout the value chain need to be ensured. For this aim different standards arise for the different domains that are involved here.

- Road operators domain: which involves from RSUs to the NAPs. They have developed an extensive standard called DATEX II. In DATEX II all relevant details of traffic information and management can be expressed. The focus of this standard is on providing an as complete information view of the road as feasible.
- Content and service aggregators domain: which involves NAPs to SAPs. Here the leading standard is TPEG, which was developed to enable the consistent provision of traffic and traffic-related information to end-users across several media.
- Service providers and OEMs: that involves the vehicle and its connections to the ITS value chain, the SAPs and RSUs. The main standards here are the C-ITS standards developed by ETSI.

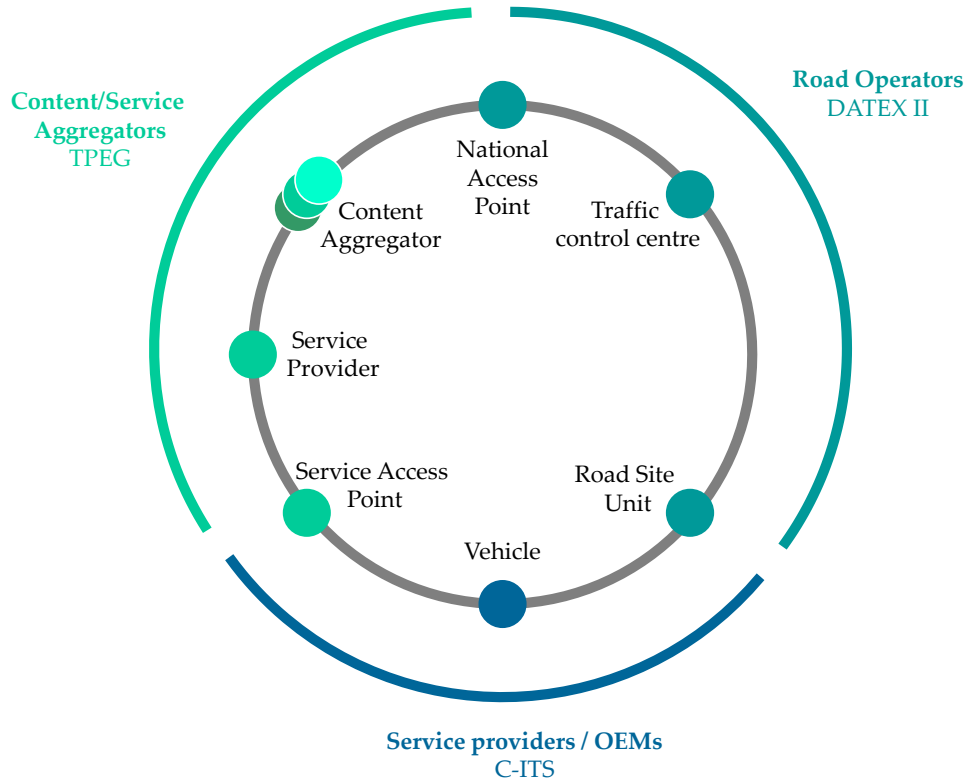


Figure 5-1. ITS value chain representation. It contains all involved actors and domains with its leading standards.

5.1.2. Data exchange via cooperative V2X communication

In a cooperative road traffic scenario, cooperative V2X communication units – so-called ITS stations – are implemented in vehicles and traffic infrastructure, and exchange data with each other via the cooperative V2X short-range ad-hoc network.

Nowadays, the vast majority of new vehicles are equipped with a navigation system and different sensing technologies. And every so often, onboard units in the vehicles broadcast data such as their position, speed and driving direction to their surroundings. Additionally, they send out event-triggered messages about special incidents, such as an emergency braking, a vehicle defect or a slippery road detected. As an example, the vehicles receiving such positioning information can calculate where the cars are going to be, predict whether a hazardous situation of a crash could occur and consequently warn

the driver to react in this situation (V2V). Similarly, an event detected by the vehicle such as a slippery road can be transmitted to the road site units of the infrastructure so that they can alert other vehicles that are about to pass through the same area (V2I).

In addition to intelligent vehicles, the road infrastructure is also starting to incorporate more and more technology (speed sensors, acoustic sensors, IP CCTV cameras, smart traffic lights, condition/weather monitoring systems, digital signage, etc.) and thus, getting smarter. The I2V communication is also used when the so-called roadside units in the traffic infrastructure inform e.g. about signal phases of traffic lights, speed limits or road works. This way, for example, traffic lights can advise the driver about green or red lights and inform of an adequate speed to find it open, thus influencing the driving behaviour to be more efficient (see Figure 5-2).

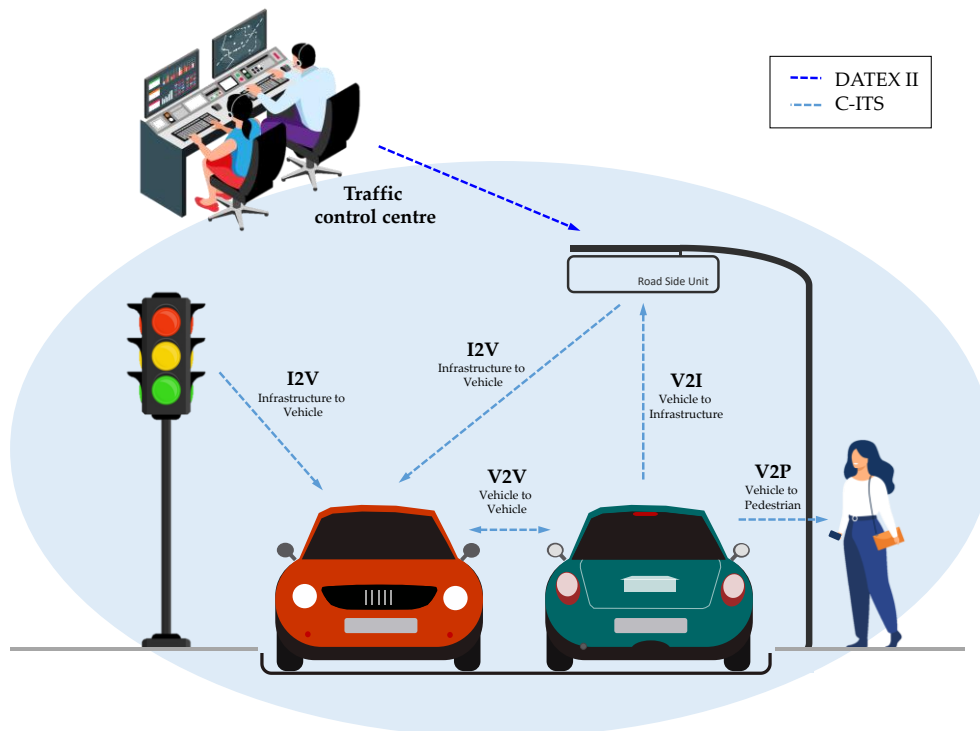


Figure 5-2. V2X communication ecosystem diagram plus the communication with the Traffic Control Centre.

Chapter 5 Application in Connected, Cooperative and Automated Mobility

Throughout the development of this thesis, V2I and I2V communications were worked on, this is, in the service provider and road operator domains. For this purpose, the C-ITS standard has been studied to transmit information obtained by the monitoring vehicle to the RSU or vice versa. On the other hand, the Datex II standard has also been analysed to communicate events of other sources from the traffic control centre to the RSU.

5.2. Message handling

In the previous chapters, the monitoring of the road using artificial vision techniques has made it possible to detect fog banks, traffic signs and damaged road markings. These events can be of great use for cooperative and connected mobility and it is therefore important to know which standards allow this information to be transmitted in a harmonised way so that other road users (V2V) or road operators (V2I/I2V) can interpret it.

5.2.1. C-ITS standard

C-ITS is a standard managed by the ETSI TC ITS committee which defines specifications for Co-operative ITS, which offers enormous potential through vehicle-to-vehicle and vehicle-to-roadside communication. Applications include road safety, traffic control, fleet and freight management and location-based services, providing driver assistance and hazard warnings and supporting emergency services. These specifications are crucial for the commercial deployment of the technology

TC ITS develops standards related to the overall communication architecture, management and security as well as the related access layer agnostic protocols: the physical layer, Network Layer, Transport Layer and Facility Layer.

In this thesis and concerning the services in the Facility layer for the communication between infrastructure and traffic participants there are two C-ITS messages to highlight that may represent the type of events that would be generated in the previous chapters:

- Decentralized Environmental Notification Message (DENM) was defined as a Basic Set of Applications (BSA) for Road Hazard Warning (RHW) application to alert road users of a detected event. It is composed of multiple use cases (see Appendix D). Its technical specification reference is ETSI EN 302 637-3.
- Infrastructure to Vehicle Information message (IVIM) that supports mandatory and advisory road signage such as contextual speeds and road works warnings. IVIM either provides information on psychical

road signs such as static or variable road signs, virtual signs or road works. Its technical specification is ETSI TS 103 301.

5.2.1.1. Weather Condition

Bad weather conditions such as fog banks are contemplated in the DENM messages as reduced visibility warning events categorized as *adverseWeatherCondition-Visibility* (see Appendix D).

The next DENM message shows an example of a fog bank event (cause code 18 – subcause code 1) with a range of 1000 meters that affects all traffic directions in the A-8 road passing through the municipality of Eibar (43.195055, -2.437543). The event was created the 06/07/2022 8:41:45 a.m. and has a validity duration of 10 minutes (see Message 5-1).

Message 5-1. DENM message example for a fog bank event.

```
{
  "header": {
    "protocolVersion": 1,
    "messageID": 1,
    "stationID": 3494000000
  },
  "denm": {
    "management": {
      "actionID": {
        "originatingStationID": 3494000000,
        "sequenceNumber": 0
      },
      "detectionTime": 1657089705,
      "referenceTime": 1657089705,
      "eventPosition": {
        "latitude": 43195055,
        "longitude": -2437543,
        "positionConfidenceEllipse": {
          "semiMajorConfidence": 1,
          "semiMinorConfidence": 1,
          "semiMajorOrientation": 0
        },
        "altitude": {
          "altitudeValue": 0,
          "altitudeConfidence": "alt-000-01"
        }
      },
      "relevanceDistance": "lessThan1000m",
      "relevanceTrafficDirection": "allTrafficDirections",
      "validityDuration": 600,
      "stationType": 5
    },
    "situation": {
      "informationQuality": 0,
      "eventType": {
        "causeCode": 18,
        "subCauseCode": 1
      }
    }
  }
}
```



5.2.1.2. Traffic Signs

C-ITS standard allows representing the traffic sign information with IVIM messages. Usually, they are composed of two different containers, the General Ivi Container (GIV) which defines the *ivitype* and the road sign code, and the Geographic Location Container (GLC) where the location of the traffic sign is detailed.

The Pictogram category code which defines the kind of traffic signal represented in the message is defined in ISO 14823:2017. These traffic signs are separated into three main category types (see Figure 5-3):

Service category code		Category code	Pictogram category code
Category number	Sub category number	Category number	
1: Traffic sign	1: Danger warning	1-9: Danger warning	
		1-3: Priority	
		4-6: Prohibition or restriction	
	2: Regulatory	7-9: Mandatory	
		1-3: Advance direction	
		4: Direction	
	3: Informative	6: Lane guidance	
		7-9: Road/place identification	
		Serial number (1-99)	
2: Public facilities	1: Public facilities	1-9: Public facilities and services	
3: Ambient/road condition	1: Ambient condition	1-9: Ambient conditions and nature	
	2: Road condition	1-9: Road condition and nature	

Figure 5-3. Table extracted from ISO 14823:2017(E) where general category codes are defined [154].

- **Traffic sign:** traffic signs are officially established pictograms in each country to control traffic using warning, regulatory or informative signs.
- **Public facilities:** public facilities indicate the existence of certain public facilities and their service details (e.g. toilets, restaurants, hospitals, etc.).

