Sveučilište u Zagrebu

Faculty of Geodesy

Ivan Brkić

OBJECT DETECTION ON GEOSPATIAL RASTER DATASETS IN SUPPORT OF ROAD SAFETY ASSESSMENTS

DOCTORAL DISSERTATION

Zagreb, 2023

Sveučilište u Zagrebu

Faculty of Geodesy

Ivan Brkić

OBJECT DETECTION ON GEOSPATIAL RASTER DATASETS IN SUPPORT OF ROAD SAFETY ASSESSMENTS

DOCTORAL DISSERTATION

Supervisor:
Assoc. Prof. Mario Miler, PhD

Zagreb, 2023

Sveučilište u Zagrebu

Geodetski fakultet

Ivan Brkić

DETEKCIJA OBJEKATA NA GEOPROSTORNIM RASTERSKIM PODLOGAMA RADI PROCJENE CESTOVNE SIGURNOSTI

DOKTORSKI RAD

Mentor:
Izv. prof. Mario Miler

Zagreb, 2023.

Declaration of authorship

UNIVERSITY OF ZAGREB

FACULTY OF GEODESY

Based on Article 19 of the Code of Ethics of the University of Zagreb and Decision no. 1_349_11 of the Faculty Council of the Faculty of Geodesy, University of Zagreb, from 26.10.2017. year (class: 643-03 / 16-07 / 03), the obligation to issue a "Declaration of Authenticity" for the doctoral dissertation which is evaluated at the postgraduate doctoral studies of geodesy and geoinformatics has been regulated in order to confirm that the work is the original result of student work and that the work does not contain sources other than those listed in them.

DECLARATION

I, Ivan Brkić, hereby declare that my doctoral thesis is the original result of my work and that I did not use any other sources other than those listed in it.

In Zagreb, October 18th, 2023

_____

Ivan Brkić

Acknowledgment

ZAHVALA.

Posveta.

Thesis information

I. Author

Name and Surname Ivan Brkić
Date of birth June 24th, 1994
Place of birth Split, Croatia

II. Supervisor

Assoc. Prof. Mario Miler, PhD
University of Zagreb, Faculty of Geodesy
Kačićeva 26, 10000 Zagreb, Croatia
mmiler@geof.unizg.hr

III. Study programme

University	University of Zagreb
Faculty	Faculty of Geodesy
Study programme	Postgraduate Doctoral Study of Geodesy and Geoinformatics
Scientific area	Technical Sciences
Scientific field	Geodesy

IV. Institution

University of Zagreb
Faculty of Geodesy
Kačićeva 26, 10000 Zagreb
Croatia
Tel: + 385 (1) 4639 222
Fax: + 385 (1) 4828 081
e-mail: pisarnica@geof.hr
WGS84: ϕ = 45°48'31.7'', λ = 15°57'49.0''
HTRS96/TM: *E*= 458 300 m, *N*= 5 074 470 m

V. Timeline and commission

Enrollment in the study programme	November 7th, 2019
Student ID number	PD-602
Defense of the theme	April 17th, 2023
Faculty Council report	May 25th, 2023
Theme approved by the University council	September 12th, 2023
Supervisors report on the Faculty Council	October 23rd, 2023
Evaluation committee	Full Prof. Damir Medak, PhD [a] Assist. Prof. Luka Rumora, PhD [a] Assoc. Prof. Marko Ševrović, PhD [b]
Thesis evaluation report on the Faculty Council	November 30th, 2023
Defense committee	Full Prof. Damir Medak, PhD [a] Assist. Prof. Luka Rumora, PhD [a] Assoc. Prof. Marko Ševrović, PhD [b]
Defense	
Thesis ID number	

[a] *Faculty of Geodesy, University of Zagreb, Zagreb, Croatia*
[b] *Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb, Croatia*

VI. Thesis in numbers

Doctoral thesis has 150 pages, 6 chapters, 35 figures, 14 tables and 166 cited references (duplicates removed).

VII. Citation

Brkić, I. (2023). Object Detection on Geospatial Raster Datasets in Support of Road Safety Assessments, Ph.D. thesis, Faculty of Geodesy, University of Zagreb.

VIII. Digital version

Available on the permanent link:
to be uploaded

Abstract

Statistics from the World Health Organization (WHO) and other global organisations point to a high number of road accidents, highlighting the need to improve road safety. Safety assessment of road infrastructure can be done by recording various road attributes. All completed studies and legal guidelines emphasise the need to establish a uniform, comprehensive framework for assessing the current state of road safety. Currently, one of the most widely used road assessment frameworks is the International Road Assessment Programme (iRAP), which defines 78 specific road attributes that affect road safety. Most of these attributes are spatial in nature and require accurate spatial identification. The current method for identifying these attributes involves georeferenced videos reviewed and annotated by accredited experts. However, this method has shortcomings, such as a very time-consuming process, and it can lead to inconsistencies when performed by multiple experts. To overcome these challenges, iRAP has developed the Advanced and Intelligent RAP (AiRAP) programme which aims to take advantage of modern technologies such as artificial intelligence, machine learning, deep learning, Lidar, etc.

In view of this, this dissertation explores recent deep learning techniques for object detection using three different geospatial datasets (UAV video, Light Detection and Ranging (Lidar), and very high-resolution satellite imagery) to address the aforementioned shortcomings of the prevailing road attributes coding method. Furthermore, the combination of different spatial data sources with recent object detection methods can lead to automation and increase the accuracy and consistency of the road attribute coding process. The work focuses on automating and improving the accuracy of the capture of six road attributes: traffic flow rate on a given road segment, roadside objects and their distance from the roadside, presence of a school zone on a given road segment, detection of four pedestrian crossing classes and identification of divided carriageways.

Keywords: Road safety, International Road Assessment Programme (iRAP), machine learning, deep learning, object detection, Faster R-CNN, Yolo, Unmanned Aerial Vehicles (UAVs), Lidar, very high-resolution satellite imagery.

Extended Abstract in Croatian language

Prema Svjetskoj zdravstvenoj organizaciji (WHO), ozljede u prometu vodeći su uzrok smrti među osobama u dobi od 5 do 29 godina. Nadalje, prema Europskom opservatoriju za sigurnost cestovnog prometa (ESRO), u Europskoj uniji (EU) 2020. godine bile su 42 smrtno stradale osobe u prometu na milijun stanovnika. Tri su glavna faktora koji utječu na nastanak prometne nesreće: vozač, vozilo te cestovna infrastruktura. Vozač ima najveći utjecaj na nastanak prometnih nesreća, a prati ga kombinacija utjecaja vozača i cestovne infrastrukture kao drugi najveći uzročnik prometnih nesreća. Kako bi se vrednovala kvaliteta postojeće infrastrukture cestovnih mreža, potrebno je definirati način vrednovanja. Iz toga je razloga, 1999. godine uspostavljen Europski program za procjenu cestovne sigurnosti (EuroRAP). EuroRAP se s vremenom proširio diljem svijeta. Stoga je 2006. osnovan Međunarodni program za procjenu cestovne sigurnosti (iRAP), koji je proizašao iz EuroRAP-a, a zamišljen je kao krovna strategija za programe procjene cestovne sigurnosti diljem svijeta. iRAP se temelji na vrednovanju cestovne sigurnosti putem zvjezdica. Pojedinoj cesti se dodjeljuje jedna do pet zvjezdica, a određuju se na temelju 78 cestovnih atributa. Službeni postupak prikupljanja cestovnih atributa temelji se na primjeni georeferenciranog videa. Obučeni stručnjaci pregledom georeferenciranog videa označavaju cestovne atribute duž cestovne mreže. Ovakav pristup ima dva nedostatka. Prvi nedostatak se odnosi na konzistentnost samog postupka označavanja. Iako iRAP jasno definira pojedine atribute, uvijek postoji mogućnost da će dva različita stručnjaka istu pojavu na cesti okarakterizirati različito. Na primjer, kamen uz cestu je definiran kao svaki objekt maksimalne visine do 60cm, dok svi veći kameni objekti se karakteriziraju kao kameni usjeci (engl. *rockface*). Nedostatak leži u činjenici da je ovakve, metrički definirane atribute, nemoguće definirati iz georeferenciranog videa. Iz navedenog je očito kako je teško osigurati konzistentnost samog procesa. Posebno je to izraženo kod većih cestovnih mreža na kojima sudjeluje više stručnjaka te je inkonzistentnost neizbježna. Drugi nedostatak ovakvog pristupa je to što je vremenski dugotrajan. S obzirom da stručnjak istovremeno pregledava video te označava primijećene atribute, dužina procesa direktno je vezana za brzinu vožnje kojom je cesta snimljena.

Zbog svega navedenog javlja se potreba za unaprjeđenjem samog procesa prikupljanja podataka. Kako bi se eliminirali gore opisani nedostatci, iRAP definira program Naprednog i inteligentnog programa za procjenu cestovne sigurnosti (AiRAP). AiRAP je zamišljen kao

program koji će se temeljiti na sintezi različitih izvora podataka (Lidar, satelitske snimke, telematika, itd.) te modernih tehnologija obrade podataka (strojno učenje, duboko učenje, itd.). U ovom radu istražene su mogućnosti kombinacije tri različita izvora prostornih podataka (snimke bespilotnih letjelica, Lidar i satelitske snimke) s metodama detekcije objekata u svrhu automatizacije, ubrzanja i osiguravanja konzistentnosti procesa prikupljanja cestovnih atributa. Metode detekcije objekata na rasterskim podlogama su se u prošlosti temeljile na analizi snimke, morfološkim transformacijama, detekciji rubova, računanju gradijenta piksela, itd. Razvoj računalnih resursa omogućio je primjenu konvolucijskih neuronskih mreža (CNN) u svrhu detekcije objekata. Tako se razlikuju jednorazinski (engl. *One-stage*) te dvorazinski (engl. *Two-stage*) detektori temeljeni na CNN-u. Jednorazinski detektori koriste jedan CNN koji je zadužen za detektiranje objekata, njihovu klasifikaciju te lokalizaciju na slici (određivanje slikovnih koordinata objekta). Najpoznatiji jednorazinski detektori su *You Only Look Once* (Yolo) grupa detektora, *Singe – Shot Detector* (SSD), RetinaNet, CornerNet, itd. Dvorazinski detektori pak koriste neki od vanjskih algoritama za izdvajanje dijelova slike koji potencijalno sadrže tražene objekte, a zatim se ti prijedlozi procesuiraju kroz CNN kako bi se dobila konačna klasa objekta te lokalizacija. Kao algoritmi za predlaganje dijelova s potencijalnim objektima mogu se koristiti tradicionalni algoritmi poput *Selective Search* algoritma ili manji CNN-ovi poput *Region Proposal Network* (RPN). Najpoznatiji dvorazinski detektori su oni iz grupe *Region – based* CNN (R – CNN). S obzirom da koriste samo jedan CNN, jednorazinski detektori su u pravilu brži, ali postižu manju točnost u odnosu na dvorazinske detektore.

Jedan od cestovnih atributa je i protok prometa koji definira količinu prometa na pojedinoj cesti. U ovoj disertaciji izrađen je i predložen okvir za obradu snimki dobivenih iz bespilotnih letjelica u svrhu određivanja parametara prometnog toka, uključujući i protok prometa. Predloženi okvir sastoji se od četiri dijela: terenska mjerenja, obrada videa (slika), detekcija vozila te računanje parametara prometnog toka. Terenska mjerenja odnose se na određivanje lokacija dviju točaka na terenu (primjenom globalnog navigacijskog satelitskog sustava - GNSS-a) te snimanje određenog segmenta ceste bespilotnom letjelicom postavljenom vertikalno iznad ceste. Dužina snimanja ovisi o dugotrajnosti baterije bespilotne letjelice. Obrada videa uključuje razdvajanje videa na slike te proces poravnanja svih slika s onom prvom kako bi se eliminirali pomaci bespilotne letjelice tokom snimanja. Detekcija vozila uključuje primjenu Faster R-CNN detektora na procesuiranim slikama te njihovo praćenje iz slike u sliku. Posljednji dio okvira odnosi se na računanje tri makro te četiri mikro parametra prometnog toka. Prednosti primjene ovakvog okvira su visoka točnost s obzirom na korištene tehnologije i metode (GNSS i Faster

R-CNN) kao i mogućnost određivanja svih sedam parametara prometnog toka jednim snimanjem. Isto tako, dosadašnje metode prikupljanja parametara prometnog toka poput induktivnih petlji zahtijevaju znatne građevinske zahvate te ne mogu prikupiti svih sedam parametara.

Dva parametra koja imaju značajan utjecaj na sigurnost cestovnog prometa su objekti uz cestu te njihova udaljenost od ruba ceste (engl. *Roadside Severity – Object* - RSS – O te *Roadside Severity – Distance* - RSS – D). Isto tako, navedeni atributi uzrokuju i velike probleme s inkonzistentnošću s obzirom da sadrže 17 klasa (13 ih postoji u Republici Hrvatskoj) koje su posebno definirane različitim metričkim vrijednostima. Za prikupljanje takvih atributa i njihovo klasificiranje izrađen je i predložen okvir koji se temelji na upotrebi mobilnog Lidar-a za prikupljanje oblaka točaka te Yolo v5 detektora. Okvir uključuje izradu 10-metarskih segmenata ceste, njihovu transformaciju u prostoru te konverziju u raster s ciljem izrade poprečnih presjeka cestovnih segmenata. Poprečni presjeci predstavljaju najbolju osnovu za detekciju objekata uz cestu te određivanje njihove udaljenosti od ruba ceste. U svrhu detekcije ceste kao i objekata uz cestu korišten je detektor Yolo v5 zbog brzine detekcije. Postupak prikupljanja ovakvih atributa je dugotrajan te je naglasak okvira na postizanju brzog i automatskog prikupljanja. Osim toga, okvir pruža mogućnost određivanja udaljenosti detektiranih objekata od ceste s visokom točnošću, što nije slučaj prilikom prikupljanja iz georeferenciranog videa.

Na kraju, u disertaciji je izrađen i predložen proces detekcije tri cestovna atributa iz satelitskih snimki visoke rezolucije. Proces se temelji na satelitskim snimkama rezolucije 30 cm, a fokusiran je na detekciju školskih zona, četiri vrste pješačkih prijelaza te detekciju fizički razdvojenih kolnika. Svi navedeni atributi su jasno definirani od strane iRAP-a. Proces uključuje izradu 20-metarskih cestovnih segmenata s preklopom od 30% kako bi se izbjegao gubitak podataka. Kao detektor je korišten Yolo v5 iz istog razloga kao i za detekciju objekata uz cestu. Prednosti predloženog procesa leže u korištenju satelitskih snimki. Naime, korištene snimke prikupljene su satelitom Pleiades Neo 3 koji opaža svaku točku na Zemljinoj površini dva puta dnevno. Na taj način je predloženim okvirom moguće automatski pratiti promjene za tri navedena atributa kroz duži vremenski period.

Ova disertacija predstavlja doprinos primjeni AiRAP-a u stvarnom svijetu. Izradom i predlaganjem opisanih okvira i procesa za prikupljanje cestovnih atributa potvrđeno je kako različiti izvori prostornih podataka u kombinaciji sa detektorima objekata temeljenim na CNN-u mogu poboljšati proces prikupljanja cestovnih atributa. S obzirom da su snimke bespilotne

letjelice, Lidar i satelitske snimke visoke rezolucije korištene za određivanje šest cestovnih atributa, buduća istraživanja mogu se fokusirati na prikupljanje preostalih cestovnih atributa

Ključne riječi: cestovna sigurnost, Međunarodni program za procjenu cestovne sigurnosti (iRAP), strojno učenje, duboko učenje, detekcija objekata, *Faster* R-CNN, Yolo, bespilotne letjelice, Lidar, satelitske snimke vrlo visoke rezolucije.

Table of Contents

Declaration of authorship ..... **i**
Acknowledgment ..... **ii**
Thesis information ..... **iv**
Abstract ..... **vi**
Extended Abstract in Croatian language **7**
Table of Contents **11**
1. Introduction **14**
1.1. Road Safety statistical indicators 15
1.2. Contributing factors to road accidents 16
1.2.1. Driver factors 16
1.2.2. Vehicle factors 17
1.2.3. Roadway factors 19
1.3. Recent action in improving road safety 19
1.3.1. Road infrastructure 20
1.4. International Road Assessment Programme (iRAP) 21
1.4.1. iRAP Star Rating 22
1.4.2. iRAP road attributes 22
1.5. Accelerated and Intelligent Road Assessment Programme (AiRAP) 24
1.6. Machine learning approach 26
1.6.1. Convolutional Neural Network (CNN) 27
1.7. Object detection 32
1.7.1. Faster R-CNN 33
1.7.2. Yolo 35
1.8. Thesis objectives 36
1.9. Expected scientific contributions 37
1.10. Chapter summary 37
2. An Analytical Framework for Accurate Traffic Flow Parameter Calculation from UAV Aerial Videos **39**

2.1. Introduction 41

2.1.1. Related Works 42

2.2. Data Collections and Methods 44

2.2.1. Terrain Survey 45

2.2.2. Image Processing 46

2.2.3. Vehicle Detection 47

2.2.4. Parameters Determination 50

2.3. Results 55

2.4. Discussion 61

2.5. Conclusion 64

3. Automatic Roadside Feature Detection Based on Lidar Road Cross Section Images 66

3.1. Introduction 68

3.1.1. Related Works 70

3.2. Materials and Methods 72

3.2.1. Object Detection 78

3.3. Results 81

3.3.1. Object Detection Evaluation 82

3.3.2. Spatial Accuracy of Detected Objects 84

3.3.3. Evaluation of road segments classification 85

3.4. Discussion 87

3.5. Conclusion 89

4. Utilizing High Resolution Satellite Imagery for Automated Road Infrastructure Safety Assessments 91

4.1. Introduction 93

4.1.1. Related Works 95

4.2. Materials and Methods 96

4.2.1. Vector Data Processing 98

4.2.2. Satellite Imagery Processing 99

4.2.3. Detection of School Zones, Pedestrian Crossings, and Divided Carriageways.. 100

4.3. Results 104

4.4. Discussion..... 109

4.5. Conclusions 111

Chapter 5 Joint Discussion 113

Chapter 6 Conclusion..... 119

Bibliography 122

Appendix A 135

List of Figures 143

List of Tables..... 146

Curriculum Vitae 147

Chapter 1
Introduction

Since the earliest days of human history, roads have played a crucial role in connecting people and places. Ancient civilisations such as the Romans and Greeks built roads for walking and for transporting goods using animals. The Romans, known for their impressive engineering skills, built vast networks of roads that were used for trade and even military operations. Some of these ancient roads still survive today, proving their continued importance and design [1]. Over time, the way travelled on roads evolved. The invention of the wheel was a crucial turning point as it made transport faster and more efficient. With the industrial revolution came another significant change with the invention of the car. This invention meant that roads had to be better planned, built, and maintained to cope with these new vehicles.

Roads are important for trade and help regions grow and prosper. They allow diverse cultures to meet and exchange ideas and play a big role in how societies have grown and developed. Today, a well-built and maintained road network is a sign of a country's economic strength. Roads support businesses, connect markets and provide jobs for people. Recent technological advances, especially the integration of artificial intelligence, have had a significant impact on road transport. This has led to more vehicles on the road, vehicles reaching higher speeds than was previously possible, and autonomous vehicles with artificial intelligence entering the market. At the same time, technological growth has social implications that lead to an increased rural exodus and subsequently to larger cities with high population densities. This urban population participates in traffic on a daily basis, either as pedestrians or using various vehicles [2]. For example, from 2012 to 2021, the number of registered vehicles in the United States (US) is expected to increase by 14%, the number of registered drivers by 10%, and Vehicle Miles Travelled (VMT) by 6% [3]. While these technological innovations offer benefits to road transport, they also bring challenges, with a focus on improving road safety [4]. For example, an increase in average speed is directly related to the likelihood of a crash and the severity of the consequences of the crash. Every 1% increase in average speed leads to a 4% increase in the risk of a fatal accident and a 3% increase in the risk of a serious accident. The risk of death for pedestrians hit by car fronts increases rapidly (by 4.5 times from 50 km/h to 65 km/h) [5].

Addressing these challenges requires rapid improvements in road safety systems and the swift establishment of an appropriate legal framework [6]. Such measures will ensure that road safety standards evolve in parallel with technological progress.

1.1. Road Safety statistical indicators

Statistical indicators published in recent years show that increasing road safety is a priority task. The United Nations General Assembly (UN) has set an ambitious target to halve the global number of deaths and injuries caused by road traffic crashes by 2030 [7]. This goal is crucial because road traffic injuries are the leading cause of death among people aged 5-29. Every year, about 1.3 million people die in road traffic crashes worldwide [8]. Alarmingly, more than half of these deaths involve vulnerable road users such as pedestrians, cyclists, and motorcyclists [5]. The global imbalance in road safety is also evident when one considers that 93% of all road deaths occur in low- and middle-income countries. This is even though these countries account for only about 60% of the total number of vehicles in the world [5].

Moreover, recent statistics from the US are equally worrying. Between 2020 and 2021, the number of traffic-related fatalities increased by 10%, from 39 007 to 42 939. Furthermore, if the number of traffic fatalities per 100 million VMT was observed, a 2% increase, from 1.34 in 2020 to 1.37 in 2021 is visible. A notable observation from 2021 shows that speed-related crashes were responsible for 29% of all traffic fatalities, which equates to 12 330 fatalities. This was an 8% increase from the previous year and the highest figure since 2007. In addition, 26 325 occupants of passenger vehicles died in 2021, a 10% increase from the 23 914 in 2020. In addition, approximately 2 092 541 occupants were injured, a 10% increase from 1 907 011 the previous year [3].

In Europe, data for 2021 reported approximately 19 900 deaths and over 900 000 injuries in road crashes for European Union (EU) Member States. A closer look at these figures revealed that rural roads were the most dangerous, accounting for 52.5% (10 451) of fatalities. In contrast, urban roads and motorways accounted for 38.7% (7 699) and 8.8% (1 747) of fatalities, respectively. Among the fatalities in these accidents, drivers and passengers of passenger cars formed the largest group with 44.6%, followed by pedestrians with 18.1% and motorcyclists and their passengers with 16.6%. However, over a longer period of time, there is also a silver lining. Over the last two decades, the number of road deaths in the EU has decreased significantly. From 2011 to 2021, the number of road deaths fell by 30.7%. This decrease was constant over the entire period. Nevertheless, it is worth noting that the EU's target of halving

the number of road deaths in 2020 compared to 2010 was not achieved, despite a significant reduction in traffic that year due to coronavirus restrictions [9].

Taking a closer look at individual member countries, the road safety scenario in Croatia shows unique findings. In 2019, Croatia recorded a total of 297 fatalities from reported road accidents. In the ranking of the 27 EU countries based on the number of road deaths per million inhabitants, Croatia ranks 4th, a ranking that highlights the increased risk associated with road traffic within the country [10]. A closer look at the data reveals that the distribution of road fatalities in Croatia deviates from the general trends in the EU in several respects. A relatively large proportion of fatalities are due to accidents on urban roads and accidents during the night. However, in contrast to the EU norm, Croatia reports a significantly lower proportion of fatalities and casualties among the elderly population (65 years and older) [10].

1.2. Contributing factors to road accidents

In order to reduce the number of road accidents and thus increase road safety, it is important to define the main causes of road accidents. Dr William Haddon Jr pointed out in 1970 that road accidents are influenced by three main categories: Factors related to the driver, aspects related to the vehicle, and elements related to the road and the roadside. According to the American Association of State Highway and Transportation Officials (AASHTO), the above categories are interrelated, as shown in Figure 1.1 [11]. While driver factors have largest part, combination of roadway and driver factors also has double figures stake in total number of road accidents.

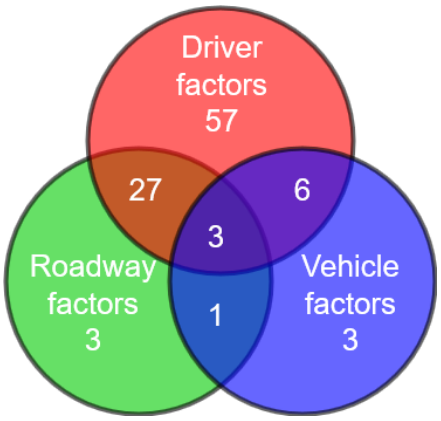


Figure 1.1 Contributing factors to road traffic accidents shown in percentage [11].

1.2.1. Driver factors

According to AASHTO driver factors can be divided into four sub-segments [11]:

- Attention and information processing – While driving, people can only process a certain amount of information. They often rely on previous experience to process the flood of new details they encounter. Information is best understood by drivers when it is consistent with their expectations, when it is given sequentially to ensure consistent demand, and when it helps to highlight important details.
- Vision – Approximately 90% of the data a driver relies on is visual in nature. Therefore, it is critical that this information is displayed with driver's varying visual abilities in mind so that they can perceive, understand, and respond to it effectively.
- Perception reaction time – The time and distance it takes a driver to react to certain influences, such as hazards on the road, road signs or signage, can vary. This depends on various human factors, such as ability to process information, level of attention, preconceived ideas, and visual acuity.
- Speed choice – Drivers depend on their perception and road indicators to choose a speed they think is safe. Peripheral vision often plays a role in picking up details that can influence a driver's decision to speed up or slow down based on proximity to objects on the road. It is also worth noting that drivers who have maintained a high speed on motorways may unknowingly drive faster than intended when moving into areas with lower speed limits.