Chapter 5 Application in Connected, Cooperative and Automated Mobility

- Ambient/road conditions: they are concerned with the ambient condition of the roadway or local conditions which may affect the flow of road traffic (such as bad weather and traffic congestion). The next message shows an example of a *regulatoryMessages* (1) ivi type which contains a 100 km/h speed limit event in the A-8 road reaching the Zarautz toll (43279035, -2.152400) (see Figure 5-4). It was created on at 06/07/2022 10:38:18 am and has a validity duration of 24 h (see Message 5-2). For this specific case the category code is constructed as follows (see Table 5-1):



Figure 5-4. Representation of the traffic sign that is codified in the IVIM message example.

Table 5-1. Breakdown of the codes defining the speed limit sign according to ISO 14823:2017 (E) [154].

Service category		Pictogram category		Full name
Category number	Sub category number	Nature	Serial number	
1: Traffic sign	2: Regulatory	5	57	Maximum speed limited to the figure indicated

Message 5-2. IVIM message example for a speed limit sign.

```

{
  "header":{
    "protocolVersion":1,
    "messageID":1,
    "stationID":0
  },
  "ivi":{
    "mandatory":{
      "serviceProviderId":{
        "countryCode":"8500",4
        "providerIdentifier":3495
      },
      "iviIdentificationNumber":1,
      "timeStamp":1657096698,
      "validFrom":1657096698,
      "validTo":1657183098,
      "iviStatus":1
    },
    "optional":[
      {
        "giv":{
          {
            "iviType":1,
            "roadSignCodes":[
              {
                "code":{
                  "iso14823":{
                    "pictogramCode":{
                      "serviceCategoryCode":{
                        "trafficSignPictogram":"regulatory"
                      },
                    "pictogramCategoryCode":{
                      "nature":5,
                      "serialNumber":57
                    }
                  }
                },
                "attributes":[
                  {
                    "spe":{
                      "spm":100,
                      "unit":0
                    }
                  }
                ]
              }
            ]
          }
        }
      ]
    ]
  }
}

```

⁴ Country code follows the standard ISO 3166-1:2020 Codes for the representation of names of countries and their subdivisions — Part 1: Country code. The representation for Spain is ES that using ITA2 Baudot-Murray code is codified as 1000010100 in binary and 8500 in hexadecimal.

```
    ]
  }
]
},
{
  "glc":{
    "referencePosition":{
      "latitude":43279035,
      "longitude":-2152400,
      "positionConfidenceEllipse":{
        "semiMajorConfidence":1,
        "semiMinorConfidence":1,
        "semiMajorOrientation":0
      }
    },
    "altitude":{
      "altitudeValue":0,
      "altitudeConfidence":"alt-000-01"
    }
  },
  "parts":[
    {
      "zoneId":1
    }
  ]
}
}
}
```

It should be noted that such C-ITS messages do not make sense in the vehicle-RSU direction but could be transmitted in the other direction in the case of the In-Vehicle Signage (IVS) service. IVS provides information about existing, fixed and dynamic traffic signs to passing vehicles employing IVI messages. For the practical case where the maintenance vehicle equipped with a signal recognition system is being used to generate/update a signal inventory, no communication standard has been established.

5.2.1.3. Road Damage

There is no C-ITS message defined for the exchange of road marking damage event data. However, other road damages such as subsidence or burst pipe are contained in DENM messages for *hazardousLocation-SurfaceCondition* cause code (see Appendix D).

The next DENM message example shows the case of a hazardous location event due to subsidence damage on the road surface (cause code 9 – subcause code 4) of less than 200m in the GI-3440 mountain road (43.330319, -1.898631). The message was created on at 06/07/2022 13:19:44 a.m and the

validity duration of this event is 24 h. Just one lane or traffic direction is affected (see Message 5-3).

Message 5-3. DENM message example for subsidence damage on the road surface.

```
{
  "header":{
    "protocolVersion":1,
    "messageID":1,
    "stationID":3494000000
  },
  "denm":{
    "management":{
      "actionID":{
        "originatingStationID":3494000000,
        "sequenceNumber":0
      },
      "detectionTime":1657106384,
      "referenceTime":1657106384,
      "eventPosition":{
        "latitude":43330319,
        "longitude":-1898631,
        "positionConfidenceEllipse":{
          "semiMajorConfidence":1,
          "semiMinorConfidence":1,
          "semiMajorOrientation":0
        }
      },
      "altitude":{
        "altitudeValue":0,
        "altitudeConfidence":"alt-000-01"
      }
    },
    "relevanceDistance":"lessThan200m",
    "relevanceTrafficDirection":"upstreamTraffic",
    "validityDuration":3600,
    "stationType":5
  },
  "situation":{
    "informationQuality":0,
    "eventType":{
      "causeCode":9,
      "subCauseCode":4
    }
  }
}
```

It should be noted that the subcause code field is an integer that supports a range of 0-255. Therefore, a new cause code (e.g. 10) could be defined to cover this specific case.

```
HazardousLocation-SurfaceConditionSubCauseCode ::= INTEGER {  
    unavailable(0),  
    rockfalls(1),  
    earthquakeDamage(2),  
    sewerCollapse(3),  
    subsidence(4),  
    snowDrifts(5),  
    stormDamage(6),  
    burstPipe(7),  
    volcanoEruption(8),  
    fallingIce(9),  
    roadMarkingDamage (10)  
} (0..255)
```

5.2.2. DATEX II standard

DATEX II is the electronic language used in Europe for the exchange of traffic information and traffic data. The development of DATEX II was initiated in the early 90s because of the need to exchange information between traffic centres of motorway operators. Soon there was the need to open this information to service providers and DATEX I was too limited for this. Therefore, DATEX II was developed in the early 2000s. Employing DATEX II, traffic information and traffic management information are distributed in a way that is not dependent on language and presentation format. This means that there is no room for misunderstandings and/or translation errors by the recipient, but the recipient can choose to include spoken text, an image on a map, or integrate it into a navigation calculation.

DATEX II is a multi-part standard created and maintained by CEN TC278. The content of these specifications can be found in CEN 16157. In addition, there is also an Exchange Specification standard that has led to several options to implement DATEX II content exchange based on different requirements from different DATEX II application fields. This last one can be found in ISO/CEN TS 19468.

The DATEX II data model includes various sub-standards or parts. Part 1 describes the rules of the standard, Part 2 describes the chosen location referencing method and Part 7 describes common information elements. Parts 3 to 6 and 8 to 13 describe the data model for the exchange of information about a certain type of information (see Figure 5-5).

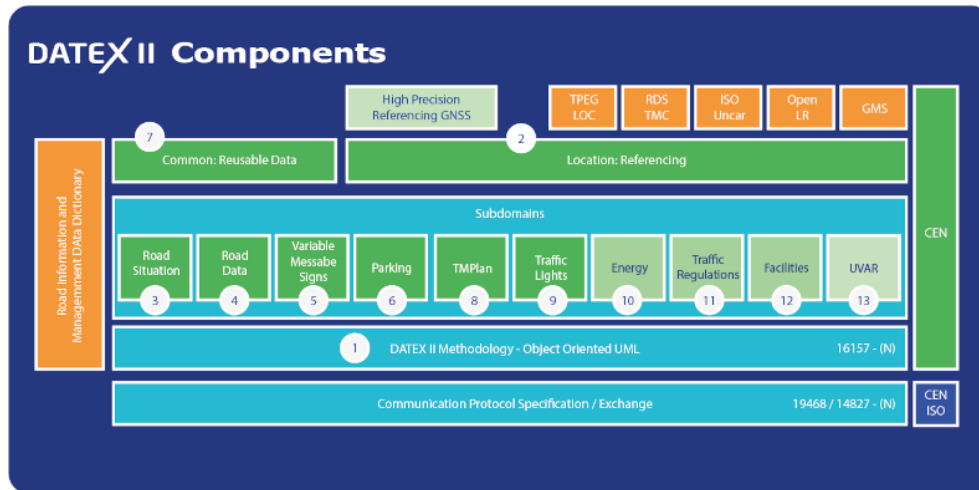


Figure 5-5. DATEX II standard's components schema. Extracted from [155].

In this thesis and for the representation of the event types generated in the chapters below the DATEX II components that will be used are Road Situation for road hazard events and Traffic Regulations for traffic signs. It is will be also necessary for the Common and the Location Referencing components for modelling and to specify their location.

5.2.2.1. Weather condition

For the representation of bad weather conditions such as fog banks, the DATEX II standard contemplates the component Situation – Traffic element where there is a specific section for *PoorEnvironmentConditions* (eg. fog, heavy rain, snowfall, etc.). For this type of event, the location reference type that will be used is *TpegAreaLocation* since this kind of event cannot be specified in more detail. For the specific case of fog representation DATEX II provides a different level definition (see Table 5-2).

The next DATEX II message shows an example of a fog bank event with a coverage area of 1000 meters in the A-8 road passing through the municipality of Eibar (43.195055, -2.437543). The event was created the 2022/07/07 11:16:22 a.m. and has a validity duration of 10 minutes (see Message 5-4).

Chapter 5 Application in Connected, Cooperative and Automated Mobility

Table 5-2. Fog levels are defined by the DATEX II standard for PoorEnvironmentType events.

Situation	PoorEnvironmentTypeEnum	Dense fog	Dense fog, limiting visibility to 50m or less.
Situation	PoorEnvironmentTypeEnum	Fog	Fog, visibility more than 50m.
Situation	PoorEnvironmentTypeEnum	Moderate fog	Misty conditions impair vision over 100m.
Situation	PoorEnvironmentTypeEnum	Patchy fog	Fog, in which intermittent areas of dense fog may be encountered.

Message 5-4. DATEX II message example for a fog bank event.

```
{
  "d2:payload": {
    "@lang": "eng",
    "@modelBaseVersion": "3",
    "publicationTime": "2022-07-07T11:16:22",
    "publicationCreator": {
      "country": "es",
      "nationalIdentifier": "ceit-brta"
    },
    "sit:situation": [
      {
        "@id": "2840",
        "sit:headerInformation": {
          "informationStatus": "technicalExercise"
        },
        "sit:situationRecord": [
          {
            "@id": "2840-00",
            "@version": "1",
            "sit:situationRecordCreationTime": "2022-07-07T11:15:49",
            "sit:situationRecordVersionTime": "2022-07-07T11:15:49",
            "sit:probabilityOfOccurrence": "riskOf",
            "sit:validity": {
              "validityStatus": "active",
              "validityTimeSpecification": {
                "overallStartTime": "2022-07-07T11:15:49",
                "overallEndTime": "2022-07-07T11:25:49"
              }
            },
            "sit:locationReference": {
              "loc:tppegAreaLocation": {
                "loc:tppegAreaLocationType": "other",
                "loc:radius": 1000,
                "loc:centrePoint": {
                  "loc:latitude": 43.195055,
                  "loc:longitude": -2.437543
                }
              }
            },
            "sit:poorEnvironmentType": [
              "fog"
            ]
          }
        ]
      }
    ]
  }
}
```

```

}
  }
    }

```

5.2.2.2. Traffic Signs

DATEX II standard has recently added a new component for traffic regulation where a specific traffic regulation issued by a competent authority can be communicated. In this component, different traffic regulation types are included such as speed limits, road or ambient warnings, traffic jams ahead, etc.

The next DATEX II message shows an event example of a traffic sign limit of 100 km/h located in the A-8 road Bilbao direction (see Message 5-5).

Message 5-5. DATEX II message example for a speed limit traffic sign.

```

"d2:payload": {
  "@lang": "eng",
  "@modelBaseVersion": "3",
  "@id": "",
  "publicationTime": "2022-07-07T16:36:40",
  "publicationCreator": {
    "country": "es",
    "nationalIdentifier": "ceit-brta"
  },
  "tro:trafficRegulationsFromCompetentAuthorities": {
    "tro:trafficRegulationOrder": [
      {
        "@id": "2842",
        "@version": "1",
        "tro:issuingAuthority": {
          "values": {
            "value": [
              "road maintenance operator"
            ]
          }
        },
        "tro:regulationId": "2842-00",
        "tro:status": "madeAndImplemented",
        "tro:implementedLocation": {
          "loc:supplementaryPositionalDescription": {
            "loc:roadInformation": [
              {
                "loc:roadDestination": "BILBAO",
                "loc:roadName": "A",
                "loc:roadNumber": "8"
              }
            ]
          }
        }
      }
    ]
  },
  "tro:trafficRegulation": [
    {

```


Section 5.2 Message handling

Antxo/Donibane/Errenteria that has an extension of about 200m (from PK 0 to PK 0.182) (see Figure 5-6). The event was created the 2022/07/07 12:23:45 a.m. and has a validity duration of 24 h. For this type of event, the location reference type selected is linear since it is clearly defined in a road section and could affect only a specific carriageway (see Message 5-7).



Figure 5-6. Road marking damage detected in GI-636 exit branch road.

Message 5-7. DATEX II message example for road in poor conditions event.

```
{
  "d2:payload": {
    "@lang": "eng",
    "@modelBaseVersion": "3",
    "publicationTime": "2022-07-07T12:23:45",
    "publicationCreator": {
      "country": "es",
      "nationalIdentifier": "ceit-brta"
    }
  },
  "sit:situation": [
    {
      "@id": "2841",
      "sit:headerInformation": {
        "informationStatus": "technicalExercise"
      }
    },
    "sit:situationRecord": [
      {
        "@id": "2841-00",
        "@version": "1",
        "sit:situationRecordCreationTime": "2022-07-07T12:23:01",

```

```

"sit:situationRecordVersionTime": "2022-07-07T12:23:01",
"sit:probabilityOfOccurrence": "riskOf",
"sit:validity": {
  "validityStatus": "active",
  "validityTimeSpecification": {
    "overallStartTime": "2022-07-07T12:23:01",
    "overallEndTime": "2022-07-08T12:23:01"
  }
},
"sit:generalPublicComment": [
  {
    "sit:comment": {
      "values": {
        "value": [
          "Road marking damage"
        ]
      }
    }
  }
],
"sit:locationReference": {
  "loc:secondarySupplementaryDescription": {
    "loc:roadInformation": [
      {
        "loc:roadName": "GI",
        "loc:roadNumber": "636-1-0"
      }
    ]
  },
  "loc:tpegLinearLocation": {
    "loc:tpegDirection": "unknown",
    "loc:tpegLinearLocationType": "segment",
    "loc:to": {
      "loc:pointCoordinates": {
        "loc:latitude": 43.3113699,
        "loc:longitude": -1.9169185
      }
    },
    "loc:name": [
      {
        "loc:descriptor": {
          "values": {
            "value": [
              "PK 0.182"
            ]
          }
        }
      },
      {
        "loc:tpegOtherPointDescriptorType": "pointName"
      }
    ]
  },
  "loc:from": {
    "loc:pointCoordinates": {
      "loc:latitude": 43.3113314,
      "loc:longitude": -1.9153143
    },
    "loc:name": [
      {
        "loc:descriptor": {
          "values": {
            "value": [
              "PK 0"
            ]
          }
        }
      }
    ]
  }
}

```


5.3. Discussion


The interconnection of all the elements of the intelligent transport environment is essential to achieve a future of connected, cooperative and autonomous mobility. Incorporating road dynamic information will facilitate anticipation of drives in vehicles with low driving automation levels as well as CAVs systems with high driving automation levels. In this way both will be able to adapt their driving task in the event of an unexpected situation – low visibility due to weather conditions, infrastructure in poor conditions, etc. -. As a result, V2X communications will have an impact on efficiency, sustainability and road safety.

For these communications to be effective, they must be secure and reliable, and the messages transmitted must be interpretable by any user. Therefore, regulations in this area are of utmost importance as well as standardisation.

In this chapter, two of the most widely used standards for communication with OEMs (C-ITS) as well as with road operators (DATEX II) have been studied. Example messages have been implemented in both standards that capture the events generated by the modules developed in the previous chapters: fog banks, vertical signalling and road marking. In this study it has been found that the fog event is defined in both standards, similarly, there is also the possibility to report on vertical signage although it would have to be checked if DATEX II includes all the signs listed in the catalogues. However, there is currently no defined C-ITS format for a vehicle capable of detecting the status of signals or road markings to transmit them. Similarly, the DATEX II standard does not allow the case defined for road surfaces in poor conditions to be detailed.

It is clear that there is still a long way to go in the regularisation and standardisation of CCAM and it is necessary to continue working and updating the existing ones in order to adapt to new needs that may arise.

Chapter 6
Conclusions and
Future Research
Directions



6.1. Conclusions

The mobility of the future will be connected, cooperative and autonomous. The defined roadmaps make it clear that, although 100% autonomous mobility is still far from being a reality, vehicles will increasingly feature a higher level of automation. For this to be sustainable, interaction between all road users, V2X, is essential. In addition, it is also key that the infrastructure is in good condition so that the perception of the environment is as clear and reliable as possible. In this way, the exchange of safety-critical information together with a robust and reliable perception system will be able to assist the behaviour of both drivers and autonomous vehicles to improve driving efficiency and safety.

The general objective of the thesis has been met, having designed a road monitoring system capable of detecting critical situations such as fog banks and infrastructure damage by employing computer vision techniques.

- It has been implemented a system for the detection of fog banks between 0 and 400 meters of visibility, differentiated into three different alert levels.
- A traffic signal recognition system capable of recognising up to 164 different signal classes has been implemented.
- A system capable of detecting road line damage has been implemented.

This road monitoring system comprises three modules that already have a TRL level of between 5 and 6, this is, a model prototype already exists and for some of the modules the field verification process has started.

In general terms, the following can be concluded with regard to the sub-objectives:

Fog bank detection module

- Two different algorithms have been developed. The first one is a rule-based algorithm with an accuracy of 80% in real scenarios and 63% for synthetic ones. Whereas the second algorithm is based on DL which has an accuracy of 96% on synthetic images but 70% on real images.

- For the comparison of the module, it has not been possible to contrast it with the state of the art since results for public datasets of real scenarios have not yet been reported.
- It is observed that current algorithms for synthetic fog generation are not realistic enough as they present a much whiter colour than what the image acquisition systems capture.
- It is concluded that by generating a synthetic dataset where the fog has a more realistic colour, a DL model could be trained to improve the performance of the rule-based method for real scenes.

Traffic sign recognition module

- A two-stage algorithm has been developed for separate detection and classification. The first stage is based on classical CV while the second stage is based on a DL classifier voting system. The classification stage obtains accuracy values of around 92.4-98.5% while the detection stage presents good precision values but a low recall (52-76%).
- Regarding the validation of the module, similar values to the state of the art have been achieved for signal classification, however, the signal detection performance could be improved.
- It is concluded that although the developed system is suitable as an inventory tool, the development of a single-stage end-to-end recognition system based on DL could improve the performance of the current system.

Road marking damage module

- Two algorithms have been developed, firstly the problem was tackled with classical vision methods but due to the impossibility of modelling the defects, an end-to-end model based on DL was trained which achieves an F1 score of 92%.
- The validation of this module demonstrates that the current state of the art has been exceeded by 25% for the detection of the line marking defect.

Operational objectives

- The designed system is modular since no algorithm depends on any other algorithm than the acquisition system.
- The system used in this thesis is compact and low-cost as it consists of an RGB camera and a processing unit that could be installed in any vehicle.
- The system designed in the thesis can be implemented in real-time. All developed algorithms can run with calculation times between 70 - 300 ms.
- The system studied in the thesis can generate and transmit events for V2I/I2V communication in common and standardised languages such as DATEX II and C-ITS messages.
- The system designed in this thesis is universal and interoperable, it has been proved that the algorithms can work with images collected in different countries.
- The road operators and public administrations we have contacted show interest in the usefulness of the system designed in this thesis for the prioritisation and management of their investments and resources.

The work carried out in this thesis will make it possible to gain a better understanding of the state of the road and to detect critical situations which are limiting to the operation of ADSs or drivers and which endanger road safety. This information will be useful to improve the management of infrastructure maintenance and to warn vehicles so that they can anticipate a dangerous situation. This will also involve the improvement and extension of ISADs and ODDs that allow further progress to be made in completing the safety cases of self-driving vehicles.

6.2. Future research directions

The results and conclusions drawn from the present work allow to think of future developments related to road monitoring:

Improvement guidelines:

- Generate a synthetic dataset where fog has a more realistic colour (not pure white). It could be improved by inserting the synthetic fog directly into the real-scenario images already collected
- Develop a single-stage end-to-end traffic sign recognition system with a wider dataset.
- Improvement of the embedded system to improve image quality and prevent dirtiness of the frontal glass.

General future research lines

- The quality improvement and extension of the current datasets with new images that have a rich variety of different scenarios to improve the performance of the algorithms. This will require making use of other public image repositories such as Google Street View and enriching images of real scenarios with synthetic fog or defects to compensate for the imbalance of such rare situations (the use of Generative Adversarial Networks (GANs) could be explored).
- Extension of the road damage module to include other types of defects that are critical to road safety such as potholes and others that can help to better manage road maintenance by prioritizing resources.
- Extension of the adverse weather detection module to include detection of spray or splash clouds effect produced by extremely wet roads that highly hinder the visibility of the drivers and the capabilities of autonomous vehicles.
- For the construction of the embedded system, migration of the execution of the algorithms to the cloud to allow for more powerful processing than on an on-board computer.

Chapter 6 Conclusions and Future Research Lines

- The universalisation and standardisation of data, so that the scientific community can work synergistically with the ultimate goal of improving road safety and sharing local developments with others to achieve richer models for a global solution.
- The dumping of the data collected by the road monitoring system into a GIS data visualisation tool. This will allow post-processing of the raw data to generate the asset inventory and improve infrastructure maintenance management.
- The physical modelling of the road (macroscale, microscale, capacities, etc.) so that in combination with the data extracted from the road monitoring system a Digital Twin (DT) of the road can be built. This will allow the study of the life cycle and predict the deterioration of the road.

Bibliography

1. ERTRAC Working Group, C. *Connected , Cooperative and Automated Mobility Roadmap*; 2022;
2. Connected Cooperative and Automated Partnership, C. *Strategic Research and Innovation Agenda (SRIA) 2021-2027 - European Leadership in Safe and Sustainable Road Transport through Automation*; 2021;
3. European Commission *EU Road Safety Policy Framework 2021-2030 - Next Steps towards ' Vision Zero '*; 2020;
4. European Commission *Sustainable and Smart Mobility Strategy - Putting European Transport on Track for the Future*; 2020;
5. SAE On-Road Automated Driving Committee, O.; TC204/WG14, I. SAE J3016 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. **2021**.
6. PIARC; Universitat Politècnica de València; GIIC; iTEAM PIARC *Special Project "Smart Roads Classification" Proposal*; 2021;
7. Wishart, J. *Fundamentals of Connected and Automated Vehicles*; SAE International, 2022;
8. Neuvition, I. LiDAR Price for Cars Available online: <https://www.neuvition.com/media/blog/lidar-price.html>.
9. Wang, B.; Zhu, M.; Lu, Y.; Wang, J.; Gao, W.; Wei, H. Real-Time 3D Object Detection from Point Cloud through Foreground Segmentation. *IEEE Access* **2021**, *9*, 84886–84898, doi:10.1109/ACCESS.2021.3087179.
10. Gupta, A.; Anpalagan, A.; Guan, L.; Khwaja, A.S. Deep Learning for Object Detection and Scene Perception in Self-Driving Cars: Survey, Challenges, and Open Issues. *Array* **2021**, *10*, 100057, doi:10.1016/j.array.2021.100057.