To reduce driver factors and consequently reduce amount and hardness of traffic accidents, AASHTO provide next strategies:

- Education – Helps reduce accidents through awareness campaigns, driver training, and professional training for engineers and doctors.
- Policy/Legislation – Can lower accidents by dictating human behaviour and setting road and vehicle design standards, such as banning phone use, setting design requirements, and making helmets and seatbelts mandatory.
- Enforcement – Aims to decrease accidents by penalizing illegal actions like over-speeding and driving under the influence.

1.2.2. Vehicle factors

As far as vehicle factors are concerned, there are several programmes that aim to evaluate vehicle safety based on vehicle attributes. In 1979, National Highway Traffic Safety Administration (NHTSA) developed a New Car Assessment Programme (NCAP), which was

the first step towards improving vehicle safety in the US and worldwide. The programme was expanded into a 5-star rating system designed to rate the safety of cars from one star for unsafe cars to five stars for the safest cars. The stars are awarded based on three tests: Frontal, Side and Rollover tests [12]. In the tests, cars are collided under controlled conditions to simulate real-world traffic accidents. The tests take into account many components such as seat belts, airbags, electronic stability control, rear view camera, blind spot detection, driver assistance, etc.

Similar to the US, Euro NCAP was founded in Europe in 1996 by the Swedish Road Administration, the Federation Internationale de l'Automobile (FIA) and International Consumer Research & Testing. The programme is based on scoring in four key areas [13]:

- Adult Occupant Protection – This score indicates how well the car shields adult occupants during various impacts, such as frontal, side, and from whiplash. It also considers the ease and safety of rescue and extraction measures after an accident.
- Child Occupant Protection – This metric evaluates three main areas: how effectively the car's child restraint systems protect during frontal and side collisions; the adaptability of the vehicle to fit various child restraint designs; and the car's features ensuring children's safe transportation.
- Vulnerable Road User (VRU) Protection – Beyond assessing occupant safety, Euro NCAP measures a vehicle's safety provisions for vulnerable individuals like pedestrians and cyclists. They study potential injury risks to various body parts of a pedestrian, such as the head and legs. Vehicles that excel in these areas can earn extra points with an autonomous emergency braking (AEB) system that identifies and protects these VRUs.
- Safety Assist – This category evaluates critical driver assistance technologies that aid in safe driving and in reducing accident risks. Euro NCAP examines the effectiveness and reliability of these systems during regular driving and potential crash situations.

As a NHTSA CAP programme, Euro NCAP awards stars from one to five. The number of stars reflects how well the car performs in the Euro NCAP tests but is also influenced by the level of safety equipment in the vehicle. A higher number of stars not only shows that the test result was good, but also that the safety equipment of the tested model is readily available to all consumers in Europe.

1.2.3. Roadway factors

Roadway factors are the third most important factor contributing to road accidents. There are many road attributes have a significant impact on road safety. For example, Wang & Zhang (2017) investigated the relationship between road environmental factors and road crash severity using six road infrastructure characteristics: road class, location, road layout, lighting condition, road surface condition and speed limit [14]. In addition, Wong et al. (2007) [15] analysed the relationship between road crash outcomes and road attributes at signalised intersections. These include road attributes such as the number of approaches, the number of approach lanes, the number of conflict points, the number of traffic flows, the average lane width, the reciprocal of the average turning radius, proportion of commercial vehicles, number of signal stages, cycle time, number of pedestrian flows, presence of tram stops and presence of right turn pockets.
From all this, there is a need to define all road attributes that have a greater or lesser influence on the occurrence, severity, and outcome of a road accident. It is also necessary to define a uniform framework for assessing the safety of road infrastructure.

1.3. Recent action in improving road safety

The EU has set itself the long-term goal of almost eliminating road deaths by 2050 and halving the number of serious injuries by 2030 [16]. The Communication “Europe on the Move” proposes to adopt “Vision Zero”, which emphasises that no loss of life is acceptable. The European Commission (EC) aims to strengthen European road safety policy, engage stakeholders, and use research to develop new solutions. These efforts also contribute to the global debate on road safety by aligning with UN's Decade of Action for Road Safety 2010-2020 and embedding road safety in the Sustainable Development Goals.
An interim evaluation in 2015 found that EU road safety policy for the period 2011-2020 is generally moving in the right direction, with EU action adding value and likely to accelerate improvements, particularly in Member States with lower road safety standards [17]. However, further efforts and the completion of certain actions are needed to achieve the strategic objectives. The evaluation found that Member States could achieve immediate improvements, in particular through stricter enforcement of traffic rules such as speed limits.
It was noted that fatalities and serious injuries have not decreased at the same rate, suggesting that while fatalities can be prevented (e.g., through safer vehicles), accidents can still result in serious injuries. Therefore, the evaluation suggested that a separate target for reducing the number of serious injuries should be set alongside the existing fatality target. It was also

recommended to focus on measures to protect vulnerable road users and to ensure coordination with other policy areas such as environment, economy, health, and social affairs.

An updated technical study from 2018 reinforced many of these findings [18]. It recognised the potential impact of EU initiatives such as advanced braking for motorbikes, cross-border enforcement of traffic rules and the eCall emergency call system. In summary, the study recommends further refining the EU's road safety objectives with a broader and evidence-based strategy. This means focusing on the prevention of deaths and serious injuries, incorporating broader societal goals, setting new interim targets for "Vision Zero", and establishing key performance indicators (KPIs) for road safety at European level, especially in relation to the prevention of deaths and injuries.

1.3.1. Road infrastructure

More than 30% of accidents are due to road infrastructure and the surrounding area [19]. Properly designed roads, especially those with safety features such as medians, can significantly reduce the risk of accidents and their resulting severity. To achieve this, systematic risk mapping and safety assessment tools are needed that go beyond simply analysing accident-prone areas. The European Road Assessment Programme (EuroRAP) plays an important role in this. It provides safety ratings (from one to five stars) for roads in different EU countries [20]. While some countries have their own assessment methods, the EU's newly revised safety rules advocate risk mapping and safety assessment, especially for major road networks such as the Trans-European Transport Network (TEN – T). The EC aims to standardise the methodology in cooperation with member states. The updated rules also address the coming era of highly automated vehicles and emphasise the performance, placement and visibility of road signs and markings, which are essential for driver assistance systems.

These advances could potentially prevent 3 200 deaths and nearly 20 700 serious injuries by 2030, according to EC [20]. Finally, a challenge is to define a KPI for road infrastructure that measures the safety quality of roads regardless of user behaviour or technology. The EC intends to elaborate an indicator that considers the percentage of kilometres travelled on roads that exceed a safety assessment threshold. The basis of the KPI will be in network assessment and will consider the distance travelled. Furthermore, it is planned to integrate the indicator into the upcoming EU infrastructure safety standards.

In legislation form, the first legal frame is EU Directive 2008/96/EC on Road Infrastructure Safety Management (RISM) [21]. The central objective of this directive is to reduce the number of fatalities and serious injuries on EU road networks. This is achieved by focusing on

improving the safety features of road infrastructure. However, there is wide variation in the way the Directive has been implemented by Member States, with many countries performing very well, going beyond the requirements of the Directive, while others lag behind. Therefore, the original Directive 2008/96/ EC has been revised under EU Directive 2019/1936, a component of the third initiative EC entitled “Europe on the Move” [22]. EU Member States have been tasked with creating and applying specific procedures. These include road safety impact assessments, safety audits, safety inspections, comprehensive safety assessments of the road network and the continuous improvement of these methods in conjunction with the exchange of best practises. The updated guideline now applies to roads part of the TEN -T, motorways, and major roads. This covers all phases of the road life cycle: Design, construction, and operation. In addition, roads outside urban areas that do not directly serve adjacent properties and are built with EU funds are also covered by this directive. However, roads not accessible to the public are excluded. In addition, the directive sets a deadline of 2024 for EU member states to carry out a thorough road safety assessment. Thereafter, according to the directive's guidelines, they are to assess the entire road network in operation every five years. The assessment process includes visual inspections of the design elements of roads and an in-depth analysis of road sections where the number of accidents is disproportionately high compared to the volume of traffic. Roads flagged in the assessment reports are either subject to targeted safety inspections or other necessary corrective measures. Finally, by the end of December 2021, EU countries are required to inform EC about the totality of major roads and motorways in their jurisdiction. Roads exempted due to proven low safety risks must also be reported.

1.4. International Road Assessment Programme (iRAP)

All completed studies and legal guidelines emphasise the need to establish a uniform, comprehensive framework for assessing the current state of road infrastructure. in 1999, the British, Dutch, and Swedish governments, with the support of European mobility clubs and safety charities, launched EuroRAP. The main objective of the programme was to focus on road infrastructure as one of the factors contributing to road safety. The same programme was launched in Australia (AusRAP) and the United States (usRAP) in 2000 and 2003, respectively. In 2006, EuroRAP established the International Road Assessment Programme (iRAP) to serve as an umbrella programme for all road assessment programmes (RAPs) in the world. Today, iRAP is active in over 110 countries worldwide, covering Europe, Asia-Pacific, North America, Latin America and the Caribbean, and Africa.

1.4.1. iRAP Star Rating

In 2004, following the example of NCAP, the iRAP Star Rating Protocol was introduced to rate road infrastructure. The protocol assesses the characteristics of road infrastructure that influence the likelihood of an accident and its potential severity. Roads are rated from one to five stars based on their safety characteristics. The safest roads, rated with four and five stars, have safety features suitable for average traffic speeds. These features include a wide median or barrier separating oncoming traffic, clear markings, efficiently designed intersections, generous lanes, paved shoulder, and roadsides without unguarded hazards such as poles. They are also well suited for cyclists and pedestrians, with footpaths, cycle lanes and pedestrian crossings. On the other hand, roads rated one or two stars lack these safety measures. They often consist of single-lane roads with numerous curves and intersections, narrow pavements, unpaved shoulder, and unclear markings. Hidden intersections and exposed roadside hazards such as trees and steep edges are often. In addition, they do not provide adequate facilities for cyclists and pedestrians, as there are no sidewalks, cycle tracks or pedestrian crossings [23].
In summary, definiton of a high number of road attributes and their clear definition is obligatory. As mentioned above, most EU Member States base their road assessment on EuroRAP, which includes the use of the attributes defined in EuroRAP.

1.4.2. iRAP road attributes

The iRAP road attributes are the core of the iRAP Star Rating protocol. The road attributes are collected during road inspections, which consist of two components:

- Road survey, where images (or videos) of the roads, GPS data, and distance information are compiled,
- Road coding, in which road features are documented using the captured images (or footage).

In road surveys, the main purpose is to collect georeferenced images of roads that can be used for coding road attribute. Sampling of traffic flows, pedestrian flows, cyclist flows and speeds on roads is also usually carried out during a survey. A range of technologies and systems can be used to georeferenced, but all must be accredited by iRAP. The minimum requirements for imagery collected during a survey are as follows are [24]:

- Image resolution of 1280 x 960 pixels.

- 140-degree field of view (centred on the travel lane). This may be accomplished with either a single camera or with multiple cameras with overlapping fields of view.
- Continuous footage (or for still images, captured at fixed intervals of 20m or less) from a height of no less than 1.2m in height from the road surface.
- Recorded with latitude and longitude data in WGS84 projection and decimal degrees units. Latitude and longitude data are recorded with a minimum accuracy of < +/- 10 meters for at least 90% of images and must not "drop-out" for any more than 500 meters at a time.
- The road and roadsides are clearly visible. Images should not be compromised by factors such as rapid changes in shade/sun, poor lighting, dirty camera lenses, fog, and blurring.
- Forward space within the images is kept clear of vehicles as much as possible to ensure the required attributes can be viewed and assessed. This may require an escort for congested urban areas.
- Recorded for a minimum of 500 meters before the official start point and 500 meters after the official end point of each section of road. Road sections should be segmented in accordance with advice from the relevant road authority.
- Collected for all divided carriageway road lengths (surveyed in both directions) regardless of length.

Also, following data must be collected for each georeferenced image:

- Unique image number,
- Road name,
- Road section,
- Distance along the road,
- Section length,
- Date,
- Time,
- Latitude and longitude.

In terms of road coding process, the primary purpose of road attribute coding is to use georeferenced images collected during a survey or road designs to record road attributes for each 100m segment of road. A set of attributes is captured for each 100-metre segment of the road. If the condition of an attribute changes within a 100-metre span, the more severe scenario in terms of road safety is documented. For example, if the first 50 metres of a segment have

roadside safety barriers and the following 50 metres have roadside hazards, the entire 100-metre segment is marked as roadside hazards [25]. Coding process is performing after survey process, in the appropriate coding system. iRAP does not define specific systems for coding. Example of coding system interface is shown in Figure 1.2. Georeferenced image is at the heart of the interface. The geolocation of the road segment is displayed in the upper left corner. Road attributes are displayed at the bottom and right of the interface. road attributes coder is trained to simultaneously georeferenced video and code the road attributes as they appear in the video.

Figure 1.2 Example of coding system interface

There are 78 road attributes in the coding process, divided into two categories: those that can be collected from road survey data (52 attributes) and those that cannot be collected from road survey data (14 attributes). These are, for example, attributes such as traffic flow (annual average daily traffic – AADT) and peak hour pedestrian traffic on the road. The remaining 12 attributes refer to the name of the coder, the x and y coordinates of the geolocation, the road name, the date of the road survey, the coding date, etc. [26]. For all of road attribute definitions and options see Appendix A.

1.5. Accelerated and Intelligent Road Assessment Programme (AiRAP)

Although the iRAP coding manual details the coding team's procedures to standardise the criteria for assessing road attributes, such an approach still has shortcomings. Practise has shown that despite all efforts to standardise the coding of road attributes, there are still a variety

of situations where two different coders characterise the same object differently. For example, a larger rock on the side of the road may be defined as a rock face or as a large boulder. Although the attribute types are clearly defined, the definitions often refer to metric values such as the height or width of an object. In a georeferenced video it is almost impossible to define metric values and therefore the same objects are interpreted differently. In addition, there are road attributes that directly refer to metric values, such as the Roadside Strength – Distance (RSS – D) attribute, whose classes are defined as < 1m, 1-5m, 5-10m and > 10m. From the above shortcomings of georeferenced video, it is clear that multiple coders will place the same object in a different class, as accurate distance measurement is not possible. Furthermore, the process of encoding georeferenced video fatigues the coder, which can lead to a decrease in concentration, fatigue and ultimately a decrease in the quality of the encoded attributes. Although the coding manual clearly defines coding rules that specify both coding time and rest time for the coder, the rules are not universal for every coder considering their cognitive abilities. Therefore, iRAP launched the AiRAP initiative in 2019 to improve the use and accessibility of current and emerging global data sources. This includes breakthroughs in areas such as artificial intelligence, machine learning, vision systems, Lidar, telematics, and others. The name "AiRAP" stands for the advanced and intelligent collection of data relevant to road safety. This is achieved through automated, consistent, and scalable techniques that help assess road safety, map accident risk and prioritise investments for the benefit of all road users. Figure 1.3 shows data sources and road attributes which can be collected by each source proposed by AiRAP.

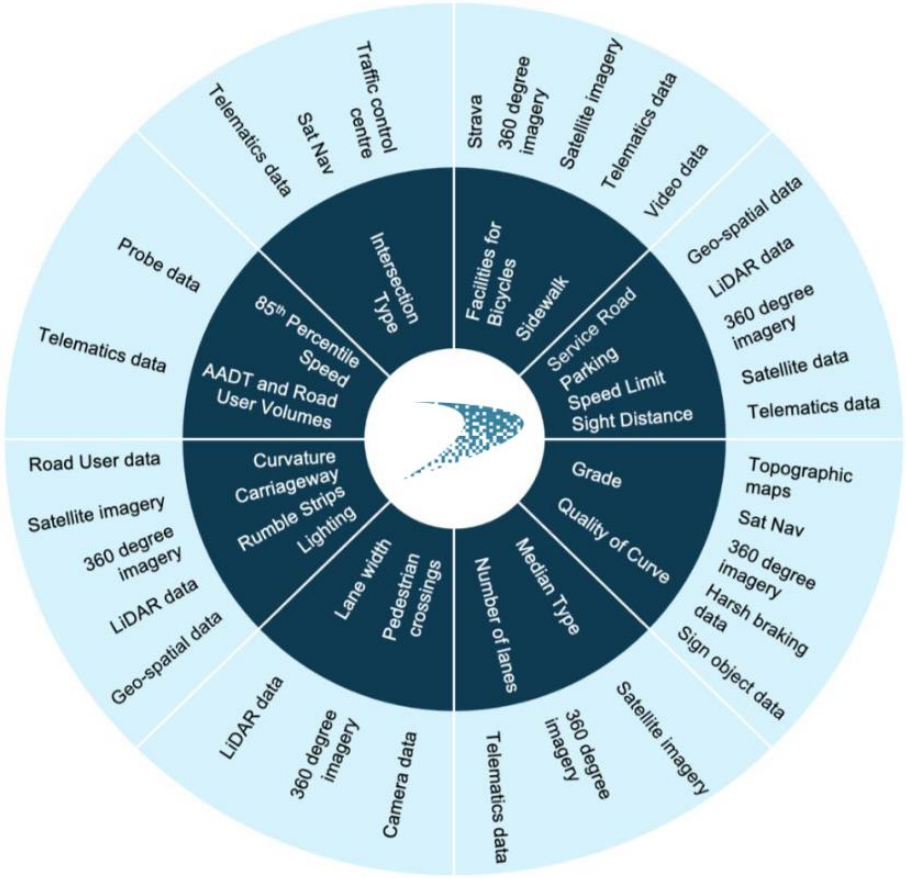


Figure 1.3 Data sources and road attributes which can be collected by each source proposed by AiRAP [27]

1.6. Machine learning approach

According to the AiRAP definition, machine learning is identified as one of the leading technologies in the future development of AiRAP. According to Alpaydin (2014), machine learning is defined as the process of programming computers to optimise a performance criterion based on sample data or previous experience [28]. The development of machine learning began with the development of more powerful computational resources as well as the increase in the amount of data at the core of machine learning. The more high-quality data fed into the system, the better the learning and the predictions or decisions that result. A machine learning model learns patterns in the data through a training process in which it is exposed to large amounts of data, iteratively makes predictions on the training data, and is corrected when the predictions are wrong so that the model learns over time. While machine learning generally aims to find patterns in training datasets to solve classification and regression tasks, there is a subfield of machine learning called Deep Learning that aims to process more complex datasets

such as images, point clouds, speech data, natural language, etc. Deep Learning is defined as a class of machine learning algorithms that use multiple layers to progressively extract higher-order features from raw input [29]. Most deep learning models are based on multi-layered artificial neural networks such as convolutional neural networks (CNN) and transformers.

1.6.1. Convolutional Neural Network (CNN)

CNNs are a class of deep learning models used primarily in computer vision to process and analyse visual data such as images and videos. They are inspired by the biological processes of the human brain, in particular the way human's visual cortex processes information. CNNs have helped to achieve peak performance in various computer vision tasks such as image classification, object detection and image segmentation. CNN's architecture is designed to automatically and adaptively learn spatial hierarchies of features from input images. Unlike conventional fully concatenated neural networks, CNNs have a unique architecture that significantly reduces the number of parameters, allowing them to be trained efficiently. A CNN consists of several layers, including convolutional layers, pooling layers, and fully connected layers, each of which performs a specific task.

The convolutional layer is the central building block of a CNN. It applies convolutional operations to the input and passes the result to the next layer. This operation allows the network to focus on the local patterns in the input data. Using multiple philtres, the network can learn and represent a variety of features in the data. The philtres glide over the input data, perform element-by-element multiplication and sum the results to create a feature map that highlights the presence of certain features in the input data. The values in the feature map must pass through the activation function, which must be defined for each convolution layer. The main role of activation functions is to introduce non-linearity into CNN, which enables CNN to learn complex patterns. There are numerous activation functions such as Sigmoid, Softmax, Rectified Linear Unit (ReLU), Hyperbolic Tangent (tanh), etc. [30]. In some cases, the activation function may be linear, which means that the input values remain unchanged. If all activation functions in the CNN are linear, the CNN becomes a linear regression model only, which limits the potential of the CNN to extract complex patterns in the dataset. A simple example of a convolutional layer with a 3x3 pixel philtre applied to a single channel, 6x6 pixel input image is shown in Figure 1.4. The convolution process begins by applying the filter (blue field) to the green subfield of the input image by multiplying and summing the values element by element. The resulting value passes through the linear activation function (remains unchanged) and is stored in the feature map as the first element (green cell). In the next step, the same process is

applied to the red sub-array. The step between the green and the red sub-array is called stride and is 1 in this example.

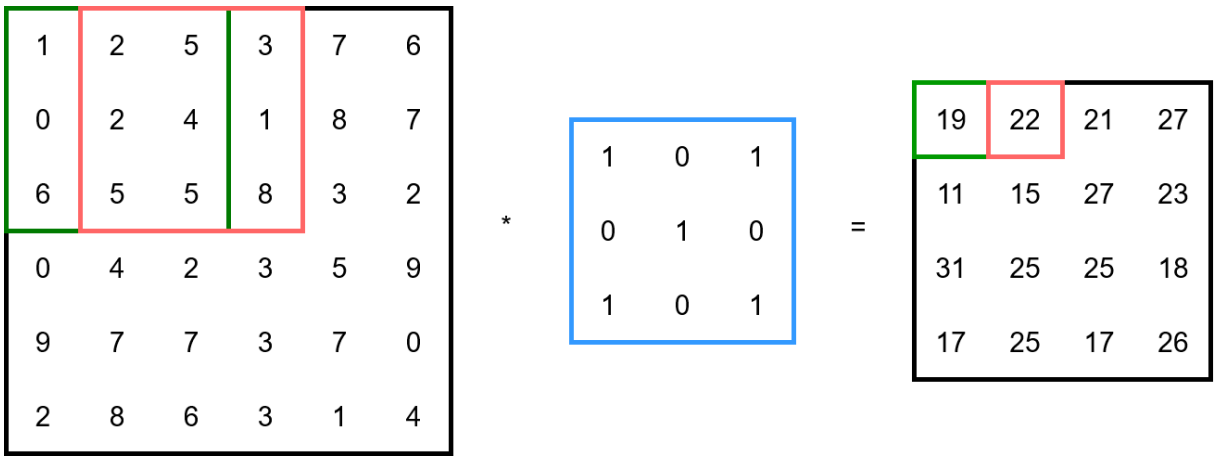


Figure 1.4 A simple example of one convolutional layer with one 3x3 pixels filter applied on input image with one channel and size of 6x6 pixels. With stride 1, output array (feature map) is 4x4 pixels.

Following the convolutional layers are the pooling layers, which reduce the spatial dimensions of the input data, thus decreasing the computational complexity of the network. Pooling layers perform down-sampling operations such as max-pooling, where the maximum value is taken from a group of values, or average pooling, where the average value is taken. A simple example of max-pooling layer is shown in Figure 1.5a, while average pooling is shown in Figure 1.5b. In both figures the pooling process is applied on the yellow sub-array of the feature map.

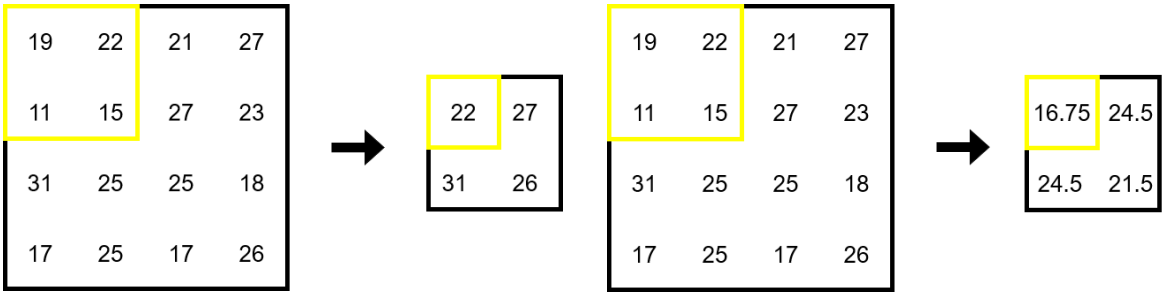


Figure 1.5 (a) max-pooling process applied on the feature map; (b) average pooling process applied on the feature map.

As the data traverses the network, the convolutional layer, and the pooling layer work together to learn and represent increasingly complex features. By the time the data reaches the fully

linked layers, i.e., the traditional neural network layers. They have been transformed into a form where the network can learn to make predictions, perform classifications or other tasks.

Towards the end of the network, there are fully linked layers that combine the high-level features learned from the previous layers into a single vector that can be used for classification or regression tasks. Finally, depending on the task, a specific activation function is applied to the values of the last layer.

The next component of the CNN training process is backpropagation. The predicted values must be compared to the ground truth values. This can be done by applying a loss function. There are a variety of loss functions suitable for different tasks. For regression tasks, the most useful are mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), Huber loss, etc. Classification tasks are usually performed with Binary Cross-Entropy Loss, Hinge Loss, Squared Hinge Loss, Multi-Class Cross-Entropy Loss, Focal Loss, etc. [31]. Each loss function has its pros and cons and can be useful for specific problems. For example, Focal Loss is suitable for binary classification with highly imbalanced dataset.

The most important step in backpropagation is to propagate the error backwards through the network to update the weights of the model. The values in the philtres of the convolutional layers are considered as weights. This process starts with the output layer and moves backwards to the input layer. For each layer, the algorithm calculates the gradient of loss with respect to the weights. This gradient indicates how much a change in each weight would affect the total error. After the gradients are calculated for each layer, the model uses an optimisation algorithm, usually gradient descent or one of its variants, to update the weights. Some of the most commonly used optimisers are Stohastic Gradient Descent (SGD), Adaptive Gradient Algorithm (AdaGrad), Root Mean Square Propagation (RMSProp), Adaptive Moment Estimation (Adam), etc. [32]. The goal of every optimiser is to adjust the weights so that the error, i.e., the loss function, is minimised. Also, each optimiser has its own hyperparameters and learning rates that need to be adjusted to achieve the best performance for specific tasks.