Bibliography

11. Hahner, M.; Dai, D.; Sakaridis, C.; Zaech, J.N.; Gool, L. Van Semantic Understanding of Foggy Scenes with Purely Synthetic Data. *2019 IEEE Intell. Transp. Syst. Conf. ITSC 2019* **2019**, 3675–3681, doi:10.1109/ITSC.2019.8917518.
12. Zang, S.; Ding, M.; Smith, D.; Tyler, P.; Rokotoarivelo, T.; Ali Kaafar, M. *IEEE Vehicular Technology Magazine*. IEEE 2019, pp. 103–111.
13. Liu, Z.; He, Y.; Wang, C.; Song, R. Analysis of the Influence of Foggy Weather Environment on the Detection Effect of Machine Vision Obstacles. *Sensors (Switzerland)* **2020**, *20*, doi:10.3390/s20020349.
14. Koschmieder, H. Theorie Der Horizontalen Sichtweite. *Beitrage zur Phys. der freien Atmosphere* **1924**, 33–53.
15. W. E. K. Middleton Vision Through the Atmosphere. In *Geophysics II*; 1957; Vol. 10, pp. 262–264 ISBN 978-3-642-45883-5.
16. Hautiere, N.; Bigorgne, E.; Bossu, J.; Aubert, D.; Hautiere, N. Meteorological Conditions Processing for Vision-Based Traffic Monitoring. In Proceedings of the The Eighth International Workshop on Visual Surveillance - VS2008; 2008.
17. Hautière, N.; Tarel, J.P.; Aubert, D. Towards Fog-Free in-Vehicle Vision Systems through Contrast Restoration. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2007**, doi:10.1109/CVPR.2007.383259.
18. Hautière, N.; Tarel, J.P.; Halmaoui, H.; Brémond, R.; Aubert, D. Enhanced Fog Detection and Free-Space Segmentation for Car Navigation. *Mach. Vis. Appl.* **2014**, *25*, 667–679, doi:10.1007/s00138-011-0383-3.
19. Hautière, N.; Tarel, J.; Lavenant, J. Automatic Fog Detection and Estimation of Visibility Distance through Use of an Onboard Camera. *Mach. Vis. Appl.* **2006**, *17*, 8–20, doi:10.1007/s00138-005-0011-1.
20. Tarel, J.P.; Hautière, N.; Cord, A.; Gruyer, D.; Halmaoui, H. Improved Visibility of Road Scene Images under Heterogeneous Fog. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV'10); 2010; pp. 478–485.

-
21. Tarel, J.; Hauti, N.; Caraffa, L.; Halmaoui, H.; Gruyer, D.; Tarel, J.; Hauti, N.; Caraffa, L.; Halmaoui, H.; Tarel, J.; et al. Heterogeneous Fog To Cite This Version : Vision Enhancement in Homogeneous and Heterogeneous Fog. *IEEE Intell. Transp. Syst. Mag.* **2012**, *4*, 6–20.
 22. Negru, M.; Nedevschi, S. Image Based Fog Detection and Visibility Estimation for Driving Assistance Systems. *Proc. - 2013 IEEE 9th Int. Conf. Intell. Comput. Commun. Process. ICCP 2013* **2013**, 163–168, doi:10.1109/ICCP.2013.6646102.
 23. Negru, M.; Nedevschi, S. Assisting Navigation in Homogenous Fog. *VISAPP 2014 - Proc. 9th Int. Conf. Comput. Vis. Theory Appl.* **2014**, *2*, 619–626, doi:10.5220/0004740006190626.
 24. Negru, M.; Nedevschi, S.; Peter, R.I. Exponential Contrast Restoration in Fog Conditions for Driving Assistance. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2257–2268, doi:10.1109/TITS.2015.2405013.
 25. Tan, R.T. Visibility in Bad Weather from a Single Image. *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR* **2008**, 1–8, doi:10.1109/CVPR.2008.4587643.
 26. Fattal, R. Single Image Dehazing. *ACM Trans. Graph.* **2008**, *27*, doi:10.1145/1360612.1360671.
 27. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353, doi:10.1109/TPAMI.2010.168.
 28. Yeh, C.-H.; Kang, L.-W.; Lee, M.-S.; Lin, C.-Y. Haze Effect Removal from Image via Haze Density Estimation in Optical Model. *Opt. Express* **2013**, *21*, 27127, doi:10.1364/oe.21.027127.
 29. Yeh, C.H.; Kang, L.W.; Lin, C.Y.; Lin, C.Y. Efficient Image/Video Dehazing through Haze Density Analysis Based on Pixel-Based Dark Channel Prior. *Proc. - 3rd Int. Conf. Inf. Secur. Intell. Control. ISIC 2012* **2012**, 238–241, doi:10.1109/ISIC.2012.6449750.
 30. Huang, S.C.; Chen, B.H.; Wang, W.J. Visibility Restoration of Single Hazy Images Captured in Real-World Weather Conditions. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 1814–1824, doi:10.1109/TCSVT.2014.2317854.
-

Bibliography

31. Pavlić, M.; Belzner, H.; Rigoll, G.; Ilić, S. Image Based Fog Detection in Vehicles. *IEEE Intell. Veh. Symp. Proc.* **2012**, 1132–1137, doi:10.1109/IVS.2012.6232256.
32. Pavlic, M.; Rigoll, G.; Ilic, S. Classification of Images in Fog and Fog-Free Scenes for Use in Vehicles. *IEEE Intell. Veh. Symp. Proc.* **2013**, 481–486, doi:10.1109/IVS.2013.6629514.
33. Spinneker, R.; Koch, C.; Park, S.B.; Yoon, J.J. Fast Fog Detection for Camera Based Advanced Driver Assistance Systems. *2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014* **2014**, 1369–1374, doi:10.1109/ITSC.2014.6957878.
34. Asery, R.; Sunkaria, R.K.; Sharma, L.D.; Kumar, A. Fog Detection Using GLCM Based Features and SVM. *Conf. Adv. Signal Process. CASP 2016* **2016**, 72–76, doi:10.1109/CASP.2016.7746140.
35. Alami, S.; Ezzine, A.; Elhassouni, F. Local Fog Detection Based on Saturation and RGB-Correlation. *Proc. - Comput. Graph. Imaging Vis. New Tech. Trends, CGiV 2016* **2016**, 1–5, doi:10.1109/CGiV.2016.10.
36. Li, C.; Lu, X.; Tong, C.; Zeng, W. A Fog Level Detection Method Based on Grayscale Features. *Proc. - 2014 7th Int. Symp. Comput. Intell. Des. Isc. 2014* **2015**, 1, 417–420, doi:10.1109/ISCID.2014.198.
37. Liu, C.; Lu, X.; Ji, S.; Geng, W. A Fog Level Detection Method Based on Image HSV Color Histogram. *PIC 2014 - Proc. 2014 IEEE Int. Conf. Prog. Informatics Comput.* **2014**, 373–377, doi:10.1109/PIC.2014.6972360.
38. Zhang, D.; Sullivan, T.; O'Connor, N.E.; Gillespie, R.; Regan, F. Coastal Fog Detection Using Visual Sensing. *MTS/IEEE Ocean. 2015 - Genova Discov. Sustain. Ocean Energy a New World* **2015**, 2–6, doi:10.1109/OCEANS-Genova.2015.7271683.
39. Chaabani, H.; Kamoun, F.; Bargaoui, H.; Outay, F.; Yasar, A.U.H. A Neural Network Approach to Visibility Range Estimation under Foggy Weather Conditions. *Procedia Comput. Sci.* **2017**, 113, 466–471, doi:10.1016/j.procs.2017.08.304.
40. Belaroussi, R.; Gruyer, D. Impact of Reduced Visibility from Fog on Traffic Sign Detection. *IEEE Intell. Veh. Symp. Proc.* **2014**, 1302–1306, doi:10.1109/IVS.2014.6856535.

-
41. Vaibhav, V.; Konda, K.R.; Kondapalli, C.; Praveen, K.; Kondoju, B. Real-Time Fog Visibility Range Estimation for Autonomous Driving Applications. *2020 IEEE 23rd Int. Conf. Intell. Transp. Syst. ITSC 2020*, 2–7, doi:10.1109/ITSC45102.2020.9294740.
 42. Geiger, A.; Lenz, P.; Urtasun, R. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2012**, 3354–3361, doi:10.1109/CVPR.2012.6248074.
 43. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016**, 2016-Decem, 3213–3223, doi:10.1109/CVPR.2016.350.
 44. Neuhold, G.; Ollmann, T.; Bulo, S.R.; Kotschieder, P. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. *Proc. IEEE Int. Conf. Comput. Vis.* **2017**, 2017-Octob, 5000–5009, doi:10.1109/ICCV.2017.534.
 45. Huang, X.; Wang, P.; Cheng, X.; Zhou, D.; Geng, Q.; Yang, R. The ApolloScape Open Dataset for Autonomous Driving and Its Application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, 42, 2702–2719, doi:10.1109/TPAMI.2019.2926463.
 46. Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; Darrell, T. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2020**, 2633–2642, doi:10.1109/cvpr42600.2020.00271.
 47. Sakaridis, C.; Dai, D.; Van Gool, L. Semantic Foggy Scene Understanding with Synthetic Data. *Int. J. Comput. Vis.* **2018**, 126, 973–992, doi:10.1007/s11263-018-1072-8.
 48. Dai, D.; Sakaridis, C.; Hecker, S.; Van Gool, L. Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding. *Int. J. Comput. Vis.* **2020**, 128, 1182–1204, doi:10.1007/s11263-019-01182-4.
-

Bibliography

49. Sakaridis, C.; Dai, D.; Van Gool, L. ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding. **2021**, 10765–10775, doi:10.1109/iccv48922.2021.01059.
50. Bijelic, M.; Gruber, T.; Mannan, F.; Kraus, F.; Ritter, W.; Dietmayer, K.; Heide, F. Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. **2020**, 11679–11689, doi:10.1109/cvpr42600.2020.01170.
51. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. **2012**, *34*, 2274–2281.
52. AFNOR; Eruopéennes, N.F. et NF P99-320. Météorologie Routière - Recueil Des Données Météorologiques et Routiers 1998.
53. Colomb, M.; Duthon, P.; Laukkanen, S. *D2.1. Characteristics of Adverse Weather Conditions*; 2017;
54. Tan, M.; Le, Q. V. EfficientNetV2: Smaller Models and Faster Training. *Int. Conf. Mach. Learn.* **2021**.
55. Coronado, G.A.P.; Muñoz, M.R.; Armingol, J.M.; De La Escalera, A.; Muoz, J.J.; Van Bijsterveld, W.; Bolaño, J.A. Detection and Classification of Road Signs for Automatic Inventory Systems Using Computer Vision. *Integr. Comput. Aided. Eng.* **2012**, *19*, 285–298, doi:10.3233/ICA-2012-0404.
56. Gomez-Moreno, H.; Maldonado-Bascon, S.; Gil-Jimenez, P.; Lafuente-Arroyo, S. Goal Evaluation of Segmentation Algorithms for Traffic Sign Recognition. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 917–930, doi:10.1109/TITS.2010.2054084.
57. Miura, J.; Kanda, T.; Nakatani, S.; Shirai, Y. An Active Vision System for On-Line Traffic Sign Recognition. *IEICE Trans. Inf. Syst.* **2002**, *E85-D*, 1784–1792, doi:10.1109/itsc.2000.881017.
58. Angel, M.; Angel, M.; Mart, E. ITSCSenales2006.Pdf. **2006**, 811–816.
59. Yang, H.; Liu, C.; Liu, K.; Huang, S. Traffic Sign Recognition in Disturbing Environments Hsiu-Ming. *Lect. Notes Comput. Sci.* **2003**, *2871*, 252.261.

-
60. Barnes, N.; Zelinsky, A. Real-Time Radial Symmetry for Speed Sign Detection. *IEEE Intell. Veh. Symp. Proc.* **2004**, 566–571, doi:10.1109/ivs.2004.1336446.
 61. Loy, G.; Barnes, N. Fast Shape-Based Road Sign Detection for a Driver Assistance System. *2004 IEEE/RSJ Int. Conf. Intell. Robot. Syst.* **2004**, *1*, 70–75, doi:10.1109/iros.2004.1389331.
 62. Gavrila, D.M. Traffic Sign Recognition Revisited. **1999**, 86–93, doi:10.1007/978-3-642-60243-6_10.
 63. Houben, S. A Single Target Voting Scheme for Traffic Sign Detection. *IEEE Intell. Veh. Symp. Proc.* **2011**, 124–129, doi:10.1109/IVS.2011.5940429.
 64. Prisacariu, V.A.; Timofte, R.; Zimmermann, K.; Reid, I.; Van Gool, L. Integrating Object Detection with 3D Tracking towards a Better Driver Assistance System. *Proc. - Int. Conf. Pattern Recognit.* **2010**, 3344–3347, doi:10.1109/ICPR.2010.816.
 65. Lim, K.; Hong, Y.; Choi, Y.; Byun, H. Real-Time Traffic Sign Recognition Based on a General Purpose GPU and Deep-Learning. *PLoS One* **2017**, *12*.
 66. Viola, P.; Jones, M. Rapid Object Detection Using a Boosted Cascade of Simple Features. *Conf. Comput. Vis. Pattern Recognit.* **2001**, doi:10.1109/ICCD.2001.8389630.
 67. Freund, Y.; Shapire, R.E. A Short Introduction to Boosting. *J. Japanese Soc. Artif. Intell.* **1999**, *14*, 771–780.
 68. Chen, L.; Li, Q.; Li, M.; Zhang, L.; Mao, Q. Design of a Multi-Sensor Cooperation Travel Environment Perception System for Autonomous Vehicle _ Enhanced Reader.Pdf. *Sensors* **2012**, *12*, doi:10.3390/s120912386.
 69. Maldonado-Bascón, S.; Laguente-Arroyo, S.; Siegmann, P.; Gómez-Moreno, H.; Acevedo-Rodríguez, F.J. Traffic Sign Recognition System for Inventory Purposes. *IEEE Intell. Veh. Symp.* **2008**, 590–595.
 70. Kouzani, A.Z. Road-Sign Identification Using Ensemble Learning. *IEEE Intell. Veh. Symp.* **2007**, 438–443.
-

Bibliography

71. Escalera, A. De; Armingol, J.M.; Salichs, M.A. Traffic Sign Detection for Driver Support Systems. In Proceedings of the International Conference on field and service robotics; 2001.
72. Hechri, A.; Mtibaa, A. Automatic Detection and Recognition of Road Sign for Driver Assistance System. In Proceedings of the 16th IEEE Mediterranean Electrotechnical Conference; IEEE, 2012; pp. 888–891.
73. Cireşan, D.; Meier, U.; Schmidhuber, J. A Committee of Neural Networks for Traffic Sign Classification. *Proc. Int. Jt. Conf. Neural Networks* **2011**, *1*, 1918–1921, doi:10.1109/IJCNN.2011.6033458.
74. Mogelmoose, A.; Liu, D.; Trivedi, M.M. Detection of U.S. Traffic Signs. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 3116–3125, doi:10.1109/TITS.2015.2433019.
75. Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-Sign Detection and Classification in the Wild Zhe. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE Xplore, 2016; pp. 2110–2118.
76. Zuo, Z.; Yu, K.; Zhou, Q.; Wang, X.; Li, T. Traffic Signs Detection Based on Faster R-CNN. *Proc. - IEEE 37th Int. Conf. Distrib. Comput. Syst. Work. ICDCSW 2017* **2017**, 286–288, doi:10.1109/ICDCSW.2017.34.
77. Shao, F.; Wang, X.; Meng, F.; Rui, T.; Wang, D.; Tang, J. Real-Time Traffic Sign Detection and Recognition Method Based on Simplified Gabor Wavelets and CNNs. *Sensors (Switzerland)* **2018**, *18*, doi:10.3390/s18103192.
78. Shao, F.; Wang, X.; Meng, F.; Zhu, J.; Wang, D.; Dai, J. Improved Faster R-CNN Traffic Sign Detection Based on a Second Region of Interest and Highly Possible Regions Proposal Network. *Sensors (Switzerland)* **2019**, *19*, doi:10.3390/s19102288.
79. Zhang, J.; Hui, L.; Lu, J.; Zhu, Y. Attention-Based Neural Network for Traffic Sign Detection. *Proc. - Int. Conf. Pattern Recognit.* **2018**, *2018-Augus*, 1839–1844, doi:10.1109/ICPR.2018.8546289.
80. Yang, T.; Long, X.; Kumar, A.; Zheng, Z.; Tong, C. Deep Detection Network for Real-Life Traffic Sign in Vehicular Networks. *Comput. Networks* **2018**, *136*, 95–104, doi:10.1016/j.comnet.2018.02.026.

-
81. Yuan, Y.; Xiong, Z.; Wang, Q. VSSA-NET: Vertical Spatial Sequence Attention Network for Traffic Sign Detection. *IEEE Trans. Image Process.* **2019**, *28*, 3423–3434, doi:10.1109/TIP.2019.2896952.
 82. Zhang, J.; Huan, M.; Jin, X.; Li, X. A Real-Time Chinese Traffic Sign Detection Algorithm Based on Modified YOLOv2. *Algorithms* **2017**, *1*–13, doi:10.3390/a10040127.
 83. Wang, L.; Zhou, K.; Chu, A.; Wang, G.; Wang, L. An Improved Light-Weight Traffic Sign Recognition Algorithm Based on YOLOv4-Tiny. *IEEE Access* **2021**, *9*, 124963–124971, doi:10.1109/ACCESS.2021.3109798.
 84. Shan, H.; Zhu, W. A Small Traffic Sign Detection Algorithm Based on Modified SSD. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *646*, doi:10.1088/1757-899X/646/1/012006.
 85. Jin, Y.; Fu, Y.; Wang, W.; Guo, J.; Ren, C.; Xiang, X. Multi-Feature Fusion and Enhancement Single Shot Detector for Traffic Sign Recognition. *IEEE Access* **2020**, *8*, 38931–38940, doi:10.1109/ACCESS.2020.2975828.
 86. Stallkamp, J.; Schlipsing, M.; Salmen, J.; Igel, C. Man vs.~Computer: {B}enchmarking Machine Learning Algorithms for Traffic Sign Recognition. *Neural Networks* **2012**, *32*, 323–332.
 87. Babić, D.; Babić, D.; Fiolić, M.; Šarić, Ž. Analysis of Market-Ready Traffic Sign Recognition Systems in Cars: A Test Field Study. *Energies* **2021**, *14*, doi:10.3390/en14123697.
 88. Stallkamp, J.; Schlipsing, M.; Salmen, J.; Igel, C. Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition. *Neural Networks* **2012**, *32*, 323–332, doi:10.1016/j.neunet.2012.02.016.
 89. Houben, S.; Stallkamp, J.; Salmen, J.; Schlipsing, M.; Igel, C. Detection of Traffic Signs in Real-World Images: The German Traffic Sign Detection Benchmark. *Proc. Int. Jt. Conf. Neural Networks* **2013**, doi:10.1109/IJCNN.2013.6706807.
-

Bibliography

90. Timofte, R.; Zimmermann, K.; Van Gool, L. Multi-View Traffic Sign Detection, Recognition, and 3D Localisation. *Mach. Vis. Appl.* **2011**, *25*, 633–647, doi:10.1007/s00138-011-0391-3.
91. Larsson, F.; Felsberg, M. Using Fourier Descriptors and Spatial Models for Traffic Sign Recognition. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2011**, 6688 LNCS, 238–249, doi:10.1007/978-3-642-21227-7_23.
92. Grigorescu, C.; Petkov, N. Distance Sets for Shape Filters and Shape Recognition. *IEEE Trans. Image Process.* **2003**, *12*, 1274–1286, doi:10.1109/TIP.2003.816010.
93. Belaroussi, R.; Foucher, P.; Tarel, J.P.; Soheilian, B.; Charbonnier, P.; Paparoditis, N. Road Sign Detection in Images: A Case Study. *Proc. - Int. Conf. Pattern Recognit.* **2010**, 484–488, doi:10.1109/ICPR.2010.1125.
94. Šegvić, S.; Brkić, K.; Kalafatić, Z.; Stanisavljević, V.; Ševrović, M.; Budimir, D.; Dadić, I. A Computer Vision Assisted Geoinformation Inventory for Traffic Infrastructure. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 2010*, 66–73, doi:10.1109/ITSC.2010.5624979.
95. Serna, C.G.; Ruichek, Y. Classification of Traffic Signs: The European Dataset. *IEEE Access* **2018**, *6*, 78136–78148, doi:10.1109/ACCESS.2018.2884826.
96. Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-Sign Detection and Classification in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; pp. 2110–2118.
97. Yang, Y.; Luo, H.; Xu, H.; Wu, F. Towards Real-Time Traffic Sign Detection and Classification. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2022–2031, doi:10.1109/TITS.2015.2482461.
98. Zhang, Y.; Wang, Z.; Qi, Y.; Liu, J.; Yang, J. CTSD: A Dataset for Traffic Sign Recognition in Complex Real-World Images. *VCIP 2018 - IEEE Int. Conf. Vis. Commun. Image Process.* **2018**, 2018–2021, doi:10.1109/VCIP.2018.8698666.

-
99. Youssef, A.; Albani, D.; Nardi, D.; Bloisi, D.D. Fast Traffic Sign Recognition Using Color Segmentation and Deep Convolutional Networks. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2016**, *10016 LNCS*, 205–216, doi:10.1007/978-3-319-48680-2_19.
 100. Møgelmoose, A.; Trivedi, M.M.; Moeslund, T.B. Vision Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1484–1497.
 101. Rehman, Y.; Riaz, I.; Fan, X.; Shin, H. D-Patches: Effective Traffic Sign Detection with Occlusion Handling. *IET Comput. Vis.* **2017**, *11*, 368–377, doi:10.1049/iet-cvi.2016.0303.
 102. Tabernik, D.; Skocaj, D. Deep Learning for Large-Scale Traffic-Sign Detection and Recognition. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1427–1440, doi:10.1109/TITS.2019.2913588.
 103. Temel, D.; Chen, M.; Alregib, G. Conditions : A Deeper Look into Performance Variations and Spectral Characteristics. **2020**, *21*, 3663–3673.
 104. Ertler, C.; Mislej, J.; Ollmann, T.; Porzi, L.; Neuhold, G.; Kuang, Y. The Mapillary Traffic Sign Dataset for Detection and Classification on a Global Scale. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2020**, *12368 LNCS*, 68–84, doi:10.1007/978-3-030-58592-1_5.
 105. Siniosoglou, I.; Sarigiannidis, P.; Spyridis, Y.; Khadka, A.; Efstathopoulos, G.; Lagkas, T. Synthetic Traffic Signs Dataset for Traffic Sign Detection & Recognition in Distributed Smart Systems. *Proc. - 17th Annu. Int. Conf. Distrib. Comput. Sens. Syst. DCOS 2021* **2021**, 302–308, doi:10.1109/DCOSS52077.2021.00056.
 106. Intel CVAT Computer Vision Annotation Tool 2019.
 107. Wang, Y.-Q. An Analysis of the Viola-Jones Face Detection Algorithm. *Image Process. Line* **2014**, *4*, 128–148, doi:10.5201/ipol.2014.104.
-

Bibliography

108. Dollar, P.; Appel, R.; Belongie, S.; Perona, P. Fast Feature Pyramids for Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1532–1545.
109. Jahnen, M.; Budweiser, P. SignKafe Model 2015.
110. Ciregan, D.; Meier, U.; Schmidhuber, J. Multi-Column Deep Neural Networks for Image Classification. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2012**, 3642–3649, doi:10.1109/CVPR.2012.6248110.
111. Jin, J.; Fu, K.; Zhang, C. Traffic Sign Recognition with Hinge Loss Trained Convolutional Neural Networks. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1991–2000, doi:10.1109/TITS.2014.2308281.
112. Chilamkurthy, S. Keras Tutorial - Traffic Sign Recognition 2017.
113. Ranyal, E.; Sadhu, A.; Jain, K. Road Condition Monitoring Using Smart Sensing and Artificial Intelligence: A Review. *Sensors* **2022**, *22*, 1–27, doi:10.3390/s22083044.
114. Kyriakou, C.; Christodoulou, S.E.; ASCE, A.M.; Dimitriou, L. Smartphone-Based Pothole Detection Utilizing Artificial Neural Networks. *J. Infrastruct. Syst.* **2019**, *25*, doi:10.1061/(ASCE)IS.1943-555X.0000489.
115. Sandamal, R.M.K.; Pasindu, H.R. Applicability of Smartphone-Based Roughness Data for Rural Road Pavement Condition Evaluation. *Int. J. Pavement Eng.* **2022**, *23*, 663–672, doi:10.1080/10298436.2020.1765243.
116. Zhang, Z.; Zhang, H.; Xu, S.; Lv, W. Pavement Roughness Evaluation Method Based on the Theoretical Relationship between Acceleration Measured by Smartphone and IRI. *Int. J. Pavement Eng.* **2021**, doi:10.1080/10298436.2021.1881783.
117. Rana, S.; Asaduzzaman. Vibration Based Pavement Roughness Monitoring System Using Vehicle Dynamics and Smartphone with Estimated Vehicle Parameters. *Results Eng.* **2021**, *12*, 100294, doi:10.1016/j.rineng.2021.100294.
118. Zhang, L.; Xu, W.; Zhu, L.; Yuan, X.; Zhang, C. Study on Pavement Defect Detection Based on Image Processing Utilizing UAV. *J. Phys. Conf. Ser.* **2019**, *1168*.