There are many tasks that can be solved by CNNs, such as:

- Image Classification – Identifying and categorizing objects or scenes in images into predefined classes or categories. For example, recognizing whether an image contains a car or a truck (Figure 1.6a).
- Object Detection – Locating and identifying multiple objects within an image while specifying their positions and classes. This is crucial for applications like autonomous driving and surveillance (Figure 1.6b).
- Semantic Segmentation – Assigning a class label to each pixel in an image, effectively dividing it into regions corresponding to different objects or object parts. It is used in applications like medical image analysis and scene understanding (Figure 1.6c).
- Instance Segmentation – Going beyond semantic segmentation, instance segmentation distinguishes individual instances of the same class. For example, identifying different cars in a parking lot (Figure 1.6d).

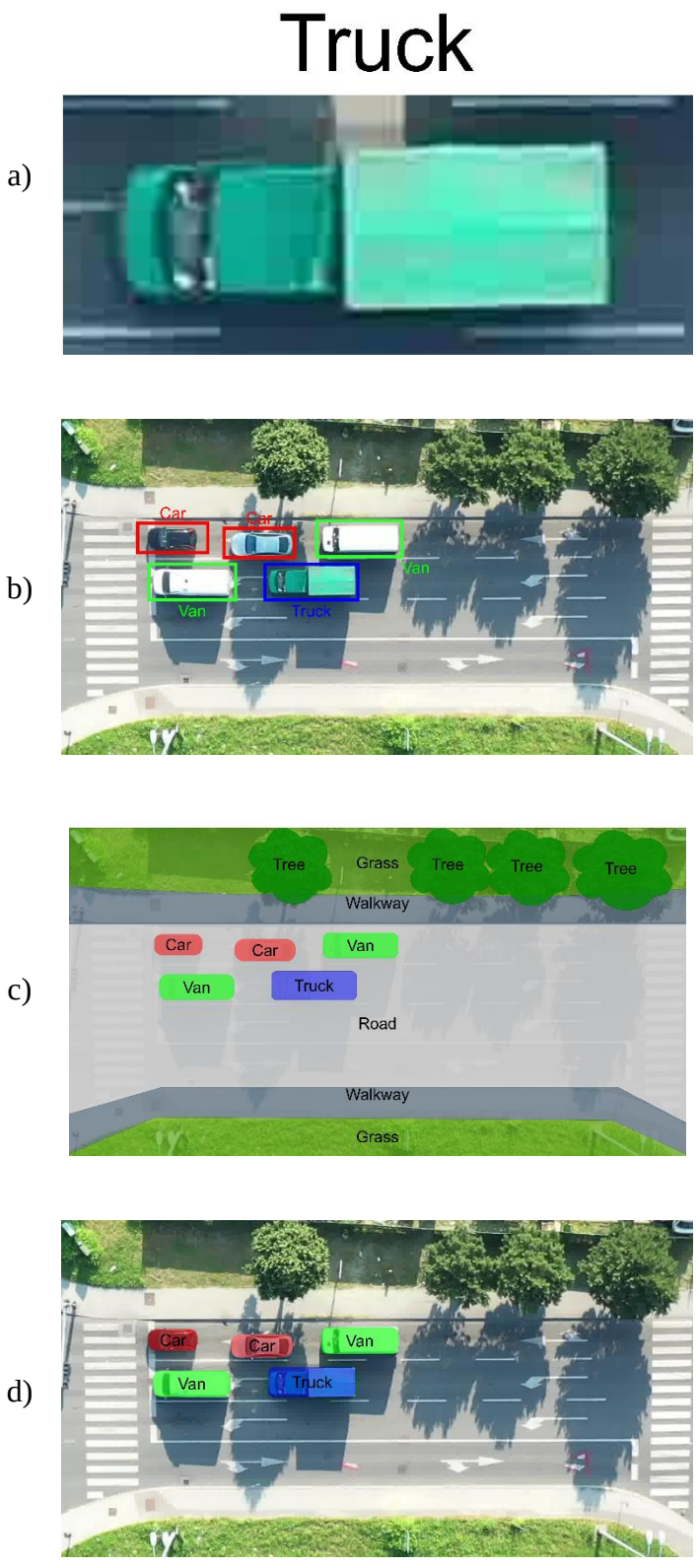


Figure 1.6 Example of four different computer vision tasks which can be solved by CNN; (a) Image classification; (b) Object detection; (c) Semantic segmentation; (d) Instance segmentation.

1.7. Object detection

Object detection is a computer vision task that involves identifying and locating multiple objects of interest in an image or video frame. The goal of object detection is not only to determine which objects are present in the scene, but also to locate them accurately by drawing bounding boxes around each detected object and associating them with specific object classes or labels. This task is fundamental to various applications in different industries, including autonomous driving, surveillance, robotics, and augmented reality.

Before the rapid development of computing resources that enabled the use of CNNs, various traditional methods were used for object detection [33]. The General Hough transform, introduced by Dana Harry Ballard in 1981, is considered a valuable technique for performing geometric feature extraction [34]. Another method, as exemplified by the Harris corner detector proposed in 1988, focuses on extracting object features by identifying corners within the image. This technique detects corner features in two images and computes the correlation between corresponding points to identify objects [33]. However, it is essential to note that these initial two methods are sensitive to the geometric characteristics of the image, including changes in size, rotation, and grayscale values, which can significantly influence the results [33]. The third approach is the SIFT (Scale and Rotation Invariant) method, introduced by Lowe in 2004 [35]. The SIFT algorithm was developed especially for the recognition and description of local features in images. It is characterised by treating each feature as an independent object, which ensures that variations in image rotation and scale do not negatively affect the results. This property increases the robustness of the algorithm and makes it suitable for various applications in computer vision and image analysis. All of the above methods have notable limitations when dealing with complex tasks, especially in terms of accuracy and performance. Therefore, the development of object detection in the last decade has shifted to the use of various CNN approaches. These approaches can be roughly divided into two-stage and single-stage detectors. Among the most widely adopted one-stage detectors today are the You Only Look Once (Yolo) family of detectors [36–41], Single Shot Detector (SSD) [42], RetinaNet [43], CornerNet [44], Fully Convolutional One-Stage Object Detection (FCOS) [45], and others [46]. These detectors directly suggest regions of interest (ROIs) from the input image, making them notably time-efficient and suitable for real-time applications.

On the other hand, two-stage detectors take a different approach. They initially generate a substantial number of ROIs, often using techniques like the Region Proposal Network (RPN) or Features Pyramid Network (FPN). Following this, CNN evaluates each ROI individually.

Prominent examples of two-stage detectors include those from the Region-based Convolutional Neural Networks (R-CNN) family, such as R-CNN [47], Fast-R-CNN [48], and Faster R-CNN [49], as well as Spatial Pyramid Pooling (SPP) Net [50], Mask R-CNN [51], and more. These two-stage detectors tend to excel in accuracy but are comparatively more complex.
In the next section Faster R-CNN and Yolo family of detectors will be explained in detail as detectors used in this thesis.

1.7.1. Faster R-CNN

Faster R-CNN is a recent object detector designed to enhance both the accuracy and efficiency of object detection tasks. It was introduced in 2015 by Ren et al. [49], building upon the foundations of previous object detection methods like R-CNN and Fast R-CNN, while effectively tackling their limitations.
A pivotal innovation in Faster R-CNN is the inclusion of a Region Proposal Network (RPN) within the model architecture. The RPN plays a critical role by eliminating the need for external methods like Selective Search [52], commonly used in R-CNN and Fast-R-CNN. The RPN operates as a fully convolutional network, processing feature maps derived from a CNN backbone, which is usually a pre-trained network such as VGG16, Inception or ResNet. The CNN backbone is responsible for extracting features from the input image. It consists of a series of convolutional layers and pooling layers, followed by fully connected layers. These layers encode hierarchical and abstract features from the image crucial for the subsequent object detection steps. The feature maps generated by the CNN backbone provide an input to the RPN. The goal of the RPN is to generate a region proposals accompanied by confidence values. To carry out this step, standard regions (anchors) must be defined. Typically, these are nine anchors for each pixel in the feature map. The goal of the RPN is to find out if some of the anchors contain an object. Therefore, the RPN has two output branches. The first is the classification branch, which has a depth of 2 * number of anchors and contains confidence values for each anchor. The confidence values indicate how likely it is that an object is present in each anchor. The second branch is the regression branch, which has a depth of 4 * number of anchors, where 4 represents the coordinates of the centre, the width and the height of the anchor. If the classification branch indicates that an anchor contains an object, the anchor is considered a proposed region. In this case, the regression branch of the RPN aims to approximate the boundaries of the anchor to the actual boundaries of the object. These region proposals serve as key points on which the model focuses its attention for further analysis.

After region proposals are generated, ROI pooling is used to align these suggestions to a fixed size, ensuring compatibility with fully linked layers. ROI pooling preserves the spatial information within each region suggestion, which is essential for accurate object localisation. The final steps involve separate branches of the network: a classification branch and a regression branch. The classification branch assigns class labels to each region proposal, determining the specific object category to which it belongs. At the same time, the regression branch refines the bounding box coordinates for each region proposal, improving the accuracy of object localisation. The architecture of the Faster R-CNN is shown in Figure 1.7.

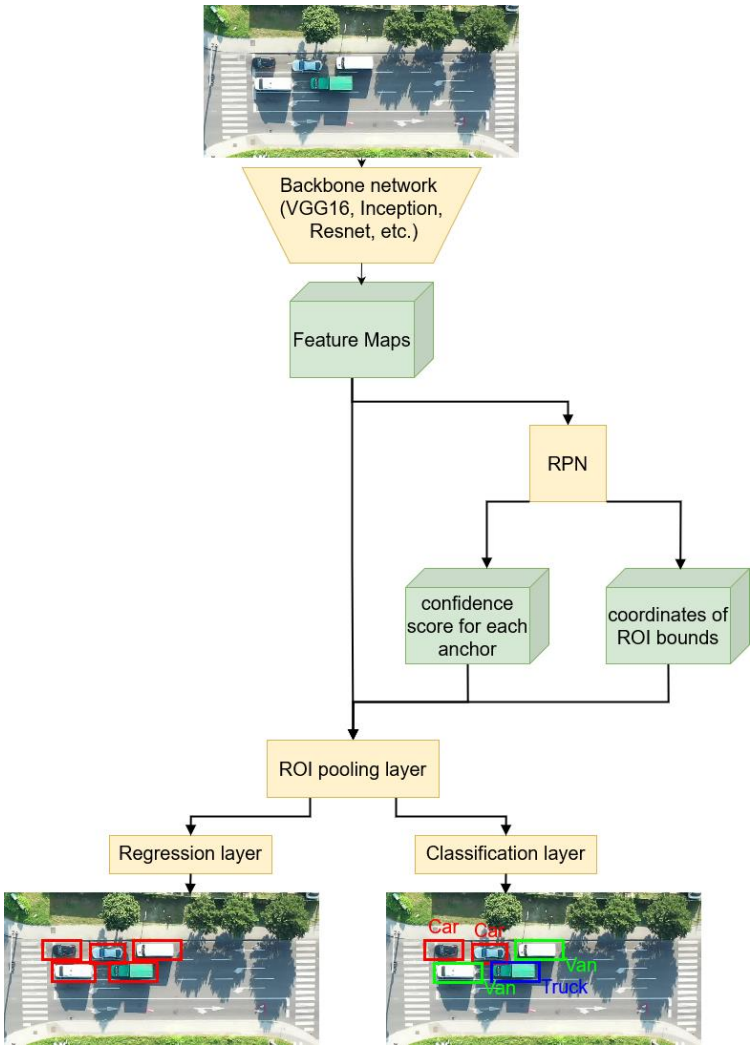


Figure 1.7 Architecture of Faster R-CNN. Yellow elements present CNNs, while green elements present output vectors of CNNs. Final regression layer provides bounding boxes of detected objects (red rectangles), while classification layer classifies detected object into one of pre-defined classes.

1.7.2. Yolo

Initially, Yolo was introduced by Joseph Redmon et al. in 2016 [36], marking a significant shift in object detection by performing the task in a single pass through the network. This was in contrast to previous methods that required multiple passes or used a two-stage process for object detection (R-CNN family of detectors). The architecture of Yolo v1 consisted of 24 convolutional layers followed by two fully concatenated layers for predicting bounding box coordinates and probabilities. A unique aspect of Yolo v1 was the simple output prediction based solely on regression, unlike Faster R-CNN which used a duo of outputs for classification and bounding box coordinates. However, Yolo v1 had some limitations, such as a higher localisation error compared to modern methods like Faster R-CNN, and it had difficulties in detecting objects with different aspect ratios.

Yolo v2 was introduced in 2017 by Joseph Redmon and Ali Farhadi [37], bringing along several improvements over its predecessor while maintaining the same real-time performance. Among the notable enhancements were the inclusion of batch normalization on all convolutional layers, high-resolution classifier fine-tuning, a fully convolutional architecture, and the introduction of anchor boxes for bounding box predictions. The backbone architecture for Yolo v2 was termed Darknet-19, which included 19 convolutional layers and five max-pooling layers. This version utilized 1x1 convolutions to decrease the number of parameters and implemented batch normalization to regularize and help convergence.

In 2018, Yolo v3 was published by Joseph Redmon and Ali Farhadi [38], presenting significant changes and a larger architecture to match the state-of-the-art while retaining real-time performance. Yolo v3 introduced Darknet-53 as the backbone, featuring 53 convolutional layers with residual connections. Key changes included multi-scale predictions, new confidence score prediction using logistic regression for each bounding box, and Spatial Pyramid Pooling (SPP) block addition to the backbone. However, Yolo v3 still struggled with detecting smaller objects, which was a persistent issue from previous versions. In February 2020 Joseph Redmon retired from the position of Yolo developer due to military applications of Yolo detector as well as privacy concerns caused by Yolo applications.

In April 2020, Yolo v4 was released by Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao [39]. This version maintained the Yolo philosophy but introduced several new features and enhancements. Yolo v4 experimented with multiple backbone architectures, ultimately adopting a modified Darknet-53 with Cross-Stage Partial connections (CSPNet) as the backbone.

A couple of months later, Yolo v5 was released by Glen Jocher, the founder and CEO of Ultralytics [53]. Yolo v5 adopted many of Yolo v4 improvements but was developed in PyTorch instead of Darknet. It introduced an algorithm called AutoAnchor for pre-training, which adjusts anchor boxes for better fitting to the dataset and training settings. The backbone of Yolo v5 is a modified CSPDarknet53, featuring a SPP Fast layer to speed up computation by pooling features of different scales into a fixed-size feature map. Yolo v5 also provides five scaled versions to cater to different hardware requirements and application specifics, making it a versatile choice for a variety of object detection tasks.

1.8. Thesis objectives

The objective of this thesis is to analyse possibilities of automated collection of road and roadside safety attributes by combination of various geospatial data sources with object detection methods.

The research hypotheses are as follows:

1. Object detection method on UAV aerial video can be utilized for calculation of traffic flow parameters.
2. Object detection method on mobile Lidar data can be used for roadside feature detection and distance measurements.
3. Combination of high-resolution satellite imagery with object detection method can be utilized for automated detection of road attributes to support road infrastructure safety assessment.

The research hypotheses resulted in the following papers:

- **Paper A** – An Analytical Framework for Accurate Traffic Flow Parameter Calculation from UAV Aerial Videos.
- **Paper B** – Automatic Roadside Feature Detection Based on Lidar Road Cross Section Images.
- **Paper C** – Utilizing High Resolution Satellite Imagery for Automated Road Infrastructure Safety Assessments.

The first research hypothesis is addressed in Paper A, which proposed a new low-cost framework for determining high-precision traffic flow parameters. A core part of the proposed framework is the combination of the state-of-the-art pre-trained Faster R-CNN for vehicle detection with UAV videos, which form the basis for calculating macroscopic and microscopic

traffic flow parameters. The traffic flow rate as a macroscopic parameter is also a road attribute defined by iRAP.

The second research hypothesis is addressed in Paper B, which proposed a new framework for determining road infrastructure attributes. The framework consists of a combination of mobile Lidar sensors to collect road infrastructure data as a point cloud and the Yolo v5 object detector to determine the final attributes. The framework focuses on the complete and automatic solution for two iRAP attributes: roadside severity-object (RSS – O) – related to the detection of roadside objects and roadside severity-distance (RSS – D) – related to the distance of the detected object from the roadside.

The third research hypothesis is addressed in Paper C, which proposes the combination of very high resolution (VHR) satellite imagery with the Yolo v5 object detector for automatic road attributes detection. The proposed approach focuses on the detection of four different types of pedestrian crossings, school zones and divided carriageways. All detected attributes are also part of the iRAP-defined road attributes.

1.9. Expected scientific contributions

The main scientific contribution of this research is automatization of road safety infrastructure attributes collection. This research has three additional scientific contributions:

- This research will lead to an automated method for calculating traffic flow parameters by applying the object detection method to UAV aerial video.
- Based on the theoretical knowledge and analysis of the research results, application of the object detection method to mobile Lidar data will be proposed for the roadside features detection.
- Based on theoretical knowledge and analysis of research results, application of object detection method on high resolution satellite imagery will be proposed for automated detection of road attributes for road infrastructure safety assessment.

1.10. Chapter summary

The doctoral dissertation is structured in six chapters. The first chapter Introduction describes the background of the research and details about road safety statistics, legislative and programmes as well as current approaches for collecting road attributes. Also, in detail explanation of CNNs and object detectors such as Faster R-CNN and Yolo family is given.

Finally, motivation, hypotheses and expected scientific contributions of the research are described.

Chapter 2 presents a low-cost traffic flow parameters estimation framework based on integration of Faster R-CNN detector with UAV video. The chapter provides the process of determination for three macroscopic flow parameters and four microscopic flow parameters, including traffic flow rate as one of iRAP-defined road attributes.

Chapter 3 presents a roadside feature detection framework based on integration of Yolo v5 detector with point cloud collected by mobile Lidar. The chapter provides the process of determination for RSS – O and RSS – D attributes which are part of the iRAP-defined road attributes.

Chapter 4 provides details about integrating Yolo v5 detector with VHR satellite imagery to detect four types of pedestrian crossings and school zones as well as making distinction between divided and undivided carriageways.

Chapter 5 provides a joint discussion on each of the published papers. Each paper contribution to research scientific contribution is explained, as well as the joint contribution of all three scientific articles. The discussion provides information about results obtained in published papers and their connection to previous research.

Chapter 6 presents the conclusions and scientific contributions of this dissertation.

Chapter 2
An Analytical Framework for Accurate Traffic Flow Parameter Calculation from UAV Aerial Videos

This chapter has been published as: Brkić, I.; Miler, M.; Ševrović, M.; Medak, D. An Analytical Framework for Accurate Traffic Flow Parameter Calculation from UAV Aerial Videos. Remote Sens. *2020,* 12, *3844. https://doi.org/10.3390/rs12223844.*

Conceptualization, I.B. and M.M.; investigation, I.B.; methodology, I.B., M.Š., and M.M.; supervision, M.M., M.Š., and D.M.; validation, I.B. and M.M.; visualization, I.B.; writing – original draft, I.B.; writing – review and editing, M.M., M.Š., and D.M.; funding acquisition, D.M.

Abstract:

Unmanned Aerial Vehicles (UAVs) represent easy, affordable, and simple solutions for many tasks, including the collection of traffic data. The main aim of this study is to propose a new, low-cost framework for the determination of highly accurate traffic flow parameters. The proposed framework consists of four segments: terrain survey, image processing, vehicle detection, and collection of traffic flow parameters. The testing phase of the framework was done on the Zagreb bypass motorway. A significant part of this study is the integration of the state-of-the-art pre-trained Faster Region-based Convolutional Neural Network (Faster R-CNN) for vehicle detection. Moreover, the study includes detailed explanations about vehicle speed estimation based on the calculation of the Mean Absolute Percentage Error (MAPE). Faster R-CNN was pre-trained on Common Objects in COntext (COCO) images dataset, fine-tuned on 160 images, and tested on 40 images. A dual-frequency Global Navigation Satellite System (GNSS) receiver was used for the determination of spatial resolution. This approach to data collection enables extraction of trajectories for an individual vehicle, which consequently provides a method for microscopic traffic flow parameters in detail analysis. As an example, the trajectories of two vehicles were extracted and the comparison of the driver's behaviour was given by speed – time, speed – space, and space – time diagrams.

2.1. Introduction

The number of vehicles in Europe is increasing every year. According to the European Automobile Manufacturers Association (ACEA), in 2019, there were 531 passenger cars per 1000 inhabitants in the European Union (EU). Comparing this number with 497 cars in 2014, it gives a 7% increase over five years [54]. This amount of growth in the number of vehicles requires new solutions in the transport infrastructure. Numerical values describing each road are indispensable and essential for performing a proper analysis. This is the main task of traffic engineers, who usually describe the flow using deterministic modelling of traffic flow parameters, depending on their requirements. These parameters can be divided into two groups: macroscopic and microscopic. According to Rao, the macroscopic parameters characterize the traffic as a whole and microscopic parameters study the behaviour of an individual vehicle in the flow relating to one another [55]. Parameters, such as the traffic flow rate, speed, and density are three fundamental macroscopic traffic flow parameters used to describe the state of continuous traffic flow [55]. Microscopic parameters are gross and net time headways (time headways and time gaps), and gross and net distance headways (distance headways and distance gaps) [56]. Collecting accurate and detailed traffic flow data and obtaining parameters at specific locations can be a very expensive, demanding, and time-consuming process. It often involves significant person-hours and expensive technologies that have limitations in collecting all the necessary data. These include traffic data acquisition techniques, such as pneumatic road tubes, induction loops, video image detection, piezoelectric sensors, and smartphone applications with their advantages and disadvantages [57–59]. However, contrary to the above techniques, Unmanned Aerial Vehicles (UAVs) represent more efficient and simpler solutions for many tasks, including the collection of traffic data [60]. The Single European Sky Air Traffic Management (ATM) Research Joint Undertaking (SESAR JU), established by the European Union Council provided a projection that more than seven million consumer leisure UAVs will operate across Europe by 2050 [61]. According to SESAR JU, one of the reasons for the large growth in the number of commercial and government UAVs is their capability to collect data from aerial points that had been inaccessible in the past. Besides, the European Union is currently investing 40 million euros through the SESAR JU project to integrate UAVs into the controlled airspace, indicating serious EU plans for UAV usage.

The main aim of this study is to propose a new framework for low-cost, location-specific traffic flow parameter measurements for the calibration of the deterministic traffic flow models. To provide a basis for deciding on new solutions, a detailed analysis of the current situation was

required. For this reason, the proposed framework had the purpose of consolidating all of the procedures needed to collect highly accurate traffic flow parameter values (flow, density, speed, headway, gap, etc.); from terrain operations, through image processing, vehicle detection and tracking to traffic flow parameters estimation. Moreover, the proposed framework included spatial analyses significant for obtaining microscopic traffic flow parameters. Since decisions about the reconstruction of the existing and the construction of new roads are time-consuming processes, it is not necessary to derive traffic flow parameters in real time. In this paper, the emphasis is on achieving high accuracy in parameter determination. For this reason, a robust state-of-the-art object detection method was integrated into the framework. It is a significant part of the study, and the results of using the stated method are supported by the evaluation metrics. Except for standard evaluation metrics of object detection methods, the displacement of the detected and ground truth vehicles is also provided in this paper.

This paper is organized as follows: after a brief introduction to the research topic and a brief overview of the related papers, the sections on data collection and methods explain in detail the parts of the proposed framework. For ease of reading and understanding, the framework is divided into four segments: terrain survey, image processing, object detection, and parameters estimation. This is followed by a section in which the results of the performed analyses are presented in the form of tables and diagrams. Moreover, the results section gives an insight into the evaluation of metric values for the proposed framework. This is followed by a proper discussion of the given results and the advantages and disadvantages of the proposed framework given the currently developed methods presented in the related papers. Finally, according to the results and outcome of the discussion, a brief conclusion is provided with future research options for traffic data collection.

2.1.1. Related Works

For the reasons mentioned above, it is evident that more and more research projects are currently relying on the use of UAVs. Many scientific articles in traffic science refer to the use of UAVs for real-time traffic monitoring and management [60,62–65]. Unlike real-time traffic monitoring and management, the collection of traffic data from UAVs is the subject of research in several papers. Khan et al., used UAVs to estimate the traffic flow parameters and to automatically identify the flow state and the shockwaves at the signalled intersections [66]. To achieve the above goals, a UAV was hovered above the road intersection on an eccentric position regarding the intersection center. For vehicle detection, Khan et al., used computer vision techniques such as optical flow, background subtraction, and blob analysis. In terms of

traffic flow parameters, all macroscopic parameters (traffic flow rate, speed, density) and one microscopic parameter (net distance headways) were derived. On the evaluation metrics side, the authors provided the evaluation metrics of speed estimation for a single vehicle. In another study, Ke et al., created a UAV-based framework for real-time traffic flow parameter estimation [67]. Their framework consisted of vehicle detection and parameter estimation. For vehicle detection, they used the Haar cascade and a Convolutional Neural Network (CNN) as an ensemble classifier and optical flow. In terms of traffic flow parameters, they also derived macroscopic parameters (traffic flow rate, speed, and density), while the microscopic parameters estimation was not performed within this study. The training process was performed on 18,000 images, while testing was performed on 2000 images. Similar study was made by Chen et al. where the authors proposed a four-step framework with the aim to extract vehicle trajectories from a UAV-based video. Vehicle detection segment was performed by ensemble Canny-based edge detector, while the tracking process is based on Kernelized Correlation Filter (KCF) algorithm. Particular emphasis is on the conversion image Cartesian coordinates to Frenet coordinates and on smoothing vehicle trajectories constructed from Frenet coordinates [68].