-
119. Pan, Y.; Zhang, X.; Cervone, G.; Yang, L. Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3701–3712, doi:10.1109/JSTARS.2018.2865528.
 120. Praticò, F.G.; Fedele, R.; Naumov, V.; Sauer, T. Detection and Monitoring of Bottom-up Cracks in Road Pavement Using a Machine-Learning Approach. *Algorithms* **2020**, *13*, 1–16, doi:10.3390/a13040081.
 121. Lekshmipathy, J.; Samuel, N.M.; Velayudhan, S. Vibration vs. Vision: Best Approach for Automated Pavement Distress Detection. *Int. J. Pavement Res. Technol.* **2020**, *13*, 402–410, doi:10.1007/s42947-020-0302-y.
 122. Maeda, H.; Sekimoto, Y.; Seto, T.; Kashiyama, T.; Omata, H. Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. *Comput. Civ. Infrastruct. Eng.* **2018**, *33*, 1127–1141, doi:10.1111/mice.12387.
 123. Vokhidov, H.; Hong, H.G.; Kang, J.K.; Hoang, T.M.; Park, K.R. Recognition of Damaged Arrow-Road Markings by Visible Light Camera Sensor Based on Convolutional Neural Network. *Sensors (Switzerland)* **2016**, *16*, doi:10.3390/s16122160.
 124. Kawano, M.; Mikami, K.; Yokoyama, S.; Yonezawa, T.; Nakazawa, J. Road Marking Blur Detection with Drive Recorder. *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017* **2017**, 2018-Janua, 4092–4097, doi:10.1109/BigData.2017.8258427.
 125. Xu, S.; Wang, J.; Wu, P.; Shou, W.; Wang, X.; Chen, M. Vision-Based Pavement Marking Detection and Condition Assessment-a Case Study. *Appl. Sci.* **2021**, *11*, doi:10.3390/app11073152.
 126. Wei, C.; Li, S.; Wu, K.; Zhang, Z.; Wang, Y. Damage Inspection for Road Markings Based on Images with Hierarchical Semantic Segmentation Strategy and Dynamic Homography Estimation. *Autom. Constr.* **2021**, *131*, 103876, doi:10.1016/j.autcon.2021.103876.
 127. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D.; Mraz, A.; Kashiyama, T.; Sekimoto, Y. Deep Learning-Based Road Damage Detection and Classification for Multiple Countries. *Autom. Constr.* **2021**, *132*, 103935, doi:10.1016/j.autcon.2021.103935.
-

Bibliography

128. Arya, D.; Maeda, H.; Sekimoto, Y.; Ghosh, S.K.; Toshniwal, D.; Kashiyama, T.; Omata, H. Crowdsensing-Based Road Damage Detection Challenge (CRDDC2022) Available online: <https://crddc2022.sekilab.global/overview/>.
129. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D. RDD2022 : A Multi - National Image Dataset for Automatic Road Damage Detection. *Pre-print* **2022**, 1–16.
130. Gopalakrishnan, K. Deep Learning in Data-Driven Pavement Image Analysis and Automated Distress Detection: A Review. *Data* **2018**, *3*, doi:10.3390/data3030028.
131. Zhang, K.; Cheng, H.D.; Zhange, B. Unified Approach to Pavement Crack and Sealed Crack Detection Using Preclassification Based on Transfer Learning. *J. Comput. Civ. Eng.* **2018**, *32*.
132. Zhang, L.; Yang, F.; Zhang, Y.D.; Zhu, Y.J. Road Crak Detection Using Deep Convolutional Neural Network. *IEEE Int. Conf. Image Process.* **2016**, 3708–3712.
133. Ochoa-Ruiz, G.; Angulo-Murillo, A.A.; Ochoa-Zezzatti, A.; Aguilar-Lobo, L.M.; Vega-Fernández, J.A.; Natraj, S. An Asphalt Damage Dataset and Detection System Based on Retinanet for Road Conditions Assessment. *Appl. Sci.* **2020**, *10*, doi:10.3390/app10113974.
134. Eisenbach, M.; Stricker, R.; Seichter, D.; Amende, K.; Debes, K.; Sesselmann, M.; Ebersbach, D.; Stoeckert, U.; Gross, H.M. How to Get Pavement Distress Detection Ready for Deep Learning? A Systematic Approach. *Proc. Int. Jt. Conf. Neural Networks* **2017**, 2017-May, 2039–2047, doi:10.1109/IJCNN.2017.7966101.
135. Amhaz, R.; Chambon, S.; Idier, J.; Baltazart, V. Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2718–2729, doi:10.1109/TITS.2015.2477675.
136. Majidifard, H.; Jin, P.; Adu-Gyamfi, Y.; Buttlar, W.G. Pavement Image Datasets: A New Benchmark Dataset to Classify and Densify Pavement Distresses. *Transp. Res. Rec.* **2020**, 2674, 328–339, doi:10.1177/0361198120907283.

-
137. Zhu, J.; Zhong, J.; Ma, T.; Huang, X.; Zhang, W.; Zhou, Y. Pavement Distress Detection Using Convolutional Neural Networks with Images Captured via UAV. *Autom. Constr.* **2022**, *133*, 103991, doi:10.1016/j.autcon.2021.103991.
 138. Mei, Q.; Gül, M. A Cost Effective Solution for Pavement Crack Inspection Using Cameras and Deep Neural Networks. *Constr. Build. Mater.* **2020**, *256*, doi:10.1016/j.conbuildmat.2020.119397.
 139. Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic Road Crack Detection Using Random Structured Forests. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1–12.
 140. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D.; Sekimoto, Y. RDD2020: An Annotated Image Dataset for Automatic Road Damage Detection Using Deep Learning. *Data Br.* **2021**, *36*, 107133, doi:10.1016/j.dib.2021.107133.
 141. Maeda, H.; Kashiya, T.; Sekimoto, Y.; Seto, T.; Omata, H. Generative Adversarial Network for Road Damage Detection. *Comput. Civ. Infrastruct. Eng.* **2021**, *36*, 47–60, doi:10.1111/mice.12561.
 142. Tzutalin LabelImg 2015.
 143. Ko, Y.; Lee, Y.; Azam, S.; Munir, F.; Jeon, M.; Pedrycz, W. Key Points Estimation and Point Instance Segmentation Approach for Lane Detection. *IEEE Trans. Intell. Transp. Syst.* **2021**, 1–10, doi:10.1109/TITS.2021.3088488.
 144. Hegde, V.; Trivedi, D.; Alfarrarjeh, A.; Deepak, A.; Ho Kim, S.; Shahabi, C. Yet Another Deep Learning Approach for Road Damage Detection Using Ensemble Learning. *Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020* **2020**, 5553–5558, doi:10.1109/BigData50022.2020.9377833.
 145. Ochoa-Ruiz, G.; Angulo-Murillo, A.A.; Ochoa-Zezzatti, A.; Aguilar-Lobo, L.M.; Vega-Fernández, J.A.; Natraj, S. An Asphalt Damage Dataset and Detection System Based on Retinanet for Road Conditions Assessment. *Appl. Sci.* **2020**, *10*, doi:10.3390/app10113974.

Bibliography

146. Jung, A.B.; Wada, K.; Crall, J.; Tanaka, S.; Graving, J.; Reinders, C.; Yadav, S.; Banerjee, J.; Vecsei, G.; Kraft, A.; et al. *Imgaug Python Library* 2020.
147. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; 2016;
148. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; et al. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* **2017**, 2017-Janua, 3296–3305, doi:10.1109/CVPR.2017.351.
149. Shaoqing Ren;Kaiming He; Ross Girshick; Jian Sun Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *A World Made Money* **2016**, 241–294, doi:10.2307/j.ctt1d98bxx.10.
150. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot Multibox Detector. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2016**, 9905 LNCS, 21–37, doi:10.1007/978-3-319-46448-0_2.
151. Tan, M.; Pang, R.; Le, Q. V. EfficientDet: Scalable and Efficient Object Detection. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2020**, 10778–10787, doi:10.1109/CVPR42600.2020.01079.
152. Everingham, M.; Van-Gool, L.; Williams, C. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results Available online: <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
153. Padilla, R.; Netto, S.L.; Da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms. *Int. Conf. Syst. Signals, Image Process.* **2020**, 2020-July, 237–242, doi:10.1109/IWSSIP48289.2020.9145130.
154. ISO 14823:2017(E) INTERNATIONAL STANDARD Intelligent Transport Systems — Graphic Data Dictionary 2017.
155. D_AT_EX II Organization Datex II Docs: Mastering D_AT_EX II Available online: <https://docs.datex2.eu/mastering/index.html>.

156. Paschotta, R. RP Photonics Encyclopedia - Color Spaces Available online: [https://www.rp-photonics.com/encyclopedia_cite.html?article=color spaces](https://www.rp-photonics.com/encyclopedia_cite.html?article=color%20spaces) (accessed on 20 October 2020).
157. Kalra, N.; Paddock, S.M. Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability? *Transp. Res. Part A Policy Pract.* **2016**, *94*, 182–193, doi:10.1016/j.tra.2016.09.010.

Publications



(Q2) Iparraguirre, O.; Amundarain, A.; Brazalez, A.; Borro, D. "Sensors on the Move: Onboard Camera-Based Real-Time Traffic Alerts Paving the Way for Cooperative Roads". *Sensors, Special Issue on Sensors for Road Vehicles of the Future*, Vol.21, No. 4, 1254. February 2021. doi:10.3390/s21041254.

(Q1) Iparraguirre, O., Iturbe-Olleta, N., Brazalez, A., Borro, D., "Road marking damage detection based on deep learning for infrastructure evaluation in emerging autonomous driving". *IEEE Transactions on Intelligent Transportation Systems*. July 2022. doi: 10.1109/TITS.2022.3192916.



Article

Sensors on the Move: Onboard Camera-Based Real-Time Traffic Alerts Paving the Way for Cooperative Roads

Olatz Iparraguirre ^{1,2,*}, Aiert Amundarain ^{1,2}, Alfonso Brazalez ^{1,2} and Diego Borro ^{1,2}

¹ CEIT-Basque Research and Technology Alliance (BRTA), Manuel Lardizabal 15, 20018 Donostia/San Sebastián, Spain; aamundarain@ceit.es (A.A.); abrazalez@ceit.es (A.B.); dborro@ceit.es (D.B.)

² Universidad de Navarra, Tecnun, Manuel Lardizabal 13, 20018 Donostia/San Sebastián, Spain

* Correspondence: oiparraguirre@ceit.com

Abstract: European road safety has improved greatly in recent decades. However, the current numbers are still far away to reach the European Commission's road safety targets. In this context, Cooperative Intelligent Transport Systems (C-ITS) are expected to significantly improve road safety, traffic efficiency and comfort of driving, by helping the driver to make better decisions and adapt to the traffic situation. This paper puts forward two vision-based applications for traffic sign recognition (TSR) and real-time weather alerts, such as for fog-banks. These modules will support operators in road infrastructure maintenance tasks as well as drivers, giving them valuable information via C-ITS messages. Different state-of-the-art methods are analysed using both publicly available datasets (GTSB) as well as our own image databases (Ceit-TSR and Ceit-Foggy). The selected models for TSR implementation are based on Aggregated Channel Features (ACF) and Convolutional Neural Networks (CNN) that reach more than 90% accuracy in real time. Regarding fog detection, an image feature extraction method on different colour spaces is proposed to differentiate sunny, cloudy and foggy scenes, as well as its visibility level. Both applications are already running in an onboard probe vehicle system.

Keywords: Cooperative Intelligent Transport Systems (C-ITS); traffic sign recognition; fog detection; intelligent roads; V2X communications; road maintenance



Citation: Iparraguirre, O.; Amundarain, A.; Brazalez, A.; Borro, D. Sensors on the Move: Onboard Camera-Based Real-Time Traffic Alerts Paving the Way for Cooperative Roads. *Sensors* **2021**, *21*, 1254. <https://doi.org/10.3390/s21041254>

Academic Editor: Kyandoghe Kyamakya

Received: 15 January 2021
Accepted: 7 February 2021
Published: 10 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

No deaths and serious injuries on European roads by 2050. This is the goal established by the European Commission (EC). Meanwhile, EU road safety targets halving these numbers by 2030 [1]. The EU has seen a substantial decrease in road fatalities in the past, but these numbers have been stagnating in recent years. The latest research studies indicate that even with the lockdown due to the COVID-19 pandemic situation, deaths on the road did not decline by the same degree as traffic volume did [2].

To address this trend and meet the targets, the EC is committed to digital technologies. Future intelligent vehicles will interact with other vehicles and with the road infrastructure. This interaction is the domain of Cooperative Intelligent Systems (C-ITS) and is expected to significantly improve road safety, traffic efficiency, environmental performance and comfort driving, by helping the driver to make better decisions and adapt to the traffic situation [3]. Additionally, the EU road safety policy framework also focuses on infrastructure safety and the updating of some legislative measures.

Over the last decades, the remarkable development in technology and the increasing digitalisation made enormous advancements in the intelligent vehicles field. Its advanced electronics and mechatronics, communications and sensors made current vehicles scale-up levels of driving automation. However, there is still a long way to go before we are faced with fully autonomous and cooperative roads. The perception of the environment, given its

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS

1

Road Marking Damage Detection Based on Deep Learning for Infrastructure Evaluation in Emerging Autonomous Driving

Olatz Iparraguirre¹, Nagore Iturbe-Olleta¹, Alfonso Brazalez, and Diego Borro², *Member, IEEE*

Abstract—The future of autonomous driving is slowly approaching, but there are still many steps to take before it can become a reality. It is crucial to pay attention to road infrastructure, because without it, intelligent vehicles will not be able to operate reliably, and it will never be possible to dispense of driver's control. This paper presents the work carried out for the detection of road markings damage using computer vision techniques. This is a complex task for which there are currently not many papers and large image sets in the literature. This study uses images from the public Road Damage Detection dataset for the D44 defect and also provides 971 new labelled images for Spanish roads. For this purpose, three detectors based on deep learning architectures (Faster RCNN, SDD and EfficientDet) have been used and single-source and mixed-source models have been studied to find the model that best fits the target images. Finally, F1-score values reaching 0.929 and 0.934 have been obtained for Japanese and Spanish images respectively which improve the state-of-the-art results by 25%. It can be concluded that the results of this study are promising, although the collection of many more images will be necessary for the scientific community to continue advancing in the future in this field of research.

Index Terms—Autonomous driving, road damage detection, road infrastructure maintenance, deep learning object detector.

I. INTRODUCTION

INTELLIGENT vehicles technology is advancing at lightning speed. However, the future of fully autonomous mobility is still unclear. In the coming decades, vehicles with different driving automation levels will coexist on our roads and traffic will be mixed with human-driven vehicles. Nowadays, the most advanced systems on our roads are the so-called level 2 autonomous vehicles defined by the Society of Automotive Engineers (SAE) and level 3 vehicles will be introduced on the market in the near future. These vehicles provide support to the drivers with Advanced Driver Assistance Systems (ADAS) that allow an automated driving experience under specific circumstances and transfer the control to the driver in case they face an unexpected event.

Manuscript received 4 January 2022; revised 20 May 2022; accepted 6 July 2022. This work was supported by the Basque Government (Elkartek Program) through the Project "Autoev@a1: Evolución Tecnológica Para la Automatización Multivehicular y Evaluación De Funciones De Conducción Autónoma" under Grant KK-2021/00123. The Associate Editor for this article was Y. Hou. (Corresponding author: Olatz Iparraguirre.)
The authors are with the CEIT-Basque Research and Technology Alliance (BRTA), 20018 Donostia-San Sebastián, Spain, and also with the Universidad de Navarra, Tecnun, 20018 Donostia-San Sebastián Spain (e-mail: oiparraguirre@ceit.es; niturbe@ceit.es; abrazalez@ceit.es; dborro@ceit.es).
Digital Object Identifier 10.1109/TITS.2022.3192916

1558-0016 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

The main challenge for driving automation is the adaptability to all circumstances that may arise along the road.

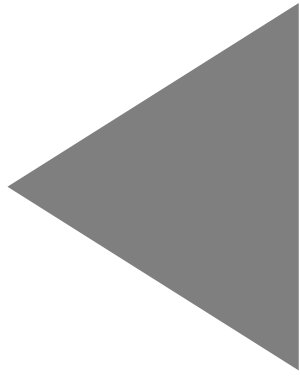
In this transformation towards automated mobility, many experts believe that roads, as we know them today with traffic control systems (signals, signs, markings), will continue to exist for at least another 50-75 years. The human driver will be able to take control of the vehicle until Level 4 and automated vehicles often directly interface with road markings to ensure that the vehicle stays on course. Therefore, regardless of the level of automation and as long as roadways continue to have a mix of human, machine and fully automated systems, there is every reason to believe that, far from disappearing, proper road markings and road maintenance practices will gain great importance in the future of the automotive industry.

In this context, there is a proposal for Smart Roads Classification (SRC), which studies the degree of adaptation to automated and/or connected driving. This classification would be valuable to determine which road sections are ready for any autonomous system and which ones are not. In order to carry out this assessment, both physical and digital infrastructure as well as the connectivity of the road itself come into play [1].

Digital infrastructure needs were set in the Commission Delegated Regulation (EU) 2015/962 (2015) which describes what information (in terms of infrastructure and traffic information) should be made available digitally and what the requirements are for updating this information. On the other hand, some efforts have been carried out for the physical infrastructure classification systems. The Connected Roadway Classification System (CRCS) was the first proposal that defines the attributes or conditions to be gathered (talking, seeing and simplifying) but does not specify any threshold for the specific road levels. Later on, SRC was defined based on two previous parameters: ISAD levels (Infrastructure Support Levels for Automated Driving) and LOSAD levels (Level of Service for Automated Driving). The first represents the readiness of the infrastructure to accommodate automated vehicles. Whereas the second one focuses on the support provided to connected vehicles (see Fig 1).

This SRC framework could help all stakeholders (road operators, administrations, vehicle manufacturers etc.) as well as users, to know what they can provide and expect from the infrastructure. On the one hand, drivers of vehicles with low levels of automation should have clear indications of where they can reliably activate their systems. On the other hand, higher levels of automation would benefit tremendously from digital information from the infrastructure and the environment, complementing what they can perceive through the

Appendix



A. Appendix – Smart Road Classification framework

Table A-1. Road categories in terms of automation and connectivity support [6]

LOSAD							
Continuous ORS. Not relevant disengagements	A	2 - Assistedway No information about dynamic parameters that could influence the ODD. Although the road segment would not trigger disengagements to Level 4 vehicles, the lack of information prevents them from an automated experience. Level 2 - 3 vehicles might experience very few disengagements.	4 - Full Automatedway Level 4 and 5 vehicles can perform autonomously. Level 2 and 3 might experience few disengagements.	5 - Autonomousway Cooperative driving. Recommended for critical road segments and junctions with traffic segregation to only allow CAVs.			
			4 - Full Automatedway The dynamic information provided by the road digital infrastructure can be used by Level 4 vehicles to manage ODDs and drive autonomously.				
Dynamic conditions may temporarily limit ORS effectiveness	B	2 - Assistedway Disengagements are not very frequent, but exist, Level 3 vehicles do not have information to foresee disengagements, so they operate like Level 2 vehicles.	3 - Automatedway The connectivity provided by the road digital infrastructure can be used by Level 3 vehicles to foresee disengagements and warn the driver to resume control in advance.				
			2 - Assistedway The physical road infrastructure triggers relatively frequent disengagements that might affect the automated experience. Under certain circumstances, drivers might be allowed to enable their driving automation systems. Due to their number, the connectivity level should not be used to foresee disengagements (for Level 3 vehicles).				
Non-continuous ORS. Most drivers may retake manual control sometimes	C	2 - Assistedway Disengagements are not very frequent, but exist, Level 3 vehicles do not have information to foresee disengagements, so they operate like Level 2 vehicles.	3 - Automatedway The connectivity provided by the road digital infrastructure can be used by Level 3 vehicles to foresee disengagements and warn the driver to resume control in advance.				
			2 - Assistedway The physical road infrastructure triggers relatively frequent disengagements that might affect the automated experience. Under certain circumstances, drivers might be allowed to enable their driving automation systems. Due to their number, the connectivity level should not be used to foresee disengagements (for Level 3 vehicles).				
No ORS or too short. No remarkable benefit of ADAS	D	2 - Assistedway Disengagements are not very frequent, but exist, Level 3 vehicles do not have information to foresee disengagements, so they operate like Level 2 vehicles.	3 - Automatedway The connectivity provided by the road digital infrastructure can be used by Level 3 vehicles to foresee disengagements and warn the driver to resume control in advance.				
			2 - Assistedway The physical road infrastructure triggers relatively frequent disengagements that might affect the automated experience. Under certain circumstances, drivers might be allowed to enable their driving automation systems. Due to their number, the connectivity level should not be used to foresee disengagements (for Level 3 vehicles).				
No ORS or too short. No remarkable benefit of ADAS	E	2 - Assistedway Disengagements are not very frequent, but exist, Level 3 vehicles do not have information to foresee disengagements, so they operate like Level 2 vehicles.	3 - Automatedway The connectivity provided by the road digital infrastructure can be used by Level 3 vehicles to foresee disengagements and warn the driver to resume control in advance.				
			1 - Humanway The road physical infrastructure triggers too many disengagements. Therefore, drivers are suggested to disable the driving automation systems. The connectivity capabilities of the road cannot be used to foresee or reduce the number of disengagements.				
		E	D	C	B	A	ISAD
		No information support	Static information support	Dynamic Digital Information	Cooperative perception	Cooperative driving	

B. Appendix – Colour spaces

In this section, there is a brief description of the three colour spaces that have been used in this work to extract characteristics from the image.

RGB colour space

This colour space is the most common one since the human eye only has colour-sensitive receptors for red (R), green (G) and blue (B). Thus, it is theoretically possible to decompose every visible colour into combinations of these three “primary colours” with different ranges of intensities from 0 to 255. Thus, this combination can be represented as a three-dimensional coordinate plane with the values for R, G and B on each axis (see Figure A-1 a). This way, it is concluded that when all channels have a value of zero, no light is emitted resulting in black colour. Whereas when all three colour channels are set to their maximum the resulting colour is white.

HSV colour space

HSV colour space is a cylindrical representation where colours of each hue (H) are arranged in a radial slice (see Figure A-1 b). The hue is the colour portion of the model that is expressed in 0-360 degrees where it can be considered colours like red, yellow, green, cyan, blue and magenta each in 60-degree increments. The central axis represents the value (V) or brightness of the colour from 0 to 100 per cent, where 0 is completely black and 100 is the brightest and reveals the most colour. Finally, saturation (S) describes the amount of grey in a particular colour, its minimum value (0) introduces more grey colour and the maximum (100) is the most similar to its primary colour.

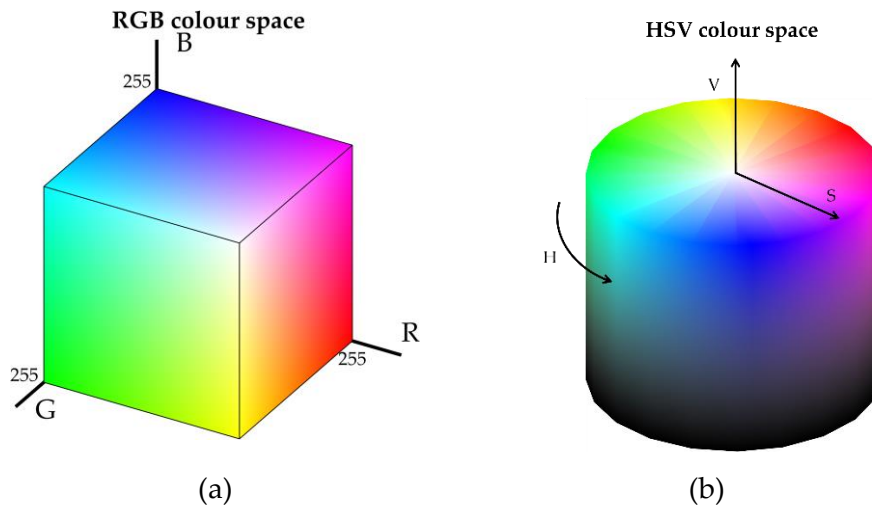


Figure A-1. Representation of (a) RGB and (b) HSV colour spaces.