Apart from the aforementioned papers aimed at collecting traffic flow data from UAVs with a precise vehicle detection method, few works have been focused explicitly on detecting vehicles or collecting traffic data from a fixed camera. Fedorov et al. applied a fine-tuned object detection network to estimate the traffic flow from a surveillance camera [69]. They used a fine-tuned Mask Region-based Convolution Neural Network (Mask R-CNN) for vehicle detection in six classes: car, van, truck, tram, bus, and trolleybus. They also provided a comparison of manually collected traffic flow rate and traffic flow rate based on their object detection network for the observed road intersection. A surveillance camera was placed high above the road intersection. In terms of traffic flow parameters, only the macroscopic parameters were presented. The fine-tuning process was performed by 786 images, while 196 images were used for evaluation. On the other hand, Wang et al., focused on detecting vehicles from UAV [70]. They developed a vehicle detecting and tracking system based on the optical flow. They also explained in detail four segments of the system: image registration, image features extraction, vehicle shape detection, and vehicle tracking. As previously listed studies, the computation of microscopic parameters was not performed in this study.

2.2. Data Collections and Methods

The motorization rate in the City of Zagreb in 2011 was 408 passenger cars per 1000 inhabitants (Ministry of the Sea; Transport and Infrastructure, 2013). The Zagreb bypass is the busiest motorway in Croatia with the traffic rate continuously rising (Ministry of the Sea; Transport and Infrastructure, 2013). The location of this research is an approximately 500 m long section of Zagreb bypass motorway (Figure 2.1). It is a part of the A3 national highway, and it extends in the northwest – southeast direction. The specified section of the road consists of two lanes with entries and exits in both directions. Figure 2.1 shows the study area on the Open Street Map (OSM) and the Croatian digital orthophoto with Ground Control Points (GCPs) marked on the image from UAV. The research area was observed between 4:00 p.m. and 4:15 p.m. assuming this is the time of increased traffic density because of the migration of workers after a working day, which is usually between 8:00 a.m. and 4:00 p.m. in Croatia. Moreover, according to the Highway Capacity Manual, a period of 15 min is considered to be a representative period for traffic analysis during the peak hour [72]. Increased traffic density presents ideal conditions for testing the proposed framework, with particular emphasis on vehicle detection and describing the traffic flow by suitable parameters.

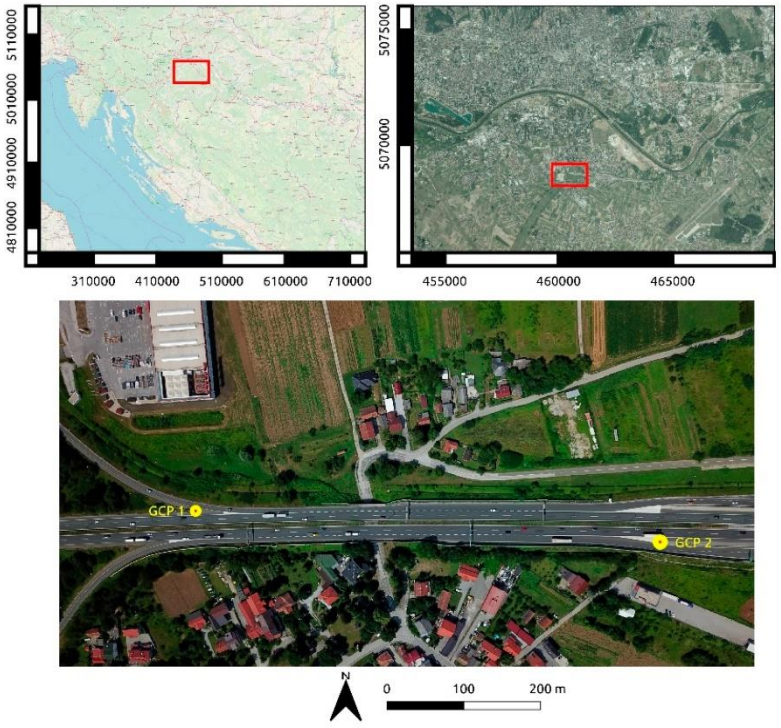


Figure 2.1 Location of the observed area on the Open Street Map and Croatian digital orthophoto from 2018 along with the image recorded from Unmanned Aerial Vehicles (UAV) whence Ground Control Points (GCPs) were marked.

The proposed framework for collecting traffic data and obtaining traffic flow parameters consists of four segments: terrain survey, image processing, vehicle detection, and determination of traffic flow parameters. All of the above segments contain sub-segments, which will be explained in detail below (Figure 2.2).

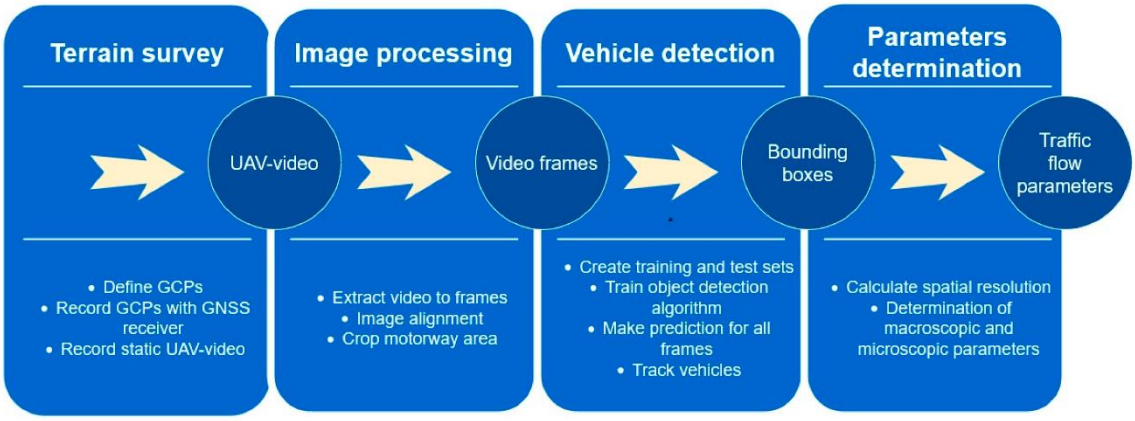


Figure 2.2 Proposed framework for the determination of traffic flow parameters.

2.2.1. Terrain Survey

The terrain survey segment includes defining two Ground Control Points (GCPs), determination of their positions, and recording a UAV-video of the observed area. GCPs are defined with unique terrain points: GCP1 on the edge of a road fence, while GCP2 was defined as the intersection of two white lanes (Figure 2.1). Considering that in this paper GCPs are used to determine the spatial resolution with high accuracy, which is a two-dimensional problem, the usage of the two GCPs is sufficient for this task. The positions of GCPs were recorded in the Croatian reference coordinate system with dual-frequency Global Navigation Satellite System (GNSS) Topcon HiPer Site Reciver (SR) obtaining GNSS corrections from the state network of referential GNSS stations – Croatian Positioning System (CROPOS). This is a very important part of the study for calculating the spatial resolution of images, which is a key detail for the determination of the traffic flow parameters. Considering that parameters such as density, speed, and distance headways are based on spatial distances, the accuracy of the mentioned parameters is directly connected with spatial resolution. After that, the observation area was recorded for approximately 14 min with a UAV. The UAV was nearly static and stabilized by an onboard GNSS. Moreover, the UAV was positioned vertically to enable a more accurate estimation of car positions. The flight height was near 50 m. The UAV was exposed to a light wind and the recorded video was not entirely stable.

2.2.2. Image Processing

The image-processing segment includes extracting frames from the UAV-video, image alignment, and cropping the motorway area. OpenCV library in Python programming language was used to make this segment. There was a 13:52 min long video, which was extracted to 19,972 frames that match the UAV frame rate of 24 fps (24 Hz). Since the video was not perfectly stable, image (frame) alignment had to be applied. Image alignment consists of applying feature descriptors to images and finding homography between images. Awad and Hassaballah described most of the existing feature detectors concisely in their book [73]. According to Pieropan et al. Oriented Features from accelerated segment test (FAST) and Rotated (ORB) and Binary Robust Invariant Scalable Keypoints (BRISK) have the best performance with motion blur videos; therefore, ORB was used in this study [74]. As its name suggests, ORB is a very fast binary descriptor which relies on Binary Robust Independent Elementary Features (BRIEF), where BRIEF is rotation invariant and resistant to noise [75]. Apart from the feature descriptors, homography is also the main part of the image alignment. Homography is a transformation that maps the points from one image to the corresponding points in the other image [76]. Moreover, feature descriptors usually do not perform perfectly, so to calculate homography, a robust estimation technique must be used [77]. For this purpose, the OpenCV algorithm using the Random Sample Consensus (RANSAC) technique was used, which Fischler and Bolles explained in detail [78]. Unlike traditional sampling techniques, which are based on a large set of data points, RANSAC is a resampling technique, which generates candidate solutions by using the minimum number observations (data points) required to estimate the underlying model parameters. Ma et al., presented more about homography [79]. In this study all frames were aligned with the first frame, which is better known as a master – slave technique, where the first frame is characterized as the master image, and all of the other frames as slaves [80].

Finally, the last step in the image-processing segment was cropping images around the motorway. The original video was recorded in 4096 × 2160 pixels dimension, which required a lot of memory while being processed. This amount of memory utilization can slow down the processing of the object detection. Depending on the hardware resources, with such high memory requirements, the algorithm may not even be able to complete the calculations. Therefore, frames must be cropped to the observation area only, i.e., a narrow area along the motorway. After cropping, the individual frame dimensions were 3797 × 400 pixels (Figure 2.3). Finally, the images were ready for the vehicle detection segment.

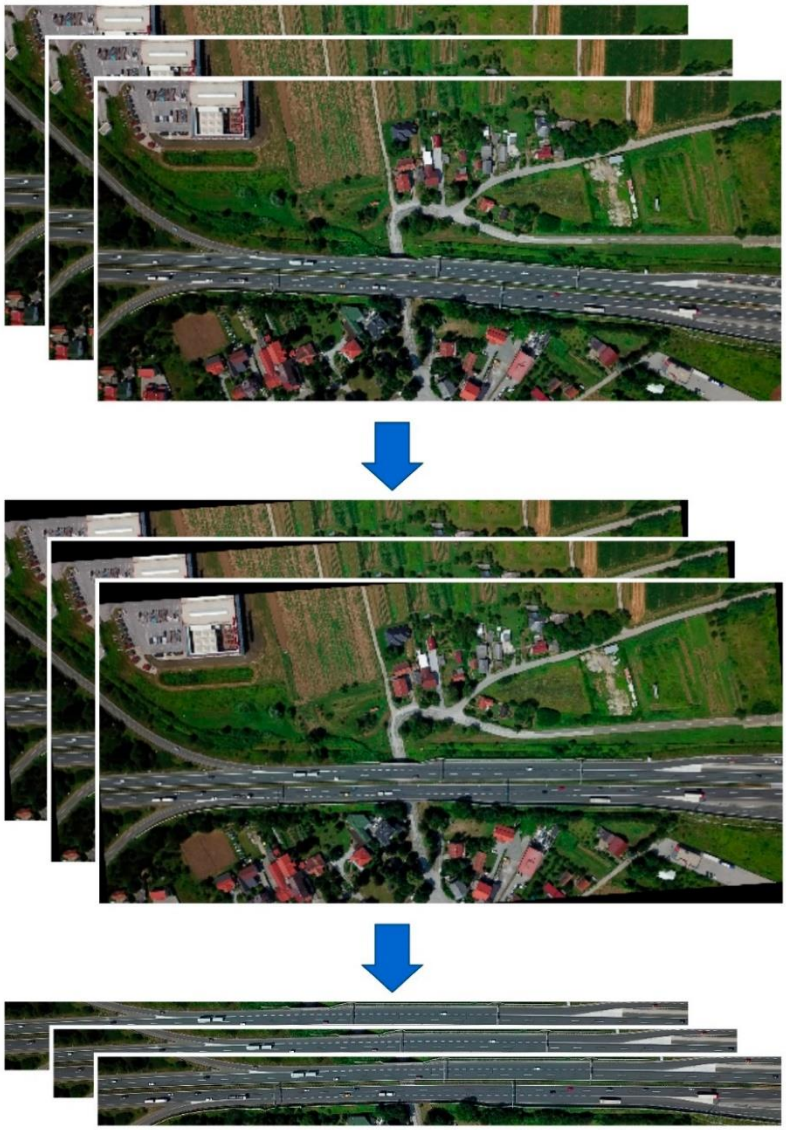


Figure 2.3 Image processing segment: firstly, frames were extracted from the video, then image alignment was applied, and finally, a narrow motorway area was cropped.

2.2.3. Vehicle Detection

This is computationally the most resource-intensive segment of the study. An Intel Core 2 Quad Central Processing Unit (CPU) with 256 GB RAM and 2 NVIDIA GP104GL Quadro P4000, 8 GB GDDR5 memory Graphical Processing Units (GPUs) was used. The deep learning part of this segment was done with a TensorFlow object detection application programming interface (API).

In this study, deep learning object detection was applied for vehicle detection. There are many object detection methods based on deep learning [81]. All of them can be divided into two-stage and one-stage detectors [82]. Two-stage detectors firstly obtain a significant number of

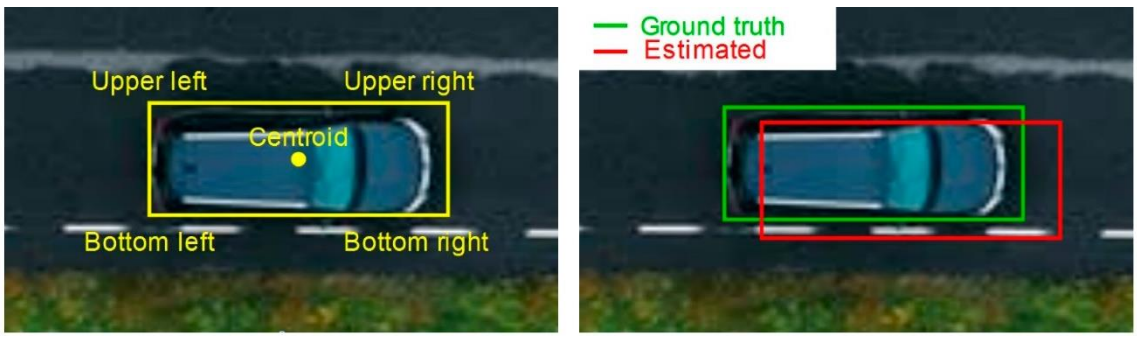
regions of interest (ROI), usually with Region Proposal Network (RPN) or Features Pyramid Network (FPN), and then use CNN to evaluate every ROI [83]. Unlike two-stage detectors, one-stage detectors suggest ROI directly from the image, without using any region proposal technique, which is time-efficient and can be used for real-time applications [82]. Because of their complexity two-stage detectors have an advantage in accuracy [84]. Since this study is not based on real-time tracking and vehicles are very small on the images, a two-stage detector was used. According to the results of the research made by Liu et al., and the availability of the mentioned hardware, Faster R-CNN with ResNet50 backbone network, pre-trained on the COCO images dataset, was selected as an optimal network for this study [85,86]. Faster R-CNN detector is described in detail by Ren et al. [49]. Since Faster R-CNN was pre-trained, this process of training is called transfer learning. It enables training with a small dataset of images, and it is a short-time process. A very important part of the Faster R-CNN detector is the selection of anchor's dimensions, aspect ratios, scales, height, and width strides. All of these parameters are explained in detail by Wang et al. [87]. Targeted selection of these parameters can greatly improve the detection time and accuracy. The height and width of anchors were defined by parameters of ResNet50 CNN, which is explained in detail by He et al. [88]. After analyzing the labeled vehicle sizes and shapes, the parameters presented in Table 2.1 were used.

Table 2.1 Table of parameters and their values used in Faster R-CNN ResNet50 COCO network.

Name of Parameters	Value
Height of anchors	16
Width of anchors	16
Height stride	16
Width stride	16
Aspect ratios	2, 4, 6
Scales	3, 7, 11

Regarding the training and test set of images, 200 images, equally distributed all over the frames, were selected. Selected images were divided into training and test dataset at a ratio of 80:20, i.e., train dataset containing 160 images (4367 vehicles) and test dataset with 40 images (1086 vehicles). According to a very small set of data with unequal distribution of vehicle types, where there are mostly passenger cars with several trucks, buses, and motorcycles, all images were marked by only one class (vehicle). Labeling of vehicles in training and test images was performed by the LabelImg software.

After labeling the vehicles, training started by using the hardware as mentioned above. The training time was 2 h and 35 min and it ended in 22,400 training steps, i.e., 140 training epochs with a batch size of 1. Minimal batch size was selected because of computation resources limitation. Afterward, the trained model was evaluated on a test set of images. From 40 test images, all of 1086 vehicles were manually labeled. For the evaluation process, the confusion matrix was used. The confusion matrix consists of two columns, which represent the numbers of true and false actual vehicles, and two rows, which represent numbers of positive and negative predicted vehicles. Other evaluation metric values, such as precision, recall, accuracy, and F1 score were derived from the confusion matrix. The process of derivation is described in detail by Mohammad and Md Nasair [89]. Finally, a trained model was applied for the prediction of vehicles on all of the frames. The prediction process resulted in frame ID, vehicle ID, confidence score, and coordinates of bounding boxes and centroids of every single detected vehicle. In this study, the bounding boxes analysis was applied to choose the best characterizing point of bounding box for tracking and determining the macroscopic traffic flow parameters. The upper left, upper right, bottom right, bottom left, and centroid points have been considered for this purpose (Figure 2.4a). The specified characteristic points of the estimated bounding boxes were compared with the same points of ground truth bounding boxes in the test dataset Figure 2.4b).



a) b)

Figure 2.4 (a) Characteristic points of bounding boxes; ***(b)*** *example of difference between ground truth and estimated bounding boxes.*

Displacements of specified points were calculated as Root Mean Square Error (RMSE), which is defined by the U.S. Federal Geographical Data Committee as a positional accuracy metric [90]. RMSE values were calculated for all the considered characteristic points with the following equation:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\overline{x_i} - x_i)^2 + (\overline{y_i} - y_i)^2}{n}} \qquad 2.1$$

where n is the number of bounding boxes in the test dataset, ($\overline{x_i}$, $\overline{y_i}$) are characteristic point coordinates of ground truth bounding boxes, while (x_i, y_i) are characteristic point coordinates of estimated bounding boxes. Based on RMSE values, the centroid point was selected for tracking and determination of macroscopic parameters.

In terms of microscopic parameters (which will be explained in 2.2.4.1), it is not sufficient to represent a vehicle with one point, not even by a centroid, but it is necessary to use a whole bounding box. The dimensions of the bounding boxes and their location on the image have great impact on the microscopic parameters. To evaluate the detection process of the bounding boxes, the Intersection over Union (IoU) metric was used. IoU is defined with the following equation:

$$IoU = \frac{E \cap T}{E \cup T} \qquad 2.2$$

where E is the estimated bounding box, while T is the ground truth bounding box. In this paper, *IoU* was calculated for every single vehicle and the detection process is evaluated by the mean value of *IoU*. Mean *IoU* with already described *RMSE* values of the characteristic point give adequate metric for microscopic parameters reliability estimation.

Finally, it is necessary to connect the same vehicles between frames by the vehicle ID number. This is done with Simple Online and Realtime Tracking (SORT) algorithm. The algorithm is based on the Kalman filter framework for determining the vehicle speed and IoU parameter between vehicles in two consecutive frames. Bewley et al. provided more details about the SORT algorithm [91]. By applying the SORT algorithm for tracking vehicles, the object detection segment is finished, and the output of this segment is also the input for the last one, i.e., the segment for parameter determination.

2.2.4. Parameters Determination

The last part of this study is closely related to traffic flow parameters measurement and calculation. As already stated, one of the aims of this study is to collect the traffic flow data and measure and calculate the traffic flow parameters. Considering this, the traffic flow parameters were measured at the beginning and at the end of the observed road section for individual through lanes on the motorway and on the entry and exit slip lanes. The location-based

parameters such as traffic flow rate, time mean speed and time headways and gaps were determined at the characteristic locations of the observed area. That makes a total of 12 locations, marked with numbers from 1 to 8, which is shown in Figure 2.5. For locations 2, 3, 6 and 7 there are separate lanes a and b, where lanes represent a slower track and b lanes a faster track. Contrary to location-based parameters, segment-based parameters such as space mean speed, traffic flow density, and distance headways and gaps were determined for each lane segment, which is shown in Figure 2.6. Each lane segment is approximately 386 m long and marked with numbers from 1 to 6.

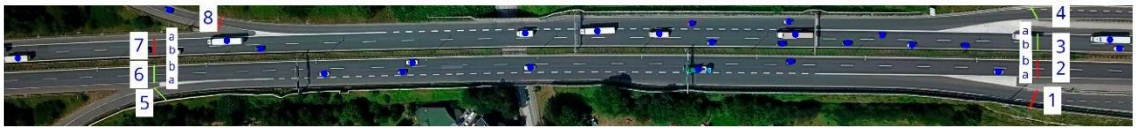


Figure 2.5 Marked characteristic locations of the observed area used for calculating location-based traffic flow parameters.

Figure 2.6 Marked lane segments of the observed area used for calculating segment-based traffic flow parameters.

Since the output of the vehicle detection segment are the coordinates of vehicles bounding boxes and centroids, it allows creating geometry objects such as points for centroids and polygons for bounding boxes. Furthermore, spatial objects allow for spatial analysis, which has been used to estimate the traffic flow parameters. Considering that every lane can be observed as a spatial object, this is the key step for estimating the parameters of each lane. The GeoPandas Python library was used to accomplish these goals. More about spatial operations and GeoPandas can be found in Jordahl [92].

2.2.4.1. Macroscopic Traffic Flow Parameters

The first of the macroscopic traffic flow parameters is the traffic flow rate. It is defined as the number of vehicles that cross the observed section of the motorway within the specified time interval [93]. The traffic flow rate is usually expressed by the number of vehicles per hour. Since the vehicles were considered as spatial objects, the traffic flow rate was measured with

simple spatial operations. This is accomplished by placing a line perpendicular to the track and counting each time the centroid of the vehicle crosses the line.

Observing the flow of traffic with a UAV allows determining the position and speed of each vehicle in each frame. Vehicle positions are defined by vehicle centroids. Based on the centroid coordinates and the given frame rate, the speed of each individual vehicle can be determined by the equation:

$$V_n = \frac{1}{N}\sum_{n=1}^{N} d_n \qquad 2.3$$

where V_n is the speed point of an individual vehicle expressed in pixels per frame, d_n represents the distance the vehicle has traveled between the two observed frames and N represents the number of consecutive frames between the two observed frames (frame interval).

Since vehicle centroids are not fixed to a single vehicle point but vary from frame to frame, what is presented by RMSE value, determining vehicle speeds for successive frames (N = 1) will result with noisy data. Contrary to consecutive frames, determining the vehicle speeds with large frame intervals will result in a too smooth speed curve, which will lose significant data. In order to define the optimum N, the positions of the one vehicle, sample 139, was manually labeled in each frame. Vehicle 139 was chosen because of its relatively constant speed during its travel through the observation area. The collected data were used as ground truth data in Mean Absolute Percentage Error (MAPE) calculation. The MAPE was calculated between the speed of the truth point on the ground of one example (vehicle 139) and the estimated speed of the same vehicle for each N in the range from 1 to 30. The following equation was used to calculate MAPE:

$$MAPE = \frac{1}{n}\sum_{i=1}^{N} \frac{\overline{V_i} - V_i}{\overline{V_i}} \times 100\% \qquad 2.4$$

where $\overline{V_I}$ represents the ground truth speed in i-th frame; n represents the number of frames, while V_I represents the estimated speed in the same frame of vehicle 139.

Based on the calculated MAPE values, N is set to 12, representing a time interval of 0.5 s (24 frames per second / 12 frames = 0.5 s). A particular N was used to estimate the speed of all vehicles in the video. The process of estimating the traffic flow speed can be calculated in two ways: speed at the point of the road (Time Mean Speed (TMS)) or at the moment (Space Mean Speed (SMS)) [55]. The TMS is the average speed of all vehicles crossing the observation spot in a predefined time interval [55]. Contrary to TMS, SMS is defined by spatial weighting given

instead of temporal [55]. According to [94], the TMS is connected to a single point in the observed motorway area, while the SMS is connected to a specific motorway segment length. According to [95], SMS is always more reliable than TMS. More about TMS and SMS can be found in [55] and [94]. In this paper, TMS is calculated for 12 characteristic locations of the observed motorway area, while SMS is calculated for each segment of the motorway lanes. Considering that vehicle speeds and their positions are available for each video, SMS was calculated using the following equation:

$$SMS = \frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} V_{ij}}{mn} \tag{2.5}$$

where m is the number of vehicles, n is the number of frames and V_{ij} is the speed of an individual vehicle in a single frame.