XYZ colour space

The CIE XYZ colour space encloses all colour sensations that are visible to the human eye. The human eye with normal vision has three kinds of cone cells that sense light, S, M and L depending on their sensitivity to short, middle or long wavelength light. The CIE colour model takes the luminance (as a measure for perceived brightness) as one of the three colour coordinates, calling it Y. The spectral response of the luminance is specified as the photopic luminosity function. The maximum possible Y value, e.g. for a colour image, may be chosen to be 1 or 100, for example. The Z coordinate responds mostly to shorter-wavelength light (blue colours on the visible light spectrum), while X responds both to shorter- and longer-wavelength light [156]. Figure A-2 shows the used colour matching functions.

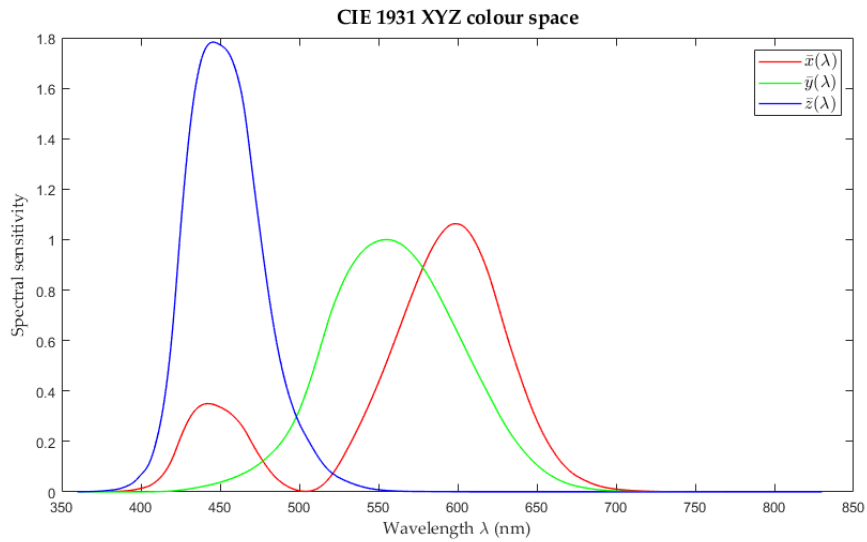


Figure A-2. The CIE XYZ standard observer colour matching functions.

Adapted from Jeff Mather (2021). Spectral and XYZ Color Functions (<https://www.mathworks.com/matlabcentral/fileexchange/7021-spectral-and-xyz-color-functions>), MATLAB Central File Exchange. Retrieved January 5, 2021

Appendix – TSR confusion matrixes

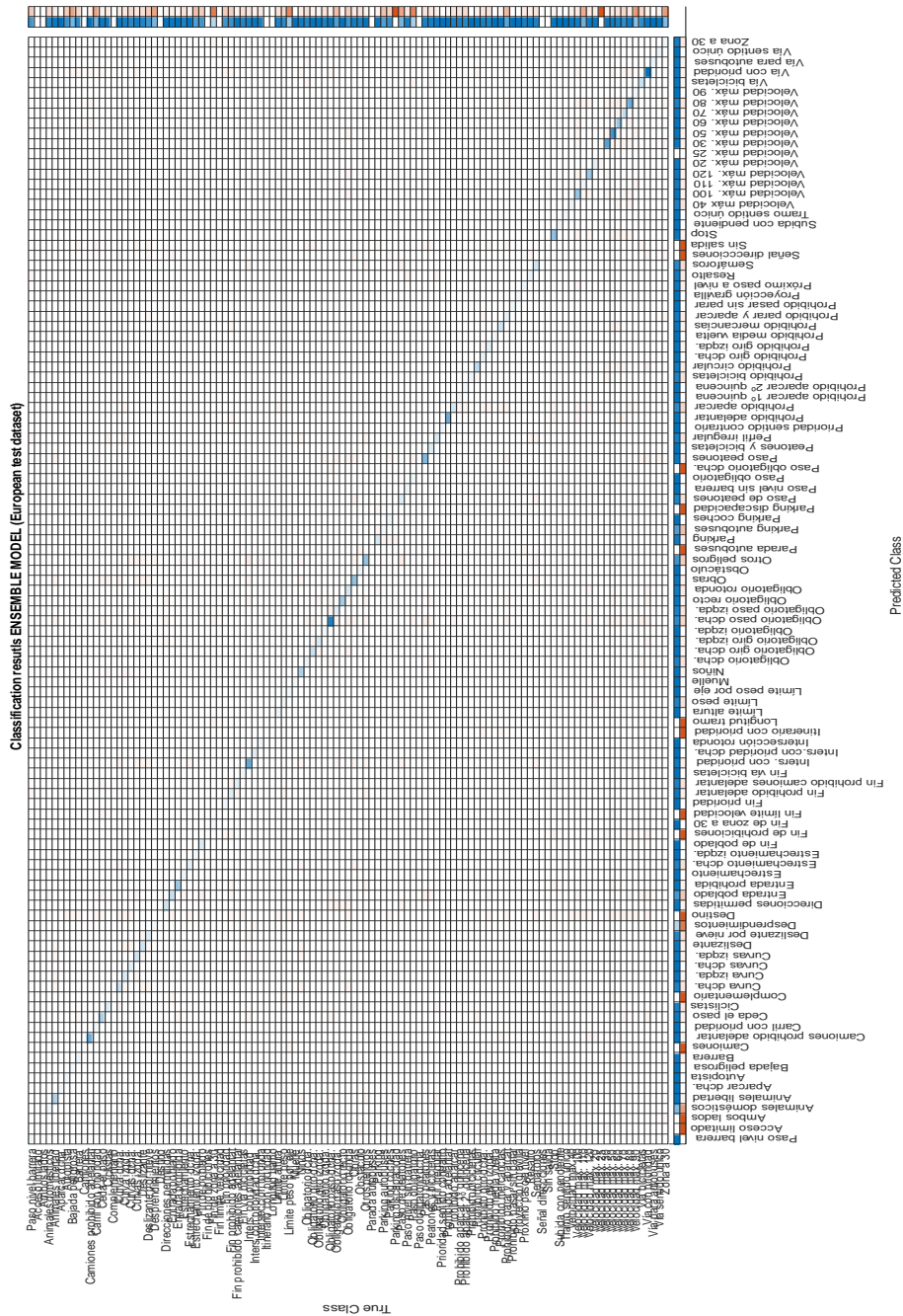


Figure A-6. Confusion matrix of ensemble classifier in ETSDB dataset.

D. Appendix – C-ITS messages

Table A-2. DENM message’s cause and subcause codes summary that covers all possible use cases for Road Hazard Warning (RHW) service.

DENM message - ETSI EN 302 637-3 V1.2.2 (2014-11)	
CauseCode	SubcauseCode
reserved(0),	
trafficCondition(1),	unavailable(0), increasedVolumeOfTraffic(1), trafficJamSlowlyIncreasing(2), trafficJamIncreasing(3), trafficJamStronglyIncreasing(4), trafficStationary(5), trafficJamSlightlyDecreasing(6), trafficJamDecreasing(7), trafficJamStronglyDecreasing(8)
accident(2),	unavailable(0), multiVehicleAccident(1), heavyAccident(2), accidentInvolvingLorry(3), accidentInvolvingBus(4), accidentInvolvingHazardousMaterials(5), accidentOnOppositeLane(6), unsecuredAccident(7), assistanceRequested(8)
roadworks(3),	unavailable(0), majorRoadworks(1), roadMarkingWork(2), slowMovingRoadMaintenance(3), shortTermStationaryRoadworks(4), streetCleaning(5), winterService(6)
adverseWeatherCondition-Adhesion(6),	unavailable(0), heavyFrostOnRoad(1), fuelOnRoad(2), mudOnRoad(3), snowOnRoad(4), iceOnRoad(5), blackIceOnRoad(6), oilOnRoad(7), looseChippings(8), instantBlackIce(9), roadsSalted(10)
hazardousLocation-SurfaceCondition(9),	unavailable(0), rockfalls(1), earthquakeDamage(2), sewerCollapse(3), subsidence(4) , snowDrifts(5), stormDamage(6), burstPipe(7) , volcanoEruption(8), fallingIce(9)
hazardousLocation-ObstacleOnTheRoad(10),	unavailable(0), shedLoad(1), partsOfVehicles(2), partsOfTyres(3), bigObjects(4), fallenTrees(5), hubCaps(6), waitingVehicles(7)
hazardousLocation-AnimalOnTheRoad(11),	unavailable(0), wildAnimals(1), herdOfAnimals(2), smallAnimals(3), largeAnimals(4)
humanPresenceOnTheRoad(12),	unavailable(0), childrenOnRoadway(1), cyclistOnRoadway(2), motorcyclistOnRoadway(3)
wrongWayDriving(14),	unavailable(0), wrongLane(1), wrongDirection(2)

Appendix – C-ITS messages

rescueAndRecoveryWorkInProgress(15),	unavailable(0), emergencyVehicles(1), rescueHelicopterLanding(2), policeActivityOngoing(3), medicalEmergencyOngoing(4), childAbductionInProgress(5)
adverseWeatherCondition- ExtremeWeatherCondition(17),	unavailable(0), strongWinds(1), damagingHail(2), hurricane(3), thunderstorm(4), tornado(5), blizzard(6)
adverseWeatherCondition-Visibility(18),	fog(1), smoke(2), heavySnowfall(3), heavyRain(4), heavyHail(5), lowSunGlare(6), sandstorms(7), swarmsOfInsects(8)
adverseWeatherCondition-Precipitation(19),	unavailable(0), heavyRain(1), heavySnowfall(2), softHail(3)
slowVehicle(26),	unavailable(0), maintenanceVehicle(1), vehiclesSlowingToLookAtAccident(2), abnormalLoad(3), abnormalWideLoad(4), convoy(5), snowplough(6), deicing(7), saltingVehicles(8)
dangerousEndOfQueue(27),	unavailable(0), suddenEndOfQueue(1), queueOverHill(2), queueAroundBend(3), queueInTunnel(4)
vehicleBreakdown(91),	unavailable(0), lackOfFuel(1), lackOfBatteryPower(2), engineProblem(3), transmissionProblem(4), engineCoolingProblem(5), brakingSystemProblem(6), steeringProblem(7), tyrePuncture(8)
postCrash(92),	unavailable(0), accidentWithoutECallTriggered(1), accidentWithECallManuallyTriggered(2), accidentWithECallAutomaticallyTriggered(3), accidentWithECallTriggeredWithoutAccessToCellul arNetwork(4)
humanProblem(93),	unavailable(0), glycemiaProblem(1), heartProblem(2)
stationaryVehicle(94),	unavailable(0), humanProblem(1), vehicleBreakdown(2), postCrash(3), publicTransportStop(4), carryingDangerousGoods(5)
emergencyVehicleApproaching(95),	unavailable(0), emergencyVehicleApproaching(1), prioritizedVehicleApproaching(2)
hazardousLocation-DangerousCurve(96),	unavailable(0), dangerousLeftTurnCurve(1), dangerousRightTurnCurve(2), multipleCurvesStartingWithUnknownTurningDirect ion(3), multipleCurvesStartingWithLeftTurn(4), multipleCurvesStartingWithRightTurn(5)
collisionRisk(97),	unavailable(0), longitudinalCollisionRisk(1), crossingCollisionRisk(2), lateralCollisionRisk(3), vulnerableRoadUser(4)
signalViolation(98),	unavailable(0), stopSignViolation(1), trafficLightViolation(2), turningRegulationViolation(3)

Appendix – C-ITS messages

dangerousSituation(99)	unavailable(0), emergencyElectronicBrakeEngaged(1), preCrashSystemEngaged(2), espEngaged(3), absEngaged(4), aebEngaged(5), brakeWarningEngaged(6), collisionRiskWarningEngaged(7)
------------------------	--