The traffic flow density is defined as the number of vehicles on the road per unit distance. For computing traffic flow density, the spatial resolution of images must be determined. Based on the coordinates of the two GCPs and the number of pixels between the two GCPs, the spatial resolution can be calculated with three simple equations:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2.6}$$

$$n = \sqrt{(nx_2 - nx_1)^2 + (ny_2 - ny_1)^2} \tag{2.7}$$

$$r = \frac{d}{n} \tag{2.8}$$

where d is the spatial distance between GCPs expressed in meters; (x_1, y_1) are the spatial coordinates of GCP 1; x_2 and y_2 are the spatial coordinates of GCP 2; n is the image distance between GCPs expressed in pixels; (nx_1, ny_1) are the coordinates of GCP 1 image; (nx_2, ny_2) are the coordinates of GCP 2 image and r is the spatial resolution of the images expressed in meters.

The traffic flow density is spatially related to the traffic section and temporally related to the current state. It is usually expressed by the number of vehicles per kilometer [96]. In this study, the traffic flow density is determined for every video frame, but given the amount of data, this paper only shows the average density for each segment of the lane.

2.2.4.2. Microscopic Traffic Flow Parameters

In contrast to macroscopic traffic flow parameters, microscopic parameters consider the interaction of individual vehicles. There are four microscopic parameters: distance headways and gaps, and time headways and gaps. Time headway is defined as the time interval from the front bumper of one vehicle to the front bumper of the next vehicle expressed in one-second unit, while distance headway is the distance between the same points of the vehicles. Opposite to the headways, time gap is defined as the time interval from the back bumper of one vehicle to the front bumper of the next vehicle, also expressed in one-second unit, while distance gap is the distance between the same points of the vehicles [97]. It is important to recognize that the headways are defined between the same points on two consecutive vehicles. Therefore, in this study, headways were calculated between centroids of the vehicle bounding boxes.

Given the coordinates of vehicle bounding boxes and video frame rates are known, and the frames have successive numeric IDs, it is easy to determine the time headways and gaps. Time headway between two vehicles is computed as the difference between the frame ID when the upper right point of the first vehicle crosses the reference line and the frame ID when the upper right point of the next vehicle also crosses the same line. The upper right characteristic point of the vehicle bounding box was selected based on the already described calculation of RMSE values. Then, the calculated number of frames is divided by the frame rate of the video. The time gap between two consecutive vehicles is calculated in a similar way as the time headway with one significant difference: instead of centroids, the characteristic points of the bounding box edges were used to calculate the number of frames. Considering the calculated RMSE values for all characteristic points of the bounding boxes, in this paper the upper left and upper right characteristic points were used to determine the time gap. Therefore, the accuracy of the time gap depends on the positional accuracy of the upper left and upper right points. From the specified definitions of time headways and gaps, the time gap cannot be larger than the time headway, and this can be a good control point when calculating time headways and gaps.

Contrary to time headways and gaps, which are location-based parameters, distance headways, and gaps are segment-based parameters. The distance headways and gaps are calculated for each single frame using spatial resolution. As with the calculation of time headways and gaps, the upper right characteristic points were used for distance headway calculation, while edges of the bounding boxes were used to calculate the distance gaps. Moreover, as in the case of time gaps, the characteristic upper left and upper right points were used, and the accuracy of the distance gaps depends on the positional accuracy of the used characteristic points. Figure 2.7

shows the difference between headways and gaps, and the reference line for measuring time headways and gaps. From Figure 2.7, it is clear that the difference between the distance headway and gap represents the length of the observed vehicle. Likewise, as with the time headways and gaps, the distance gap cannot be larger than the distance headway, so that it can be a good control point when calculating the distance headways and gaps.

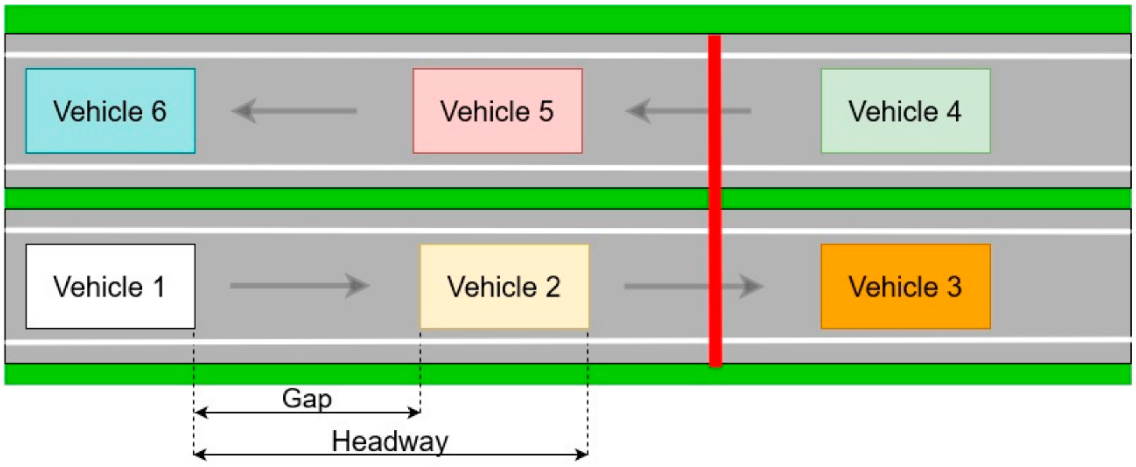


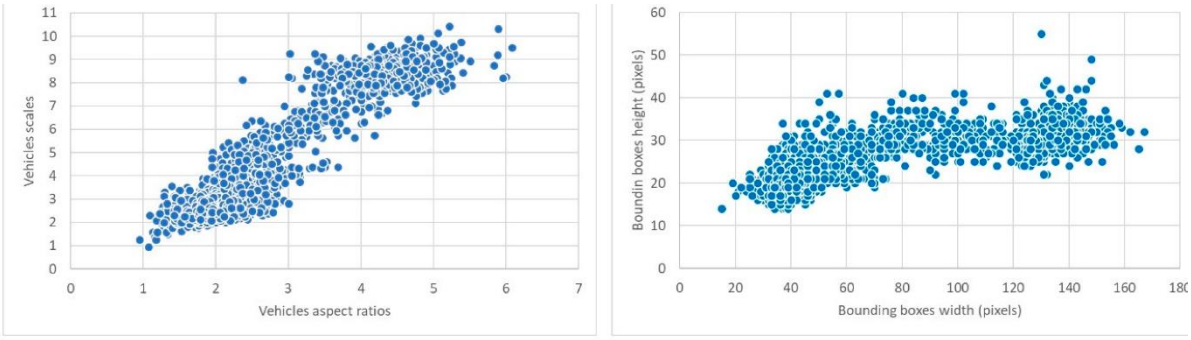
Figure 2.7 The difference between headways and gaps; red line marks the reference line for measuring location-based microscopic parameters such as time headways and gaps.

The above-described approach to estimate the macroscopic and microscopic traffic flow parameters allows the determination of the position and speed of each detected vehicle in the video at a frequency of 24 Hz. From certain data, it is possible to analyze the behavior of each individual vehicle. The combination of UAV-based video recording and object detection methods allows analyzing the path, speed, and travel time for each detected vehicle. It is usually performed by creating the diagrams with data about vehicle speed, traveled distance (space) and time of travel. The speed – time diagram represents the change in vehicle speed over time, while the space – time diagram represents the distance traveled in time. Opposite to speed – time and space – time, the speed – space diagram is derived from the data of the speed and traveled distances of observed vehicles.

2.3. Results

This study was conducted in the order given by the proposed framework, which is explained in the data collection and methods section. After the image processing part, object detection was performed. The parameters of the Faster R-CNN are defined based on the scale and aspect ratio of the vehicle. Figure 2.8a shows the distribution of vehicles according to their scales and aspect

ratios if the anchor size is already defined as 16 × 16 pixels. The anchor strides are defined based on the dimensions of vehicle bounding boxes. The distribution of the vehicle bounding boxes can be seen in Figure 2.8b.



a) b)

Figure 2.8 (a) Distribution of vehicles in the training set based on their scales and aspect ratios if anchor size is set to 16 × 16 pixels; ***(b)*** *distribution of vehicle sizes based on height and width of their bounding boxes.*

After the vehicle scales, aspect ratios and strides of anchors were selected, the Faster R-CNN network was trained and tested. The evaluation was performed on 40 images, which contain 1076 vehicles. Table 2.2 shows a confusion matrix with the numbers of true and false actual objects and positive and negative predicted objects. Out of the 1076 ground truth vehicles appearing in the 40 test set images, 1070 were detected while six were missed. Furthermore, a fine-tuned network detected 13 false vehicles. All that provides the evaluation metrics such as precision, recall, accuracy and F1 score values shown in Table 2.3.

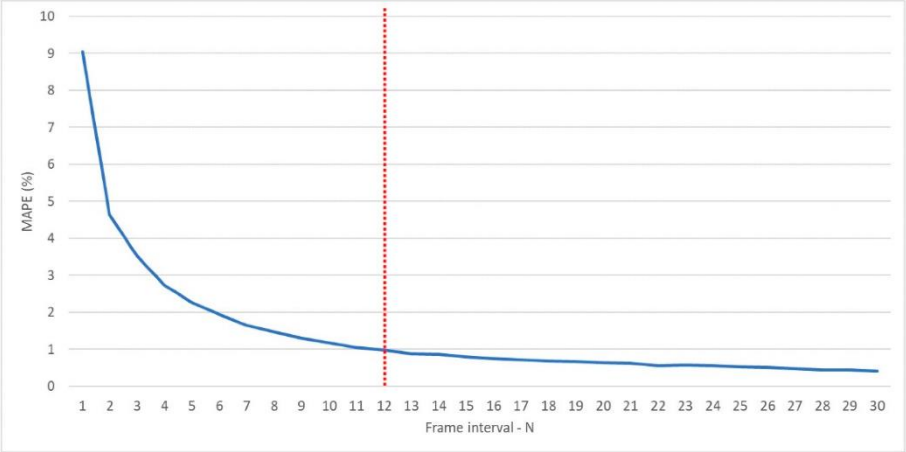
Table 2.2 Confusion matrix of test dataset.

		Actual	
		Vehicle	**Not vehicle**
Predicted	**Vehicle**	1070	13
	Not vehicle	6	0

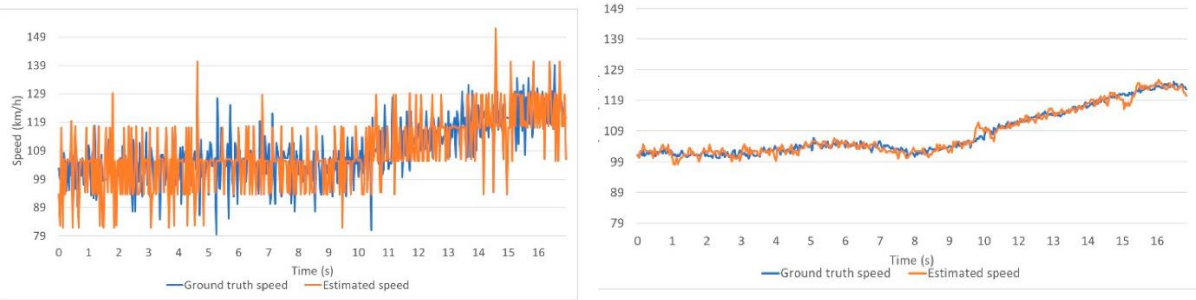
Table 2.3 Evaluating metrics of test dataset.

Evaluate Metric	**Value**
Precision	0.988
Recall	0.994
Accuracy	0.983
F1 score	0.991

To determine which characteristic point of the bounding box will be used for tracking, RMSE was calculated for every characteristic point of the vehicle bounding box. Centroids proved to be the best choice with the lowest RMSE value (0.28 m). In addition to the characteristic point, frame interval (N) was determined to estimate the vehicle speed. Figure 2.9a shows the MAPE change for N in a range from 1 to 30 frames. According to Figure 2.9a, N was set to 12. Figure 2.9b shows a comparison of the ground truth speed of vehicle 139 and the estimated speed for the same vehicle with N = 1, while Figure 2.9c shows the same comparison but with N = 12.



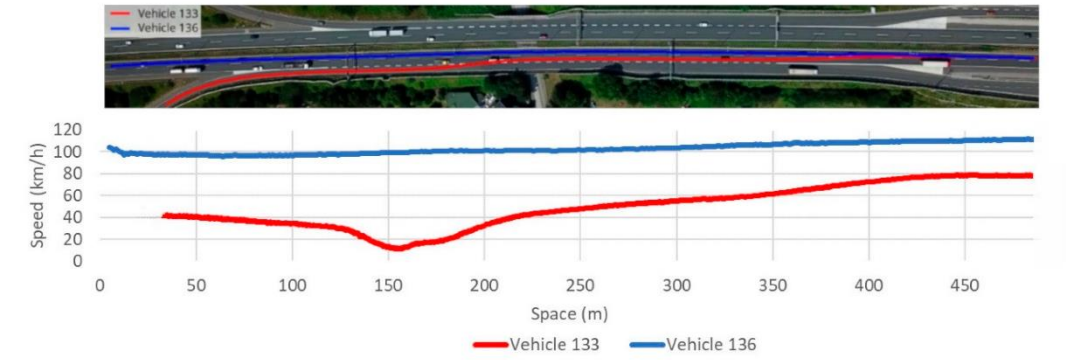
a)



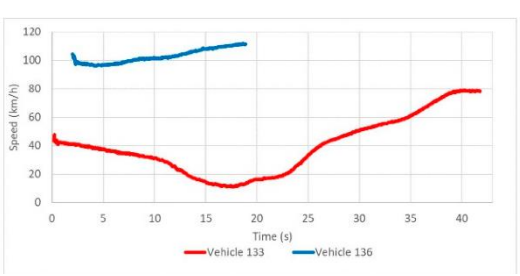
b) c)

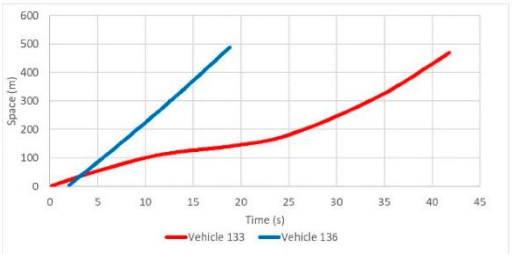
Figure 2.9 (a) MAPE – Frame interval diagram showing a decrease in MAPE as the number of frame intervals increases, the selected optimal frame interval is determined by the red line. (b) Time – speed diagram for ground truth and estimated data of vehicle 139 with N = 1; it is visible that speeds of estimated and ground truth bounding boxes have high rate of noise. (c) Time – speed diagram for ground truth and estimated data of vehicle 139 with N = 12; it is visible that the increase in frame interval causes a smoother speed curve.

Subsequently, to demonstrate the possibilities of this framework, the trajectories of vehicles 133 and 136 are presented in this paper. For both selected vehicles, speed – space, speed – time and space – time diagrams were created (Figure 2.10). From Figure 2.10b, it is evident that vehicle 133 traveled approximately 480 m, while vehicle 139 traveled approximately 500 m. Moreover, it can be seen that vehicle 136 appears in the video approximately two seconds after vehicle 133. It can be seen from Figure 2.10c that vehicle 133 passed vehicle 136 at approximately the third second of the video.



a)





b) c)

Figure 2.10 (a) Trajectories of vehicles 133 and 136 and their change of speed during travel. (b) Speed – time diagram of vehicles 133 and 136. (c) Space – time diagram of vehicles 133 and 136.

Afterwards, the spatial resolution was calculated based on equations 2.6 – 2.8. The spatial distance between GCP1 and GCP2 is 486.24 m, while the image distance between the same points is 3740.31 pixels, resulting in a spatial resolution of 0.13 m. The calculated resolution was used to determine the traffic flow density as well as to determine the microscopic parameters. The calculated RMSEs for the characteristic points of the bounding boxes are shown in Table 2.4. Location-based parameters were calculated for all individual characteristic locations (Table 2.5). It is important to note that the upper left and upper right points of the bounding boxes were used to estimate the net time headway instead of the lower left and lower right points for better RMSE. The total IoU metric parameter is 0.876. Contrary to location-based parameters, the segment-based parameters were calculated for every single lane segment (Table 2.6).

Table 2.4 Root Mean Square Error (RMSE) value of characteristic points.

Characteristic Point	RMSE Value (m)
Upper left	0.432
Upper right	0.360
Bottom left	0.405
Bottom right	0.397

Table 2.5 Estimated location-based traffic flow parameters.

Location ID	Traffic Flow Rate		TMS (km/h)	Time Headways and Gaps	
	Counted Number of Vehicles	Estimated Number of Vehicles (vehicles/h)		Time Headway (s)	Time Gap (s)
1	96	411	75.03	8.61	8.06
2a	194	831	75.76	4.37	3.7
2b	300	1286	98.14	2.87	2.5
3a	191	819	80.64	4.26	3.75
3b	243	1041	106.36	3.33	3.11
4	92	394	74.48	8.34	8.03
5	91	390	51.44	8.52	7.95
6a	206	883	81.91	3.97	3.51
6b	293	1256	101.67	2.79	2.58
7a	184	789	79.22	4.65	3.92
7b	247	1059	104.63	3.38	3.07
8	95	407	80.85	8.83	8.33

Table 2.6 Estimated segment-based traffic flow parameters.

Lane Segment ID	Average SMS (km/h)	Average Density (vehicles/km)	Distance Headways and Gaps	
			Distance Headway (m)	Distance Gap (m)
1	64.88	5	70.33	64.29
2	77.76	10	76.06	66.69
3	98.89	13	56.77	51.17
4	104.70	10	70.98	65.32
5	81.83	10	82.51	72.44
6	70.08	5	73.40	67.14

2.4. Discussion

This framework provides a new approach to traffic flow parameter estimation. The proposed framework integrates UAV, high-precision GNSS technology, state-of-the-art object detection, and spatial operations to provide highly accurate microscopic traffic flow parameters. Particular emphasis was placed on achieving high accuracy. This is done with the Faster R-CNN object detection network, which is pre-trained on the COCO dataset. Although only 160 images were used for the fine-tuning process, the evaluation process resulted in a precision of 0.988 and a recall of 0.994. Although this paper does not propose a new detection method and does not use the same dataset as datasets in the related papers, there are given comparisons with object detection metrics of the related papers. The reason for that is based on the fact that success of the novel frameworks largely depends on the accuracy of object detection. In comparison with similar studies, Ke et al. achieved slightly better precision (0.995) and significantly lower recall (0.957) [67]. Considering that Ke et al. used 18,000 images for the training process, it consequently yields a very lengthy process. The exact duration time of the training process is not given by Ke et al. [67]. Since recall represents the ratio of detected and ground truth vehicles, in terms of vehicle detection it is more important to achieve a higher recall. False vehicles are usually detected at the same location of the observation area, so in the fluid traffic flow this can be eliminated by post-processing. In addition, false detected vehicles do not appear in the consecutive frames, so they cannot be tracked and therefore the estimation of the traffic flow parameters cannot be affected. Contrary to false detected vehicles, undetected vehicles cannot be detected in post-processing and can cause major errors in estimating the traffic flow parameters. For this reason, L. Wang et al., give only the number of detected and “lost” vehicles instead of the confusion matrix, while the precision value is not provided [70]. From the number

of detected and "lost" vehicles it is clear that the recall of their vehicle detection is 0.979, which is less than the recall of vehicle detection used in this study. In addition to recall and precision, our study provides other evaluation metrics such as accuracy and F1 score that can be used in the future studies as benchmarks.

Since Faster R-CNN is a robust object detection network, it is advisable to analyse the ground truth objects and, according to the analysis, set the parameters for fine-tuning. In order to reduce the training time, it is important to specify the scales and aspect ratios of anchors with regards to the capacity of available hardware. For this reason, scales and aspect ratio are set to three values chosen to include all ground truth vehicles. Based on the distribution of vehicle sizes and aspect ratios, the scales were set to 3, 7, and 11, while the aspect ratios were set to 2, 4, and 6. Setting more values can even improve the evaluation metrics of vehicle detection, but this will yield an extension of the training time, with little benefit to precision.

To estimate the TMS and SMS, it is necessary to determine the frame interval. This type of analysis for the selection of appropriate frame interval for speed estimation has not yet been applied. For example, Ke et al., selected the frame interval of 5 in 25 Hz video, without a detailed explanation and mathematical analysis [67]. In this paper, the frame interval was selected based on the MAPE values of sample vehicle speed. As already established, a smaller interval will result in a higher MAPE rate, while larger intervals will result in a lower MAPE rate. The analysis of MAPE and frame intervals shows a decreasing MAPE rate while the frame interval is increasing. According to the MAPE – frame interval diagram, the frame interval is set to 12, which represents half of a second. Speed estimation, concerning the last half of a second, has a MAPE slightly below 1%, which represents a satisfactory rate. This study provides the difference between the speed values estimated for the frame interval from 1 to 12. The speed has higher amplitudes when the frame interval is 1, resulting in a higher MAPE rate. There are two reasons for high amplitudes. The first one is a spatial resolution for ground truth speed determination. The high spatial resolution allows the identification of characteristic details on the vehicle (rearview mirror, vehicle lights, etc.), and therefore allows the vehicle speed to be determined manually, with always tracking the same detail on the vehicle frame-by-frame. The described procedure for determining the speed of a ground truth will result in lower amplitudes of ground truth speed. Contrary to the high spatial resolution, the low resolution does not allow tracking vehicles by characteristic details. The characteristic detail points on the vehicle in each frame are determined by personal assessment. Another reason for the high amplitudes is the variable position of the vehicle centroid for each frame. The bounding

boxes of one vehicle do not have the same relative position concerning the vehicle in each frame. This may cause large amplitudes of estimated speed when the frame interval is 1.

This framework provides the determination of trajectories for every single vehicle. This study provides trajectories of vehicles 133 and 136. These vehicles were selected to show the differences in speed and trajectory, given that vehicle 136 had a relatively light flow, while vehicle 133 slowed down when entering the highway. Our results provide the speed – space, speed – time, and space – time diagrams for the same vehicles. Considering all these diagrams can be drawn from the results of the proposed framework for vehicles 133 and 136, it is possible to analyze the relations between any vehicle in the video with frequency of 24 Hz. These relations are important for some other types of traffic analysis, such as determining the critical gap and shockwaves, or for energy and environmental analysis. For shockwave analysis, Khan et al. proposed a framework with optical flow, background subtraction and blob analysis for detecting and tracking vehicles from a UAV-based video [66]. They do not provide any metrics on detection accuracy, so from this point of view these frameworks cannot be compared. However, they did provide a MAPE rate for one vehicle of 5.85%. Given that the MAPE rate of speed estimation for our framework is 0.92%, it is a significant difference. The reason for the lower MAPE rate may be that we analyzed the decrease in MAPE rate with increasing the frame interval, and according to the already described analysis, we selected 12 frames as the optimal frame interval. They also recorded a traffic flow with the UAV placed eccentrically concerning the observed area, while we placed the UAV approximately perpendicular to the observed area. Fedorov et al., used a surveillance camera for recording the road intersection and highlight the problem of overlapping vehicles caused by oblique recording [69]. Therefore, placing the UAV approximately perpendicular to the observed area has great influence on determining the spatial relations between vehicles, especially in the analysis of critical gaps and shockwaves.

Moreover, this framework enables the estimation of microscopic traffic flow parameters. Except for Khan et al., other related papers do not provide microscopic parameters [66]. The reason for this may be a highly accurate determination of the spatial resolution, which is significant for the measurement and estimation of microscopic parameters. Our proposed framework includes GNSS technology for determining a highly accurate spatial resolution. Since UAVs usually have a single-frequency GNSS, it is more accurate to use a dual-frequency GNSS receiver to determine the spatial resolution instead of spatial resolution calculation from the sensor dimensions, focal length, and UAV flight altitude.

Finally, the implementation of the proposed framework allows the determination of traffic flow parameters for each individual lane. The boundary boxes of detected vehicles and motorway lanes can be considered as spatial objects, i.e., polygons. This allows spatial analysis to be applied to them. The use of spatial analysis enables the automatic estimation of microscopic parameters of the traffic flow. A review of the related literature shows that this approach to determining microscopic parameters has not yet been applied [66,67,69,70].

In addition to all these advantages, the proposed framework has several limitations. First, the flight time of the UAV is limited, which does not allow the observation of the area of interest for more than a short period of time, depending on the capacity of the battery. Short flight times can be a significant problem in estimating reliable traffic flow parameters. This problem can be partially solved by observing the road over several periods of time during the day and it is highly dependent on the traffic volume at the given locations. Shorter observations provide a sufficient number of detections at high volumes while on low-volume sections multiple flights are required. This can be performed if multiple batteries are available, but the complete solution will only be achieved with an increase in the UAV battery capacity. Second, the framework is not fully automated. The vehicle detection segment of the framework requires manual labeling of the vehicle to fine-tune the Faster R-CNN networks. Although a small set of images is required for the fine-tuning process, it is short-lived, and it prevents frame automation. Third, UAVs represent a significantly more affordable solution for determining the flow parameters than the currently used technologies such as inductive loops, pneumatic road pipes, etc.

2.5. Conclusion

This paper provides a new framework for determining the traffic flow parameters. The proposed framework represents a more affordable and more efficient approach for typical standard traffic flow parameters determination than the techniques in current use. Furthermore, this method provides a simple and accurate method for plotting vehicle trajectories and continuous headway measurements at road sections not available with traditional traffic flow survey methods. The proposed framework can be segmented into a terrain survey, image processing, vehicle detection, and parameter estimation. Particular emphasis is placed on achieving high accuracy. To achieve high accuracy, the Faster R-CNN network was used. It is a robust state-of-the-art network, which was pre-trained with COCO image dataset and fine-tuned with only 160 training images. Using a fine-tuned Faster R-CNN network for vehicle detection achieved a recall of 0.994, which is significantly higher than the detection recalls in the related papers.

Moreover, the proposed framework provides a vehicle speed estimation with a MAPE of 0.92%, which is satisfactory and allows the estimation of the trajectories for each individual vehicle in the video.

Future studies will focus on addressing these shortcomings and improving the proposed framework. Particular emphasis should be placed on creating large datasets containing labeled vehicles in images from different videos and different contexts.

Chapter 3 Automatic Roadside Feature Detection Based on Lidar Road Cross Section Images

This chapter has been published as: Brkić, I.; Miler, M.; Ševrović, M.; Medak, D. Automatic Roadside Feature Detection Based on Lidar Road Cross Section Images. Sensors *2022*, 22, *5510. https://doi.org/10.3390/s22155510.*

Conceptualization, I.B. and M.M.; investigation, I.B.; methodology, I.B., M.Š., and M.M.; supervision, M.M., M.Š., and D.M.; validation, I.B. and M.M.; visualization, I.B.; writing – original draft, I.B.; writing – review and editing, M.M., M.Š., and D.M.; funding acquisition, D.M.

Abstract

The United Nations (UN) stated that all new roads and 75% of travel time on roads must be 3+ star standard by 2030. The number of stars is determined by the International Road Assessment Program (iRAP) star rating module. It is based on 64 attributes for each road. In this paper, a framework for highly accurate and fully automatic determination of two attributes is proposed: roadside severity-object and roadside severity-distance. The framework integrates mobile L idar point clouds with deep learning-based object detection on road cross-section images. The You Only Look Once (YOLO) network was used for object detection. Lidar data were collected by vehicle-mounted mobile Lidar for all Croatian highways. Point clouds were collected in. las format and cropped to 10 m-long segments align vehicle path. To determine both attributes, it was necessary to detect the road with high accuracy, then roadside severity-distance was determined with respect to the edge of the detected road. Each segment is finally classified into one of 13 roadside severity object classes and one of four roadside severity-distance classes. The overall accuracy of the roadside severity-object classification is 85.1%, while for the distance attribute it is 85.6%. The best average precision is achieved for safety barrier concrete class (0.98), while the worst AP is achieved for rockface class (0.72).

3.1. Introduction

According to World Health Organization (WHO), road traffic injuries are the leading cause of death of people aged 5 to 29 years [98]. Road network infrastructure is strongly linked to the consequences of road accidents and the number of fatalities [8]. Therefore, the United Nations (UN) Member States have agreed on 12 new Voluntary Global Road Safety Performance Targets to drive action across the world. Two of the targets (Targets 3 and 4) include ensuring all new roads are built to a 3-star or better standard for all road users (Target 3), and that 75% of all travel is conducted on the equivalent of 3-star or better roads for all road users by 2030 (Target 4). UN estimates that 450,000 lives will be saved every year if these targets are applied in practice [7]. The number of stars is usually determined by the International Road Assessment Programme Star Rating (iRAP Star Rating). iRAP is the umbrella strategy for Road Assessment Programmes across the world (Europe – EuroRAP, Australia – AusRAP, New Zealand – KiwiRAP, China – ChinaRAP, USA – UsRAP, Brazil – BrazilRAP, South Africa – SARAP, Thailand – ThaiRAP, and India – IndiaRAP). iRAP Star Ratings is one of the five iRAP protocols, designed to collect road attributes on a particular road segment [26]. It is applicable for use also in low- and middle-income countries where data of road crashes is difficult to obtain. Likewise, iRAP Star Ratings is intended to assess infrastructure-related risk based on crash modification factors considering the likelihood and severity of individual user accidents with respect to the infrastructure features. The most dangerous roads, where the probability of a serious traffic accident with a fatal outcome is very high, are rated with 1 star, while the safest roads, where the probability of a fatal accident is zero, are rated with 5 stars [99]. In addition to the plans of the UN Global Road Safety Performance Targets, the collection of road attributes is important for European countries to comply with the European Union (EU) Directive 2019/1936 amending the 2008/96 Directive (RISM), which requires more detailed collection of road attributes to improve the safety of road infrastructure in EU member states [100]. iRAP Star Rating protocol correspond to approximately 95% of the indicative elements set out in the Annex III of the amended RISM Directive and thus can be successfully used to produce network classification in the Network-wide road safety assessment procedures.

In order to achieve the requirements of both iRAP and the EU directives there is a need for high-quality road data collection and extraction of road features. Various research efforts have approached this problem in different ways. The approaches to determining road attributes differ primarily in the selection of the sensors used to collect the data and in the selection of the techniques used to determine the attributes from the collected data. Since not all required road

attributes can be automatically collected with one type of sensor, it is necessary to use different sensors to collect different attributes. For example, sensors mounted on Unmanned Aerial Vehicles (UAVs) are the most-used to collect traffic flow data [66–68,101]. In addition, traffic flow data can be collected using instruments such as pneumatic road tubes and induction loops [57–59]. When it comes to attributes related to road infrastructure, georeferenced videos [102,103] or standard videos [104–106] are mostly used for data collection. Stated sensors are used for data collecting, but the standard process of attribute determination is still done manually by coding attributes from the collected data [107]. For this reason, iRAP is making efforts to develop new methods for automated data collection and attribute determination by taking advantage of state-of-the-art technologies such as machine learning, telematics, and Light Detection and Ranging (Lidar) [27]. The number of studies aimed at improving data collection and attribute determination methods using these technologies is increasing [104,105,108,109].

The main objective of this study is to provide a new framework for determining road infrastructure attributes. The framework consists of a combination of mobile Lidar sensor for collecting road infrastructure data as a point cloud and deep learning techniques for object detection for final attribute determination. Both the mobile Lidar sensor and deep learning-based object detection are advanced technologies proposed by iRAP that can improve accuracy and reduce the time required to determine road infrastructure attributes. As mentioned earlier, not all attributes can be determined by a single sensor. Therefore, this paper focuses on the complete and automatic solution of only two iRAP attributes: roadside severity-object (RSS – O) – related to roadside object detection – and roadside severity-distance (RSS – D) – related to the distance of the detected object from the road edge.

The proposed framework has multiple advantages with regards to related papers and standard processes of road infrastructure attributes determination. Firstly, using Lidar sensor enables spatial consideration, which is hard to achieve from video images. The importance of spatial considerations in terms of road safety is described in [110]. Since the vast majority of iRAP RSS – O classes are defined by dimensions (length, angles, etc.) as well as the distance of an individual object from the edge of road for the RSS – D attribute, this approach improves the determination of road attributes. Secondly, this paper proposes fully automated flow for the determination of the stated attributes. It is an improvement over manual attribute coding, mostly in the significant shortening of the duration of the process, but also in the consistency of attribute determination. Thirdly, the object detection part of study enables the detection of 13 different classes specified by the iRAP Coding Manual [26] and the determination of their

relative position to the road edge. Considering the complexity of iRAP RSS – O classes definitions, the classification of detected objects is achieved with high accuracy.

This paper is structured as follows: after a brief introduction to the research topic, the mention of the main contributions of this paper, and a brief overview of recent related studies, the proposed framework is described in detail. The framework is divided into two parts: data collection with point cloud processing and the object detection process on road cross section images. For better understanding of the whole process, the framework is presented with a corresponding diagram. Both parts of the framework have subsections that describe each step of the proposed framework in detail. This is followed by a results section, which presents the results of the object detection process as well as the spatial accuracy of road and roadside object detection and the final classification of road segments in terms of RSS – O and RSS – D attributes. The results are presented in the form of tables and confusion matrices. This is followed by a discussion of the results and the main advantages and disadvantages of the system in comparison with related works. Finally, based on the results and discussion, a brief conclusion is given, indicating future research options to improve the determination of iRAP attributes.

3.1.1. Related Works

There are a growing number of studies investigating ways to improve and automate the process of determining road infrastructure features. While road attributes represent the characteristics of the road segment in the database, road infrastructure features represent physical objects on the road and in the roadside area. The methods differ depending on the sensor used to collect the road data and how the road attributes are extracted from the collected data. Regarding the sensors used, videos are the most used [104,105,111,112], but there are few studies that use Lidar [108,109,113].

Sanjeewani and Verma (2017) [114] proposed a Fully Convolutional Network (FCN) to automatically find all AusRAP roadside objects: lanes, poles, sign boards, trees, metal barriers, warning signs, rumble strips, guideposts, concrete medians, etc. A Fully Convolutional Network (FCN) is a neural network that only performs convolution (and subsampling or upsampling) operations. Simplified, an FCN is a Convolutional Neural Network (CNN) without fully connected layers [115]. The technique is based on vehicle-based video data extracted into frames (images). The images were divided into homogeneous regions, which were used for image segmentation into AusRAP object classes on pixel base. Segmentation is performed by automated deep learning feature extraction based on a neural network with a classifier in the

last layer of the FCN. This means that one FCN was used to classify all attributes. As for the evaluation metrics, the paper reports the pixel-wise attribute classification accuracy after 10,000, 15,000, and 20,000 iterations of the FCN training procedure. Jan Z. et al. (2019) [112] proposed a CNN for the identification of all roadside objects. The technique is based on videos extracted into images. The approach is divided into three parts: image segmentation into nine AusRAP object classes by applying CNN, calculation of the distance between the road and the detected object, and evaluation of the proposed approach. As for the evaluation, a confusion matrix of the detected objects was provided, but there are no evaluation metrics for the calculated distances. In addition, the authors suggest that the use of Lidar could improve the detection results. Sanjeewani and Verma (2021) [105] improved their research from 2019, and the proposed model is also based on video data and FCN, but only one FCN is used for the detection of a single object class. Finally, an improvement is achieved by fusing all FCNs. The proposed approach is applied to 13 roadside object classes, such as speed signs, poles, trees, warning signs, etc. The paper provides an evaluation of the proposed approach with attribute-wise and pixel-wise accuracy. A comparison of the number of iterations with attribute-wise and pixel-wise accuracy is also provided. Sanjeewani and Verma (2021) [104] performed an optimization of the FCN-based approach for AusRAP attribute classification proposed in [105]. The FCN optimization is applied to four AusRAP road objects: Guidepost, Signal light, Flexipost, and Rumble strip. The optimization is based on finding the best combination of hyperparameters of the FCN, such as number of convolutional layers, activation function, pooling type, image size, number of iterations, and the optimization algorithm used. The paper also provides attribute-wise and pixel-wise accuracy for a single attribute.

When it comes to Lidar-based approaches for road data collection and determination of road attributes, Martin-Jimenez et al. (2018) [108] used mobile Lidar point clouds to assess road safety and estimate risk potential in Spain. The proposed approach is divided into four segments: classification of mobile Lidar point clouds based on geometric and radiometric properties of the point clouds; extraction of horizontal alignment and main road parameters based on geometric design consistency index; estimation of potential risk by a new predictive tool based on tree induction algorithm; and verification of the results in comparison with data from road safety experts considered as ground truth. The authors suggest that this approach may be suitable for the EuroRAP approach to risk assessment. Zhong M. et al. (2019) [109] proposed a point cloud classification framework for roadside safety attributes and distance detection. The framework consists of three stages: roadside point cloud data labeling; point cloud classification network; and object center approximation technique for distance calculation. The authors

developed a system for seven roadside object classes: pole, tree, road, guard rail, sign, vehicle, and other. The object-wise accuracy and the confusion matrix are given for only two object classes – pole and tree – while the pixel-wise accuracy is given for all detected object classes. For the same two object classes, the accuracy of determining the distance to the road is also given.

Similar research has been made about road and roadside features extraction from the fusion of Lidar and images independent of iRAP standards. Ural et al. (2015) [116] proposed an approach that incorporates data from the air: Color Infrared Orthophotos and Lidar Point Clouds. They applied Support Vector Machine (SVM) to segment the road surface from orthophotos. They eliminated the main obstacles such as buildings that have color similarities with the road surface using Lidar point clouds and ground filtering based on the Tri-angular Irregular Network (TIN) compaction method. They extracted 90.25% of all studied roads. Han et al. (2017) [117] proposed a road detection method based on the fusion of Lidar and image data. First, Lidar point clouds were projected onto monocular images. Then, color features were extracted from color images and used with the corresponding pixels in monocular images generated from the Lidar point cloud. These data were used for pixel-wise classification of roads using Adaboost classifier. The authors achieved acceptable performance but noted a large number of false positive road pixels. As a second limitation, they found that the number of false positive pixels increases with the increasing distance from the sensor due to the limited accuracy of the sensor. Zeybek, M. (2021) [118] propose a method for automatically extracting lane markings from Lidar data. The proposed method includes many different algorithms such as Cloth Simulation Filtering (CSF) to distinguish ground and non-ground data. Moreover, Random Sample Consensus (RANSAC) method was used to filter road surface from ground points. Finally, the Canny edge operator was used to extract the contours of the lanes.

3.2. Materials and Methods

This study was conducted with mobile Lidar data from the Croatian highway network. The Croatian highway network has a length of 1306 km, i.e., 2612 km in both directions. Point clouds for the highway network in both directions form an extensive and heterogeneous basis for the process of determining road infrastructure attributes. iRAP provides a list of 64 attributes [26]. All attributes must be collected for 100 m-long segments of the observed road for the road to be rated instars from 1 to 5. Croatian highways network consists of eight main parts. Every direction of every part is considered as single road. For single road segments are created from its start point every 100 m. Furthermore, every segment is coded with appropriate codes for all

of 64 iRAP attributes. As mentioned earlier, despite various research efforts on automation, the process of attribute determination for road infrastructure in practice consists of manually determining attributes for a given road segment from a georeferenced video.

The proposed framework is divided into two parts: mobile Lidar data collecting with point cloud processing and object detection process on crossroad images. Moreover, the mentioned parts are divided into more detailed parts, which are shown in Figure 3.1. Both parts and their corresponding subparts are explained in detail below.

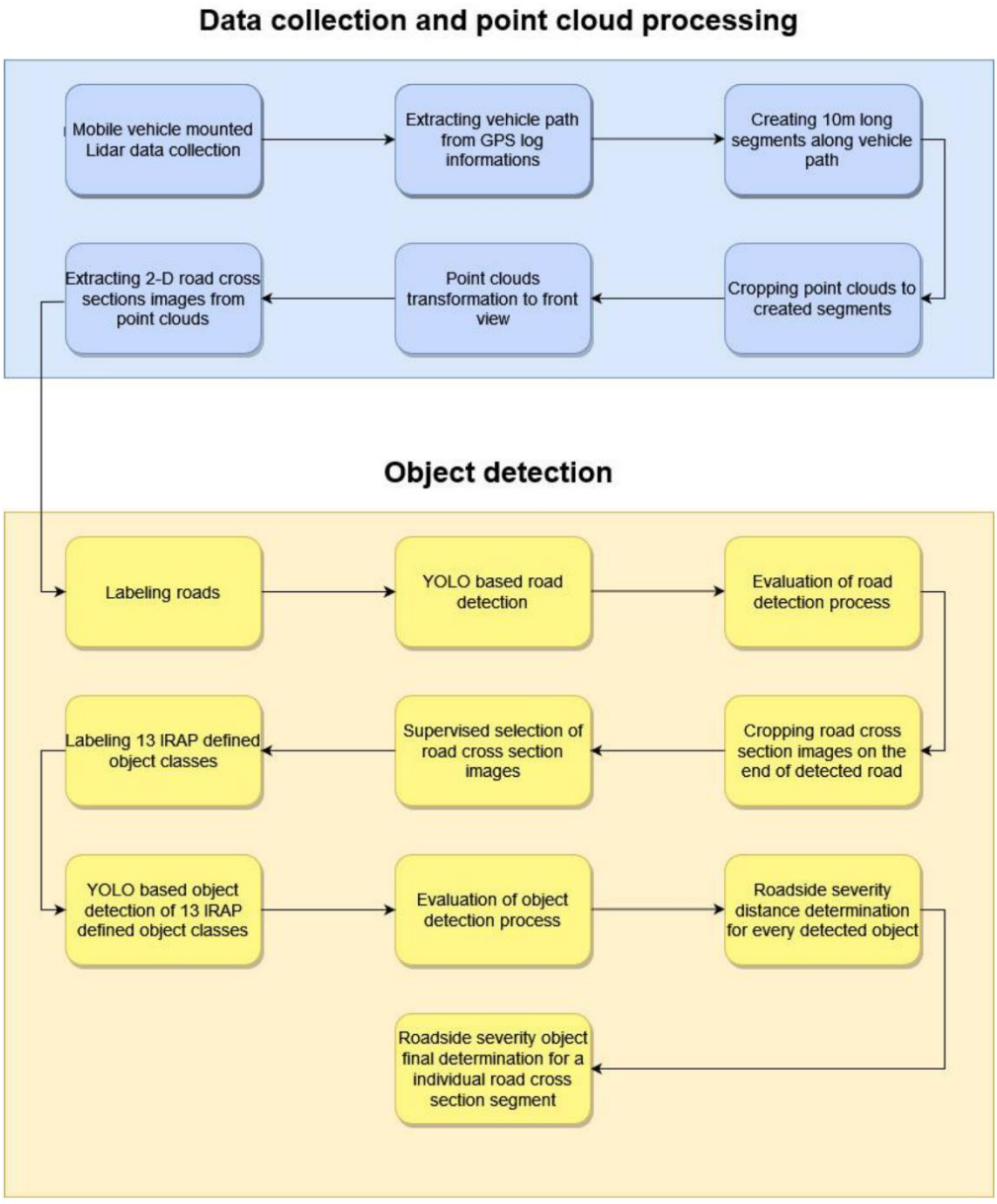


Figure 3.1 Proposed framework for the determination of RSS – O and RSS – D attribute.

2.1. Data Collection with Point Cloud Processing

A point cloud is a set of geometric points with coordinates sampled from 3D space [119]. Point clouds are usually generated by computer graphics or acquired by Lidar to represent 3D objects.

Lidar is based on a laser that is directed at the target, and the light beam is reflected from the surface. The sensor records the reflected light to measure the distance. Combining the laser distances with data from the integrated global navigation satellite system (GNSS) and the inertial measurement system (IMU) produces a dense, detailed group of points in space, i.e., a point cloud. Data collection for the study is performed using a Trimble MX 8 Land Mobile Mapping System (Trimble, Sunnyvale, California, USA). The technical specifications of the used system are listed in Table 3.1.

Table 3.1 Technical specification of Trimble MX8 Land Mobile Mapping System.

SYSTEM MODULE	Parameter	Value
Laser scanning	Accuracy	10 mm
	Precision	5 mm
	Frequency	Variable: 50 – 300 kHz (×2)
	Range	@50 kHz: 180 m σ ≥ 10%; 500 m σ ≥ 80%
		@300 kHz: 75 m σ ≥ 10%; 200 m σ ≥ 80%
Imaging modules	15 MP Forward panorama	Yes
	15 MP Rear panorama	Optional
	5 MP Oblique Surface	Yes
Positioning		POS LV 420

The collected point clouds are stored in las format. Since the Mobile Mapping System has two Lidar sensors, the point clouds for each of the sensors were collected separately, so they had to be merged to obtain a higher point density and larger field of view (FOV). The merging process was performed using the Python PDAL library.

Although the iRAP Manual Guide prescribes the assignment of a single object to a 100 m road segment, this paper performs this process for 10 m road segments to make more detailed determination of RSS – O and RSS – D attributes. Finally, 10 m road segments can be upsampled to 100 m road segments by selecting the most hazardous object among 10 road segments that are 10 m long. The hazardousness of an individual object is also defined in the iRAP Manual Guide by the type of roadside object and its distance from the roadside. Roadside hazards are listed in the iRAP Coding Manual [26] (page 50) in order from highest to lowest risk. The road segments are 10 m long and 40 m wide: 10 m on the left side of vehicle path and 30 m on the right side of vehicle path. The vehicle path was extracted from the Lidar GNSS log file. The data from GNSS were collected in International Terrestrial Reference Frame (ITRF) and then converted to the Croatian Reference Coordinate System (HTRS96/TM). An example

of a created road segment is shown in Figure 3.2, while an example of upsampling 10 road segments of 10 min length to an iRAP-defined 100 m road segment is shown in Figure 3.3.

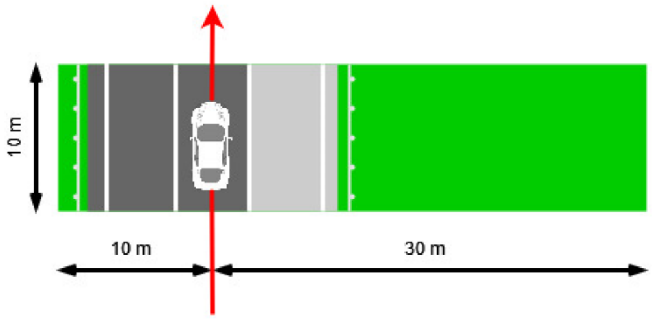


Figure 3.2 Example of one created road segment with appropriate dimensions. Red arrow represents driving direction, while black arrows represent dimensions of segment.

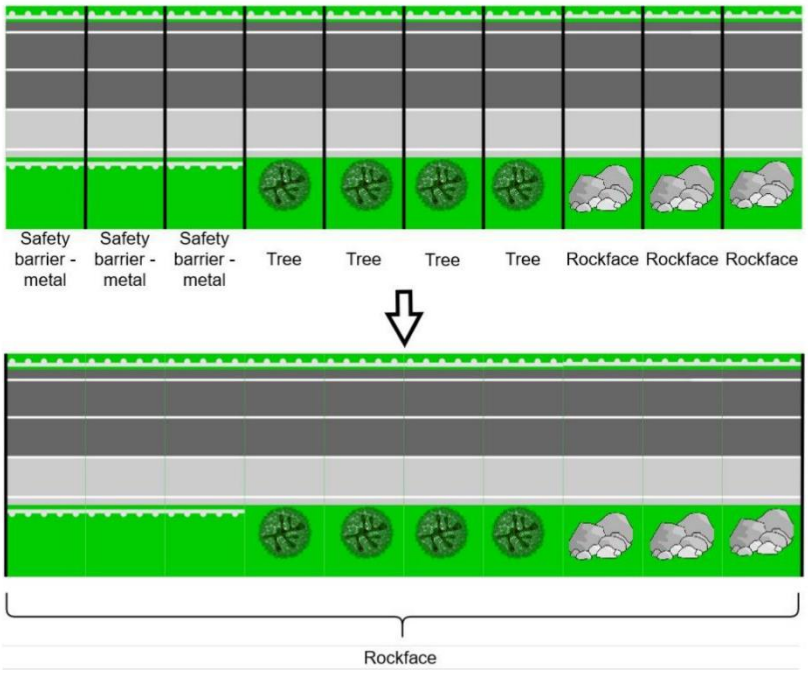


Figure 3.3 Example of upsampling 10 m road segments to one 100 m iRAP defined road segment.

After road segments were created, the collected point clouds were clipped to the boundaries of the road segments using PDAL python library.

To perform object detection on point clouds, the detection method must be carefully selected. According to [120], there are three main methods for object detection on point clouds: projection-based methods (front view and bird's eye view methods), voxel-based methods, and point-based methods. In this work, the frontal view method was chosen for object detection

because the RSS – O and RSS – D are best distinguished from the road cross-section. In order to obtain front view images of road sections, i.e., images of road cross sections, the point clouds must be transformed from the Croatian Reference Coordinate System to a local coordinate system. The origin of the local coordinate system is at the center of the Lidar sensor, the x-axis is orthogonal to the vehicle path and the y-axis is in the direction of the vehicle path. The transformation process was performed by applying the equation:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} cos\,Rz & -sin\,Rz & 0 & Tx\,cos\,Rz - Ty\,sin\,Rz \\ sin\,Rz & cos\,Rz & 0 & Tx\,sin\,Rz + Ty\,cos\,Rz \\ 0 & 0 & 1 & Tz \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \qquad 3.1$$

where vector [x', y', z',1] represents the coordinates of the single point in the point cloud after the transformation, Tx represents the translation in the x-axis direction, Ty represents the translation in the y-axis direction, Tz represents the translation in the z-axis direction, Rz represents the rotation angle about the z-axis, and the vector [x, y, z, 1] represents the coordinates of the single point in the point cloud before the transformation. The transformation process is shown in Figure 3.4.

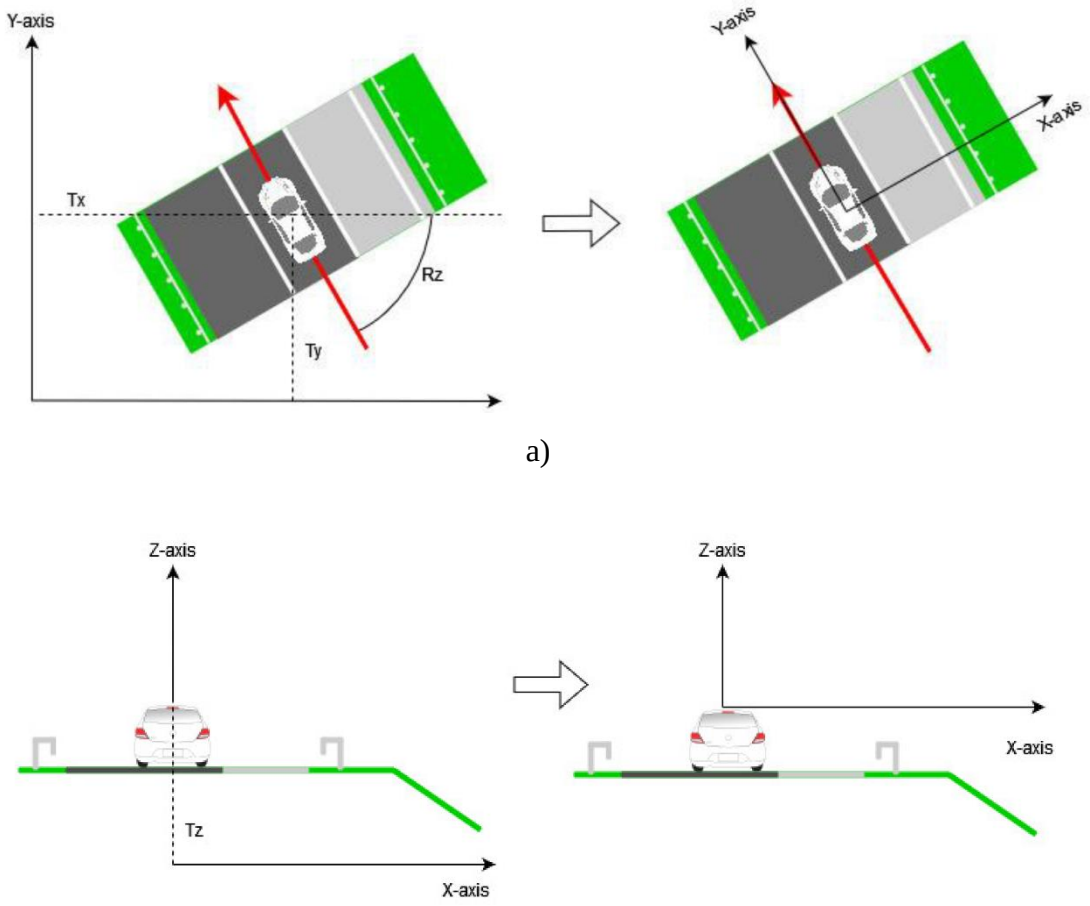
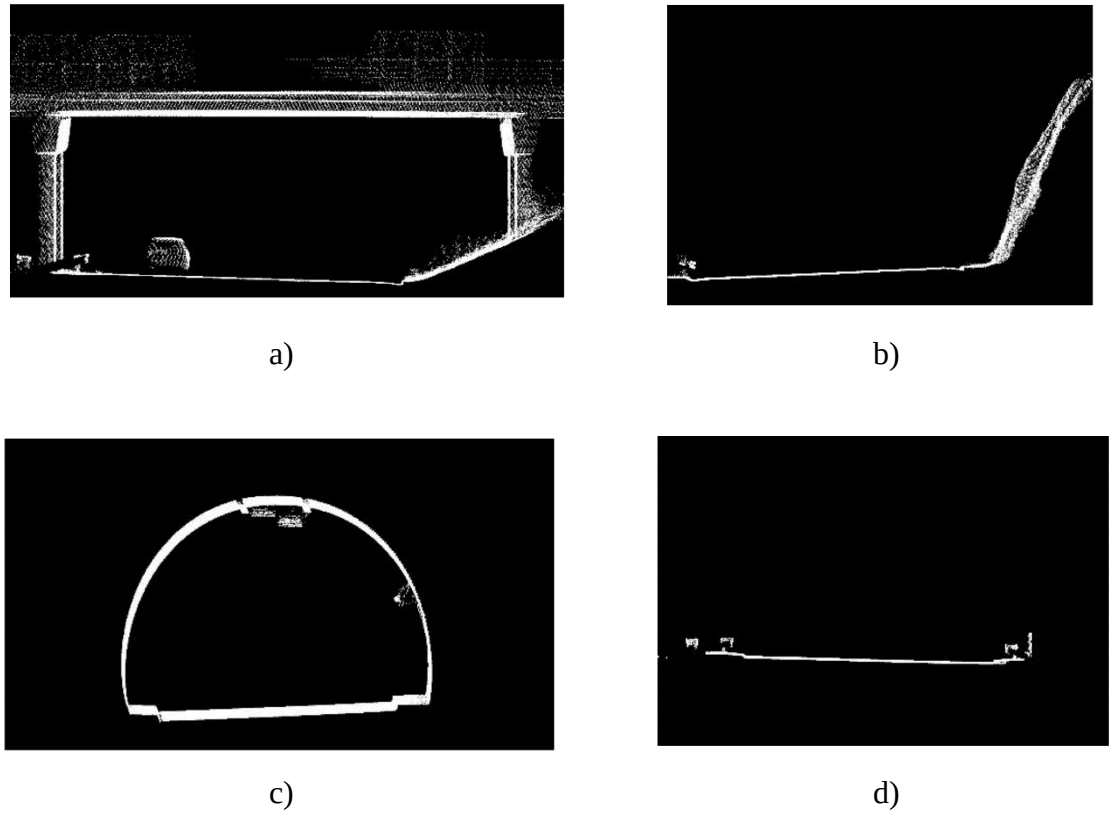


Figure 3.4 (a) Process of translation in direction of x-axis and y-axis and rotation around z-axis for Rz angle; (b) process of translation in direction of z-axis.

After the transformation process, the PDAL library was used to export point clouds into orthophoto images with depth of 10 m, i.e., road cross sections for 10 m of road. The exported images have a band with values of reflectance. The spatial resolution of the exported images is 1 cm × 1 cm. In terms of height, a range of 15 m above and 10 m below the Lidar sensor is covered. According to the height profile, the dimensions of the road segments and the spatial resolution, the dimensions of the images are 2500 px × 4000 px (2500 px × 1 cm = 25 m; 4000 px × 1 cm = 40 m). Examples of four exported road cross section images are shown in Figure 3.5.



a) b)

c) d)

Figure 3.5 (a) Example of road cross section on part of road with overpass; (b) example of road cross section on part of road with irregular rockface; (c) example of road cross section on part of road with tunnel; (d) example of road cross section on part of road with both safety barriers.

3.2.1. Object Detection

For the RSS – O attribute, iRAP defines 17 classes of objects, 13 of which are found on Croatian highways. The definitions of the individual object classes are given in the iRAP coding manual [26] (pages 52 – 54).

Regarding the RSS – D attribute, iRAP defines four classes: 0 – 1 m, 1 – 5 m, 5 – 10 m, and > 10 m from the roadside. To determine the RSS – D attribute, it is necessary to determine the road bounding box, focusing on the coordinate of the right edge of the road (Xmax). An example of a road bounding box is shown in Figure 3.6.

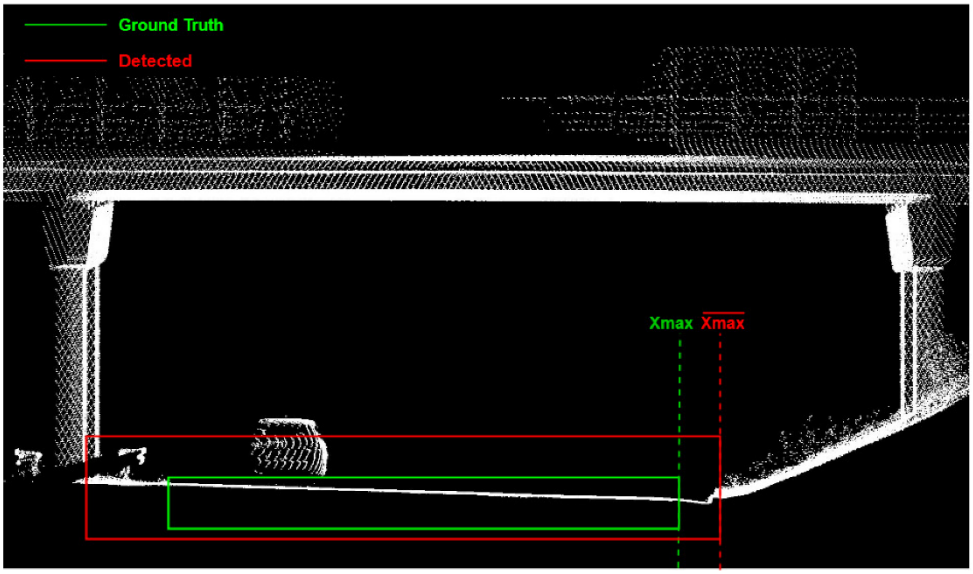


Figure 3.6 Example of road cross section with ground truth and detected road bounding boxes.

Road detection is performed by deep learning-based object detection algorithm. The You Only Look Once (YOLO) algorithm was used. This is a unified model for object detection which is trained on a loss function that directly corresponds to detection performance, and the entire model is trained jointly [36]. In recent years, five versions of YOLO have been released, each time with significant changes in the algorithm structure that improved both the inference time and the accuracy of the algorithm [121]. In terms of inference time and detection accuracy, YOLO was used in this work mainly because of the need to automate the process of classifying iRAP road segments which requires a fast inference time.

As for any object detection algorithm, it is necessary to have a sufficiently large, labeled dataset that allows adequate object detection. In this study, road labeling was performed using the labeling application on 5000 images of road cross sections. The labeled road cross-section images were divided into a train and a test dataset with a 75:25 ratio, i.e., 3750 train images and 1250 test images. The training process was performed in 1000 epochs with a batch size of 2 images. The training time was 15 h and 45 min on an NVIDIA GeForce RTX 2080 Ti GPU (NVIDIA Corporate, Santa Clara, CA, USA). Prediction process results with the confidence score of each detected road and the coordinates of the road bounding box (Xmin, Ymin, Xmax, Ymax).

The mean Average Precision (mAP) is calculated to evaluate the object detection process. It is the cross-class average of the interpolated Average Precisions (AP) [122]. AP represents the area under the recall-precision curve. It is the de facto standard for evaluating object detection

performance [123]. The computation of recall and precision and their meaning is described in detail in [124]. In this part of the framework, only one class (road) was detected, so mAP is equal to AP.

To evaluate the spatial accuracy of road detection, Root Mean Square Error (RMSE) value was used. RMSE is defined by equation:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Xmax_i - \overline{Xmax_i})^2}{n}} \qquad 3.2$$

where n is the number of road bounding boxes in the test dataset, $Xmax_i$ is the right road edge of a single ground truth road and $\overline{Xmax_i}$ is the right road edge of a single detected road.

Road cross-section images were used to label 13 iRAP-defined object classes. Supervised selection of road cross-section images was applied to label as many different object classes as possible on as few images as possible. Finally, 7804 images with 12,987 labeled objects were selected. The images were split into train dataset with 5853 images and a test dataset with 1951 images. An example of a road cross-section with labeled objects is shown in Figure 3.7. The training process was performed on the same graphics processor as the road detection. The training time was 35 h and 30 min with a batch size of 2 and within 350 epochs.

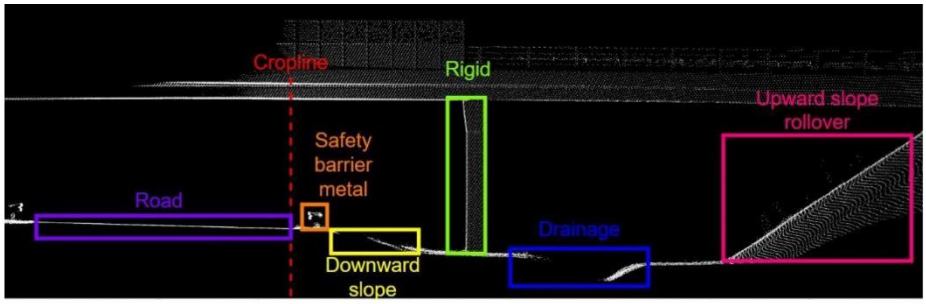


Figure 3.7 Example of road cross section with labeled iRAP defined objects.

In terms of object detection evaluation considering 13 iRAP object classes, AP was calculated for each object class as well as mAP. Except for AP and mAP, the confusion matrix is calculated based on the predicted and ground truth objects. Spatially, the RMSE was calculated for the reference X coordinate of each object class. The reference X coordinate for each object class represents the X coordinate of the detected bounding box, which is used to determine the distance between the road edge (Xmax of the road bounding box) and detected bounding box. For the tree, rigid, and semi-rigid object classes, the reference X coordinate is the center of the detected bounding box (Xcenter), while the left edge of the detected bounding box (Xmin) is

the reference X coordinate for other object classes. An example of reference X coordinate for the rigid and upward slope rollover classes is shown in Figure 3.8.

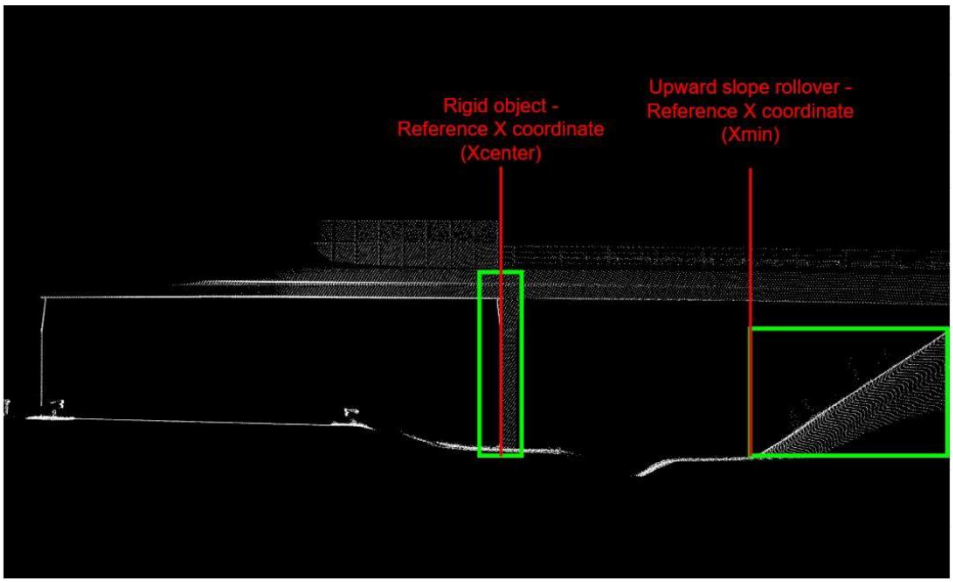


Figure 3.8 Example of reference X coordinate of rigid object and reference X coordinate of upward slope rollover object.

After detecting roadside objects and determining the distance to road edge, it is necessary to code only one object class for the whole 10-m road segment. It is based on the list of roadside hazards from the iRAP Coding Manual [26] (page 50). The list of roadside hazards is based on the object class defined by iRAP and its distance from road edge. After RSS – O is determinated, RSS – D is the distance from the reference X coordinate of RSS – O to road edge. The final values of RSS – O and RSS – D represent the predicted data in the evaluation process of this framework. For the ground truth data, the RSS – O attribute is manually coded for 1951 images in the test dataset. The manual coding is based on the combination of georeferenced video with road cross-section images. The RSS – D is measured manually on road cross-section images. To evaluate the final classification of road segments, the confusion matrix of predicted and ground truth data and other statistical values such as accuracy, precision, and recall are provided in this paper.

3.3. Results

The core part of framework is based on YOLO object detection of the road and iRAP attributes: RSS – O and RSS – D. Therefore, results are divided into object detection evaluation, spatial accuracy of detected objects, and evaluation of road segments classification.

3.3.1. Object Detection Evaluation

Road detection is performed with recall of 0.956, precision of 0.960 and AP 0.949. In terms of RSS – O, 12,987 iRAP defined objects were labeled. Distribution of labeled RSS – O classes is shown in Figure 3.9.

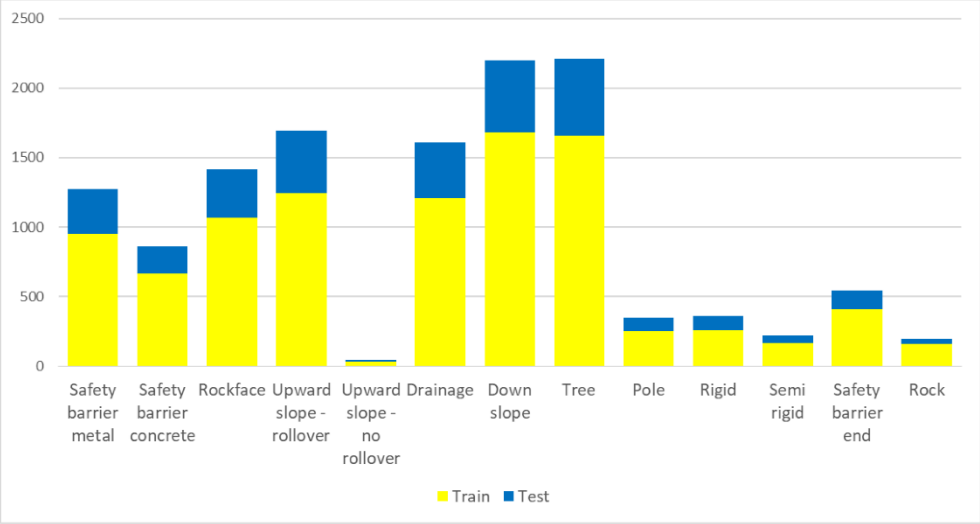


Figure 3.9 Distribution of labelled RSS – O classes.

To present performance of YOLO object detection on RSS – O, confusion matrix, recall, precision and AP are provided for each object. Stated evaluation metrics are based on a test dataset. Confusion matrix is shown in Table 3.2, while recall, precision and AP are shown in Table 3.3. Moreover, inference time for one image is 10.1 ms.

Table 3.2 Confusion matrix for predicted and ground truth objects presented with percentage.

		Ground truth													
		Safety Barrier Metal	**Safety Barrier Concrete**	**Rockface**	**Upward Slope-Rollover**	**Upward Slone-No Rollover**	**Drainage**	**Down Slope**	**Tree**	**Pole**	**Rigid**	**Semi rigid**	**Safety Barrier End**	**Rock**	**Background False Positive**
Predicted	**Safety barrier metal**	92.9	0.5										8.3		7.4
	Safety barrier concrete		98.0												0.2
	Rockface			78.5	3.6										10.7
	Upward slope-rollover			10.9	82.7	8.3									14.5
	Upward slope-no rollover					83.3									0.2
	Drainage						94.3	0.2							12.1
	Down slope							81.3							19.9
	Tree								90.2						23.5
	Pole									75.0					2.4
	Rigid					8.3					79.8				1.8
	Semi rigid										1.0	89.3			2.6
	Safety barrier end	2.8										1.8	88.0		4.0
	Rock													75.0	0.6
	Background False Negative	4.3	1.5	10.6	13.7		5.7	18.5	9.8	25.0	19.2	8.9	3.8	25.0	

Table 3.3 Recall, precision, and AP for each class as well as mean recall, precision and AP.

	Recall	**Precision**	**AP**
Safety barrier metal	0.87	0.92	0.91
Safety barrier concrete	0.99	0.98	0.98
Rockface	0.81	0.77	0.72
Upward slope-rollover	0.77	0.81	0.76
Upward slope-no rollover	0.90	0.83	0.83
Drainage	0.87	0.93	0.91
Down slope	0.78	0.89	0.85
Tree	0.81	0.88	0.82
Pole	0.87	0.71	0.73
Rigid	0.88	0.78	0.76
Semi rigid	0.78	0.89	0.87
Safety barrier end	0.79	0.88	0.85
Rock	0.90	0.75	0.74
Mean	0.85	0.85	0.83

3.3.2. Spatial Accuracy of Detected Objects

Spatial accuracy of detected object is presented by RMSE value calculated on the test dataset. The right edge of the road is detected with an RMSE of 0.08 m. RMSE values for every class of detected objects are presented in Table 3.4.

Table 3.4 RMSE value for every RSS – O class.

	RMSE (m)
Safety barrier metal	0.05
Safety barrier concrete	0.06
Rockface	0.27
Upward slope-rollover	0.65
Upward slope-no rollover	0.12
Drainage	0.73
Down slope	0.36
Tree	0.65
Pole	0.09
Rigid	0.08
Semi rigid	0.15
Safety barrier end	0.05
Rock	0.05
Mean	0.25

3.3.3. Evaluation of road segments classification

Evaluation of final classification of road segments into one RSS – O class and one RSS – D class is presented separately by accuracy, precision, recall, and confusion matrix.

Final classification of road segments into one RSS – O is conducted with an accuracy of 85.1%, precision of 0.888, and recall of 0.853. Confusion matrix of road segments classification into RSS – O classes is shown in Table 3.5. The “None” class in the predicted part of the matrix stands for those road cross sections where none of the iRAP-defined classes are detected.

Table 3.5 Confusion matrix for final road segments classification into one of RSS – O classes presented with percentage.

		Ground truth													
		Safety Barrier Metal	**Safety Barrier Concrete**	**Rockface**	**Upward Slope-Rollover**	**Upward Slone-No Rollover**	**Drainage**	**Down Slope**	**Tree**	**Pole**	**Rigid**	**Semi rigid**	**Safety Barrier End**	**Rock**	**Background False Positive**
Predicted	**Safety barrier metal**	78.6		0.4	0.6		0.5	1.1	1.3	2.8			2.4		78.6
	Safety barrier concrete		93.0		0.3			1.3	0.6						
	Rockface			84.1	5.7			0.2	2.6	1.4	6.1		0.8	20.0	
	Upward slope-rollover	1.9		8.3	88.1		0.5	0.2	3.9		2.0		0.8		1.9
	Upward slope-no rollover					90.0									
	Drainage	2.9	0.9	0.7	0.3		85.2	4.2	1.3	1.4					2.9
	Down slope	4.9	6.1	0.4			10.1	84.4	8.4	1.4			3.1		4.9
	Tree	1.9		1.1	1.8		1.1	5.0	77.3	1.4	6.1		2.4		1.9
	Pole	1.9			0.3				0.6	86.1			0.8		1.9
	Rigid				0.3	10.0				1.4	81.6				
	Semi rigid	1.9									0.0	93.1			1.9
	Safety barrier end	5.8		2.2	0.3		1.1	0.7	0.6		2.0	6.9	89.8		5.8
	Rock								1.3					77.1	
	Background False Negative			2.9	2.4		1.6	2.9	1.9	4.2	2.0			2.9	

In terms of road segment classification into one RSS – D class, an accuracy of 85.6%, precision of 0.825, and recall of 0.810 are achieved. Confusion matrix of road segments classification into RSS – D classes is shown in Table 3.6. Looking at both attributes collectively, 81.1% of all road cross sections in the test dataset are classified correctly by both RSS attributes.

Table 3.6 Confusion matrix for final road segments classification into one of RSS – D classes presented with percentages.

		Ground Truth			
		0 – 1 m	**1 – 5 m**	**5 – 10 m**	**>10 m**
Predicted	**0 – 1 m**	86.59	3.80	3.76	2.65
	1 – 5 m	9.96	88.43	11.65	15.04
	5 – 10 m	2.03	4.17	78.20	6.19
	>10 m	0.81	1.30	4.14	70.80
	None	0.61	2.31	2.26	5.31

3.4. Discussion

This paper proposes a new framework for determining two road attributes defined by iRAP: RSS – O and RSS – D. Compared to related papers, this work has several improvements in detecting roadside objects and their distance from road edge. First, this framework is based on Lidar data, i.e., point clouds that allow spatial observation. Several authors in related studies [104,105,111,112] have used vehicle-mounted cameras for a similar task and performed roadside object detection from videos. This approach is not suitable for determining the RSS – D attribute because high accuracy in distance determination cannot be achieved. This thesis is supported by the fact that only Jan Z. et al. (2019) [111] attempted to calculate the distance between roadside objects and the road edge from images but did not evaluate the distance determination. Additionally, using point clouds allows different views to road segments, including the front view, i.e., considering road segments as road cross sections. Apart from detection from videos, there are few studies [108,109,113] that have used Lidar data to determine RSS – O and RSS – D. While Martin-Jimenez et al. (2018) [108] do not focus on distance determination and evaluation of its accuracy, Zhong M. et al. (2019) [109] have

evaluated the accuracy of distance determination. They perform the evaluation of distance determination only for two classes: poles and trees. The given evaluation is expressed in terms of average error distance. The average error distance for poles and trees is 0.1 m and 0.5 m, while our average error distance for the same classes is 0.07 m and 0.38 m, respectively. It is obvious that proposed framework achieved much better accuracy in distance determination, especially for the tree class. Moreover, spatial observation allows very easy detection with high accuracy of those iRAP RSS – O classes defined by some spatial parameters such as angles for classes down slope, upward slope – rollover and upward slope – no rollover or width and height for the classes drainage and rock. For example, the down slope class is defined by iRAP as a roadside slope if the slope is less than −15°. The absence of these types of classes in papers [104,105,111,112] suggests that it is not possible to identify these classes from videos. While works based on Lidar data [108,109] did not solve this problem, this research proved that classes defined by spatial parameters can be detected from point clouds with high accuracy. Second, except for spatial observation, our framework is fully automated, which is achieved by a high level of inference time. In related works, any information on inference time has not been found. Therefore, our framework is not comparable to similar works in terms of inference time. Moreover, automating the process enables consistent determination of RSS – O and RSS – D classes which is a significant improvement over manual coding of road segments currently used in practice but prone to errors due to [107,109]. Third, the proposed framework includes detection of all RSS – O classes existing in Croatia. These are 13 classes, including those defined by spatial parameters. Related works [104,109,111,112] focus on the detection of only a few iRAP-defined classes such as poles, trees, roads, metal guideposts, warning signs, speed signs, etc. The mentioned classes are very easy to detect from images due to the large number of publicly available datasets containing these classes.

In the detection evaluation, proposed framework achieved high scores for precision, recall, and AP, i.e., mAP, especially when a large imbalance in the dataset is considered. The class with the highest AP is safety barrier concrete, followed by safety barrier metal (>0.92). The rockface, pole and rock classes have the worst AP (<0.75). Classes with clearly defined boundaries and shapes, such as safety barriers, are detected with high accuracy, while classes whose boundaries and shapes are not clearly defined, such as rockface and rock, are very difficult to detect. An exception is the pole class, which is not precisely defined in iRAP. In the iRAP Manual Guide [26], it is defined as any pole with a diameter greater than 10 cm. This includes everything from light poles to large guideposts. The mentioned objects are not even similar in shape and size, making detection very difficult. The problem can be solved by dividing objects from the pole

class to sub-classes (for example: poles, large signposts, small signposts, traffic signs, etc.) and placing them back into the pole class after the detection process. We assume that processing will increase the AP of the pole class defined by iRAP. Despite these obstacles, the final classification of road segments by RSS – O classes is 85.1%.

In terms of spatial evaluation calculated using the RMSE, the safety barrier metal, the safety barrier end and rock classes have the highest accuracy, while the tree and upward slope-rollover classes have low spatial accuracy. It is clear that classes with larger size have larger error, i.e., larger RMSE value, while classes with small size have small RMSE value. The RMSE value is directly related to the RSS – D attribute, which is correctly classified for 85.55% of the road segments.

Apart from all the improvements presented with this framework, there are some drawbacks. First, the price of the Lidar system is still much higher than that of cameras for video-based coding of road segments. There are some low-cost Lidar systems, but their accuracy does not match needs for solving this type of tasks. Considering that [125–127] some cars already use Lidar for autonomous driving [128–130], there is an intention that it will become cheaper and more affordable. In addition to the price and affordability of Lidar, another potential problem is the size of the dataset. iRAP RSS – O classes are very specific and there are not enough large datasets to help achieve higher accuracy. A large dataset would largely solve the problem of data imbalance, which in turn would solve the problem of many background false negative objects, i.e., unrecognized objects. Unrecognized objects can have a big negative impact on the final accuracy of the whole process, but also on road safety. This work is based on manually annotated objects, but the presence of a larger, balanced dataset would simplify and improve performances of the whole process.

3.5. Conclusion

In this paper, a framework for determining object and distance attributes for iRAP-defined roadsides is proposed. The framework is based on the integration of Lidar point clouds and deep learning-based object detection. It can be divided into two parts: mobile Lidar data collecting with point clouds processing and object detection process on road cross-section images. Compared to standard iRAP coding methods and recent works, this framework allows the spatial consideration of road segments, which is equally important for both RSS – O and RSS – D attributes. In addition, the framework is fully automated with a high level of inference time. It provides consistency in determining both attributes, which is a great improvement over the manual coding of road segments that is common practice today. It also allows road segments

to be classified into 1 of the 13 RSS – O classes defined by iRAP, as well as into 1 of the 4 RSS – D classes. Although most of the RSS – O classes are classified with a high AP, some classes have a lower AP value due to class imbalance in the dataset and fuzzy definitions of the classes by iRAP. Nevertheless, the final classification of road segments is achieved by the RSS – O attribute with an accuracy of 85.09%, while the classification accuracy related to the RSS – D attribute is 85.55%.

There is still much room for progress in the field of automated road infrastructure safety assessments. To improve the classification of the RSS – O attribute, a larger annotated dataset needs to be created to reduce the imbalance between classes and consequently achieve a high level of AP for all classes. Moreover, the augmentation process with an existing dataset can be explored to find out if it will result in the improvement of AP by single class. Furthermore, in addition to these two attributes, iRAP defines another 62 attributes that are necessary for road assessment. There is much room for exploring possible solutions for automating the process of determining other attributes by using different sensors and processing techniques.

Chapter 4 Utilizing High Resolution Satellite Imagery for Automated Road Infrastructure Safety Assessments

This chapter has been published as: Brkić, I.; Ševrović, M.; Medak, D.; Miler, M. Utilizing High Resolution Satellite Imagery for Automated Road Infrastructure Safety Assessments. Sensors *2023*, 23, *4405. https://doi.org/10.3390/s23094405.*

Author Contributions: Conceptualization, I.B. and M.M.; investigation, I.B.; methodology, I.B., M.Š., and M.M.; supervision, M.M., M.Š., and D.M.; validation, I.B. and M.M.; visualization, I.B.; writing – original draft, I.B.; writing – review and editing, M.M., M.Š., and D.M.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

Abstract

The European Commission (EC) has published a European Union (EU) Road Safety Framework for the period 2021 to 2030 to reduce road fatalities. In addition, the EC with the EU Directive 2019/1936 requires a much more detailed recording of road attributes. Therefore, automatic detection of school routes, four classes of crosswalks, and divided carriageways were performed in this paper. The study integrated satellite imagery as a data source and the Yolo object detector. The satellite Pleiades Neo 3 with a spatial resolution of 0.3 m was used as the source for the satellite images. In addition, the study was divided into three phases: vector processing, satellite imagery processing, and training and evaluation of the You Only Look Once (Yolo) object detector. The training process was performed on 1951 images with 2515 samples, while the evaluation was performed on 651 images with 862 samples. For school zones and divided carriageways, this study achieved accuracies of 0.988 and 0.950, respectively. For crosswalks, this study also achieved similar or better results than similar work, with accuracies ranging from 0.957 to 0.988. The study also provided the standard performance measure for object recognition, mean average precision (mAP), as well as the values for the confusion matrix, precision, recall, and f1 score for each class as benchmark values for future studies.

4.1. Introduction

According to the annual statistical report of the European Safety Road Observatory (ESRO) [131], there were 42 road fatalities per million inhabitants in the European Union (EU) in 2020, while the number of road fatalities was 67 in 2010. The ESRO pedestrian thematic report also informs that 20% of all traffic fatalities are pedestrians, with this percentage increasing to 38% in urban areas, a percentage that has been stable from 2010 to 2018 [132]. The European Commission (EC) has published an EU Road Safety Framework for the period 2021 to 2030 to reduce the above figures [20]. It is a set of intermediate targets to be achieved by 2030 to reach the long-term goal of zero road fatalities by 2050. One of the interim targets is to reduce traffic fatalities by 50% between 2021 and 2030 [20]. To achieve the stated goal, the EC defines Key Performance Indicators (KPIs) to measure progress toward the goals. Road infrastructure and environment are key factors for 30% of road crashes [19]. One of the KPIs therefore relates to road infrastructure and is defined as the percentage of distance traveled on roads with a safety rating above an agreed threshold [20]. Until the methodology and the threshold for safety rating are established, the KPI is defined as the percentage of distance traveled on roads with opposing traffic separation (by barriers or surfaces) relative to the total distance traveled [20]. Currently, the development of the methodology for safety ranking methodology is left to the EU member states. The EC is developing a common methodology based on the Road Infrastructure Safety Management (RISM) Directive 2008/96 and its amendment form directive 2019/1936, wherein indicated infrastructure elements used for road infrastructure safety assessments were defined [22]. While some EU member states have developed their own road assessment methodologies, many EU member states rely on the European Road Assessment Programme (EuroRAP), a European non-profit organization of automobile clubs, road authorities, and researchers [20]. This program results in safety ratings for roads between 1 and 5 stars, with 1-star roads having a high likelihood of road accidents resulting in serious injury or fatality, while 5-star roads have no likelihood of fatal outcomes in accidents [99]. EuroRAP is also a part of the International Road Assessment Programme (iRAP). Currently, iRAP (and therefore EuroRAP) is based on the collection of road and roadside attributes to provide a road safety rating. There are 66 defined road attributes and their classes, which are defined in detail in the iRAP coding manual [26]. Stated attributes can be divided into those which are collected from terrain data (52 attributes) and post-coding attributes (14 attributes).

Considering all the above, there is an obvious need to collect road attributes; they are a crucial factor in determining the level of road safety according to iRAP and in accordance with the EU

directive 2019/1936 [22]. Currently, iRAP-defined road attributes are collected manually from georeferenced video [107]. Lately, the development of technology enables using machine and deep learning techniques for collecting road attributes from different sources such as Unmanned Aerial Vehicles (UAVs), georeferenced video (front-view video from vehicles), Light Detection and Ranging (Lidar), etc. UAVs are used for traffic flow analysis [66,67,101,133], including number of vehicles' estimation on inspected roads, which is also one of the iRAP road attributes. In addition, UAVs are used for road marking detection [134,135], as well as road surface distress detection [136]. Vehicle-mounted georeferenced videos are a source of data mostly for traffic sign detection [137–139], but also for roadside feature detection [105,111,112,114] and road marking detection [140,141]. Lidar collects high numbers of spatial points as well as other attributes such as color, intensity, number of returns, etc. This makes it suitable for collecting a wide range of road attributes such as roadside feature detection [142,143] road surface distress detection [144], traffic sign detection [145], intersection detection [146], and determination of road geometry characteristics such as slope and curvature [147–149]. In terms of satellite imagery and road attributes, most research are focused on road extraction from optical as well as Synthetic-Aperture Radar (SAR) sensors [150–152]. They can be used in road attribute collecting processes but are not part of the road attributes defined by iRAP. There are studies that are focused on the direct detection of road attributes from satellite imagery such as road intersection detection [153] and pedestrian crossings [154–156]. This study proposed a new approach for road attribute detection using Very High Resolution (VHR) satellite imagery combined with deep learning techniques for object detection. The study focused on the detection of different pedestrian crossing types. Object detection was performed on spatially transformed road segments to distinguish between pedestrian crossings on inspected roads and on side roads. All detected classes of pedestrian crossings are defined in the iRAP Coding Manual [26]. In addition, apart from pedestrian crossings, object detection was performed to determine whether the road was divided or undivided, i.e., to detect objects or areas between carriageways where the carriageways were divided. Finally, school areas were also detected as part of iRAP-defined attributes.

This research has several scientific advantages over existing studies that focus on road attributes detection. First, satellite imagery allows easy global data access and road attribute detection without a physical presence on the road, which is a significant advantage over UAVs, georeferenced video, and Lidar. This is especially important for linear objects such as roads, which are difficult to cover with a UAV, but also very expensive with mobile Lidar or vehicle-mounted georeferenced video.

The second contribution of this research is the direct harmonization of the detected object classes with the road attributes defined by iRAP without the need for post-processing and reclassification.

The third contribution of this research concerns providing detection of divided and undivided carriageways, which directly relates to the preliminary definition of road infrastructure KPIs from the EC. In short, the detection of these objects allows the evaluation of road safety at this time, even if the methodology and thresholds for road safety are not yet defined. This is also one of the main attributes of a road, as many other attributes are defined differently depending on whether the carriageways are divided or not.

This paper is structured as follows: after a brief introduction of the research topic, the mention of the main contributions of this paper, and a brief overview of recent related studies, the proposed approach is described in detail. The approach is divided into three parts: vector data processing, satellite imagery processing and detection of school zones, pedestrian crossings, and divided carriageways. For a better understanding of the whole process, the framework is presented with a corresponding diagram. Every part of the framework has subsections that describe each step of the proposed approach in detail. This is followed by a results section, which presents the results of the object detection process. The results are presented in the form of tables and confusion matrices. This is followed by a discussion of the results and the main advantages and disadvantages of the approach in comparison to related works. Finally, based on the results and discussion, a brief conclusion is given, indicating future research options to improve the determination of iRAP attributes.

4.1.1. Related Works

Regarding the use of satellite imagery to detect road attributes, there are several works that focus on road markings and pedestrian crossings. Prakash et al. (2015) [154] proposed a framework for road markings detection. The method was based on satellite imagery from GeoEye and WorldView-2 satellites with a spatial resolution of 0.4 – 0.5 m. Open Street Map (OSM) data were used as a source of intersection locations. After extracting image tiles with intersections, the images were rotated to align the driving axis on the road with the vertical axis of the images. Finally, a periodic analysis was performed to decide whether a pixel represented an intersection or not. Overall, the recall rate and precision in the test sections were 63% and 89%, respectively. Ahmetovic et al. (2017) [157] proposed a two-stage framework for pedestrian crossing detection. In the first stage, crossing candidates are detected on satellite images provided by Google Map API. Crossings are detected by ZebraLocalizer, an algorithm

previously developed by the authors based on the geometric attributes of pedestrian crossings. This algorithm was implemented with a high recognition rate, so the process of generating crossing candidates contained many false positive examples. In the second phase, at the location of the intersection candidate, the authors developed a method for accessing Google Street View panoramic images. This was used for the final detection of whether the pedestrian crossing was at that location or not. The proposed approach achieved 77% to 95% of precision, while the achieved recall ranged from 90.2% to 97.1%. Berriel et al. (2017) [156] proposed a method for pedestrian crossing detection from satellite images. The method consisted of two stages: automatic data acquisition and annotation; model training and classification. OSM data were used to acquire crossing locations, while Google Maps API was used to access satellite imagery. Over 245,000 image tiles with positive and negative examples of crossings from more than 20 cities were created. ConvNet (Convolutional Neural Network) was used for binary image classification and achieved 97.11% accuracy. The spatial resolution of the satellite images used in this study was not reported. Ghilardi et al. (2018) [155] proposed a method to assist visually impaired people. The method was based on satellite imagery provided by Google Map API and a mobile application that warned people of nearby pedestrian crossings. The spatial resolution of the satellite images used was ~0.13 m. Crossing detection was performed using Google Maps road tiles (which are used to extract roads and mask the environment). Then, an SVM classifier was used for binary classification. The generated dataset consisted of 370 images with pedestrian crossings and 570 images without crossings and achieved 94.6% accuracy. Chen et al. (2021) [158] proposed a method for pedestrian crossing detection by fusing object detection tasks and image segmentation tasks. Image segmentation was performed using a U-net structure of CNN with the goal of extracting roads. The segmented images were used in combination with object detection techniques such as You Only Look Once v3 (YOLO v3), Faster R-CNN, and YOLO v3 based on DenseNet 121 to locate crossings. DenseNet-based YOLO v3 achieved the highest accuracy of 94.61%. Satellite images for London suburbs with a spatial resolution of 0.15 m were used in this study, but the source of the images was not provided.

4.2. Materials and Methods

This study was conducted in the wide area of Split. It is the second largest city in Croatia, with 161,312 inhabitants [159]. Therefore, this study focused on the collection of specific road attributes in urban and suburban areas. The total length of roads and highways surveyed was 83.5 km. The study area is shown in Figure 4.1, with the observed roads marked.

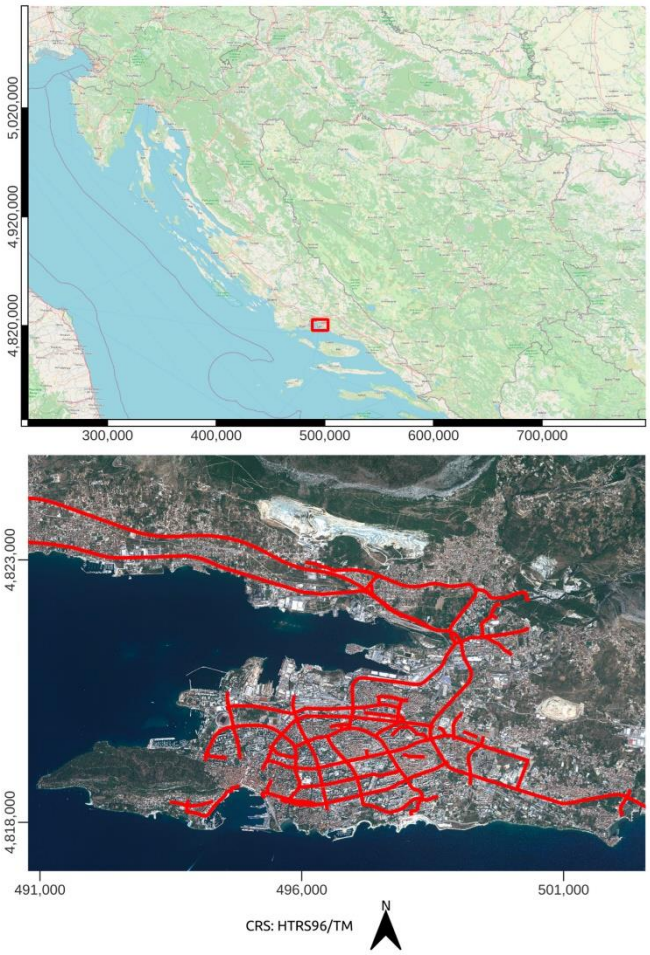


Figure 4.1 Study area on Open Street Map (OSM) and satellite imagery used in this study with observed roads plotted with red line (created by QuantumGIS software v3.22).

The study was divided into three phases. The first phase consisted of vector data processing, vectorization of road centerlines, and segmentation of observed roads with specific dimensions of road segments and percentage of overlap. The second phase included satellite image processing, band selection, cropping of the images with the boundaries of the road segments, and transformation of the cropped images to align the road centerline with the Y axis of the coordinate reference system (CRS), i.e., generating road-oriented images. Finally, the road segment images were used to detect school zones, physically divided roadways, and four types of pedestrian crossings. A complete overview of the workflow is shown in Figure 4.2.

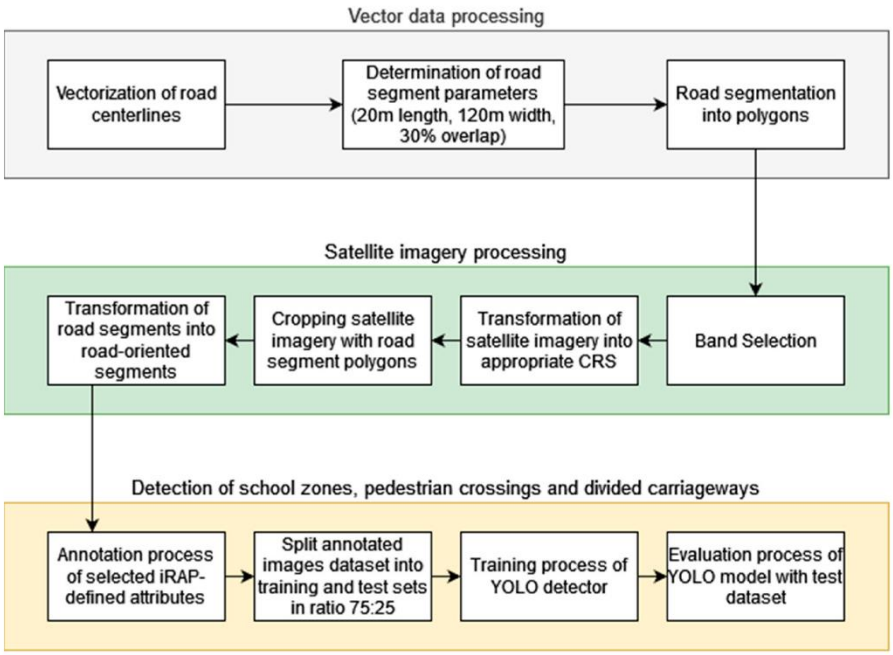


Figure 4.2 Three-stage workflow for detection of school zones, pedestrian crossings, and physically divided carriageways.

4.2.1. Vector Data Processing

The centerlines of all roads and highways were vectorized manually. According to iRAP Manual Coding [26], the road assessment is based on individual roads. Furthermore, the input datum for the assessment of a single road is the centerline, which must be vectorized manually. The single road that is the subject of the assessment is called the inspected road [26]. The process of vectorizing centerlines must comply with the rules established by iRAP. Parts of the inspected road where the carriageways are divided for a length of more than 400 m in a row are coded separately, which means that the centerlines for both carriageways must be vectorized. This role can also be clearly explained by the following equation:

$$nc = \begin{cases} 1, if\ l(x) \leq 400m \\ 2, otherwise. \end{cases} \quad 4.1$$

where nc is the number of centerlines to be vectorized and l(x) is the length of the undivided road sequence.

In this work, all observed roads were vectorized with a single centerline, since one of the results was to detect whether the road was divided or not. After vectorizing the centerlines of the observed roads, each road was divided into 20 m long and 120 m wide segments with 30% overlap. Although the iRAP coding manual defines a 100 m road segment as the basic unit for

road evaluation, smaller dimensions were used in this work to facilitate the fitting of the segment images into the YOLO network. With respect to the iRAP coding process, the conversion of smaller segments into 100 m segments was described in our previously published work [142]. When there were several different types of the same attribute, the riskiest attribute was coded. Since this study focused on capturing road attributes in urban and suburban areas, there were roads with different widths. To address this issue, segments were cut with a width of 120 m (60 m on the left and 60 m on the right side of the centerline of the road). In this way, both narrow urban roads and wide highways were included in the segments. An overlap of 30% was performed to avoid losing road attributes collected at the boundaries of the segments. Finally, 5846 road segments were created. The process of creating road segments was performed using the Python programming language and spatial vector-based packages such as GeoPandas and Shapely, while the centerlines were vectorized manually using QuantumGIS. The vectorization process was performed in the Croatian national CRS.

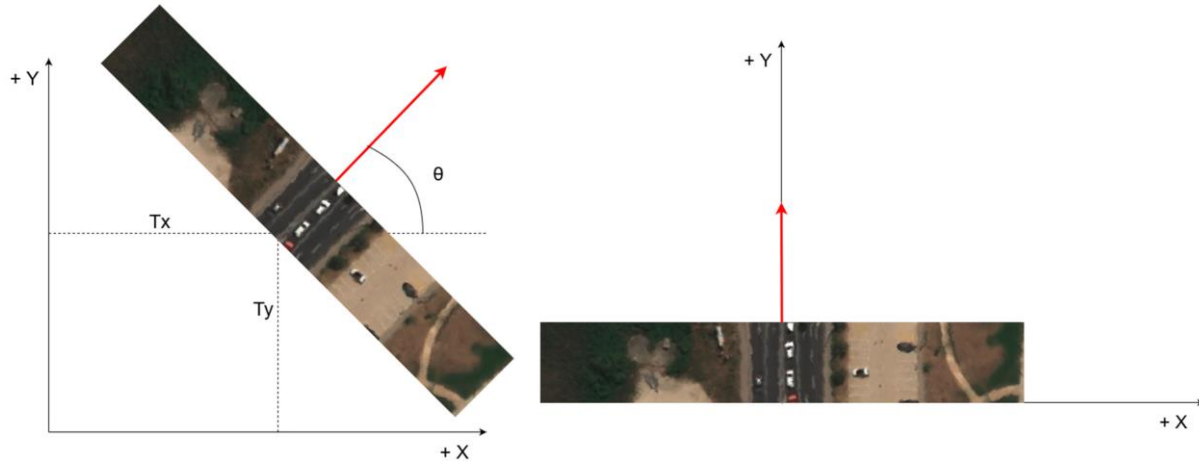
4.2.2. Satellite Imagery Processing

After preparing the vector data, where the road segments were polygons, the satellite images were processed. The source of the satellite images was the Pleiades Neo 3 satellite launched by Airbus Defence and Space in 2021. The satellite observes every point on Earth twice a day, which allows frequent temporal analysis of objects on Earth, including roads. The date used in this study to acquire the satellite images was 18 August 2022 and the satellite images covered 146.87 km^2. The spatial resolution of the Pleiades Neo 3 images was 30 cm, while the spectral resolution included seven spectral channels (panchromatic, deep blue, blue, green, red, and near infrared) [160]. In this study, the visible spectral bands were used (blue, green, and red). The satellite images were acquired from the WGS84/UTM zone 33N CRS. The first step was to convert the image position from the source CRS to the Croatian national CRS. Then, the imagery was cropped with the polygons of the created road segments. After cropping, the road segments were transformed into road-oriented segments. The transformation process included translation and rotation operations in the horizontal plane. This can be explained by the equation:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} cos\,\theta & -sin\,\theta & Tx \\ sin\,\theta & cos\,\theta & Ty \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad 4.2$$

where vector [x′, y′,1] represents the coordinates of the point in the road segment after the transformation, Tx represents the translation in the x-axis direction, Ty represents the

translation in the y-axis direction, θ represents the rotation angle, and the vector [x, y, 1] represents the coordinates of the point in the road segment before the transformation. The transformation process is shown in Figure 4.3. Finally, to fit the YOLO network, road segment images were converted from a GeoTIFF 16-bit format into a JPEG 8-bit format.



a) b)

Figure 4.3 Transformation process of cropped road segments. Red arrow presents centerline of road, while Tx, Ty, and θ are transformation parameters; (a) translation by Tx and Ty distances and rotation for θ angle; (b) transformed road segment with centerline aligned with Y-axis (road-oriented).

4.2.3. Detection of School Zones, Pedestrian Crossings, and Divided Carriageways

After the transformation of the road segments, an annotation process was performed to create an object detection dataset. To annotate objects, it is necessary to define them clearly. In this study, the focus was on school zone road markings, pedestrian crossings, and divided carriageways. All these attributes are defined by the iRAP Coding Manual [26].

A school zone attribute is divided into four types: school zone area without warnings, marked with road markings or appropriate speed limit signs, school with flashing beacons and appropriate speed limit signs, and areas without school zone. For every road segment, one of the above types must be coded. In this study, school zone areas were annotated with road marking types using satellite imagery.

Pedestrian crossing attributes presented most detected objects. According to the iRAP Coding Manual [26], pedestrian crossings can be divided into two classification tasks. The first classification task refers to whether the pedestrian crossing is on the inspected road or on a side road. This can be distinguished after the processing of vector data and satellite imagery, where each road segment is converted into a road-oriented segment. In addition, the second

classification task involves the classification into 11 classes related to the presence of pedestrian crossings, refugee islands, speed bumps, etc. All these classes are defined and described in detail in the iRAP Coding Manual [26]. In this study, two of these classes were found. The first class was the marked pedestrian crossing, which was defined as a clearly marked crossing without a refugee island. The second class was a pedestrian crossing with a refugee island. A refugee island is defined as a purpose-built safe stopping point for pedestrians at the halfway point. It must provide adequate space and protection from passing vehicles and must be seen by drivers. For a better understanding of the pedestrian crossing classification tasks and annotated classes, a diagram of the classification tasks and annotated classes with corresponding examples is shown in Figure 4.4.

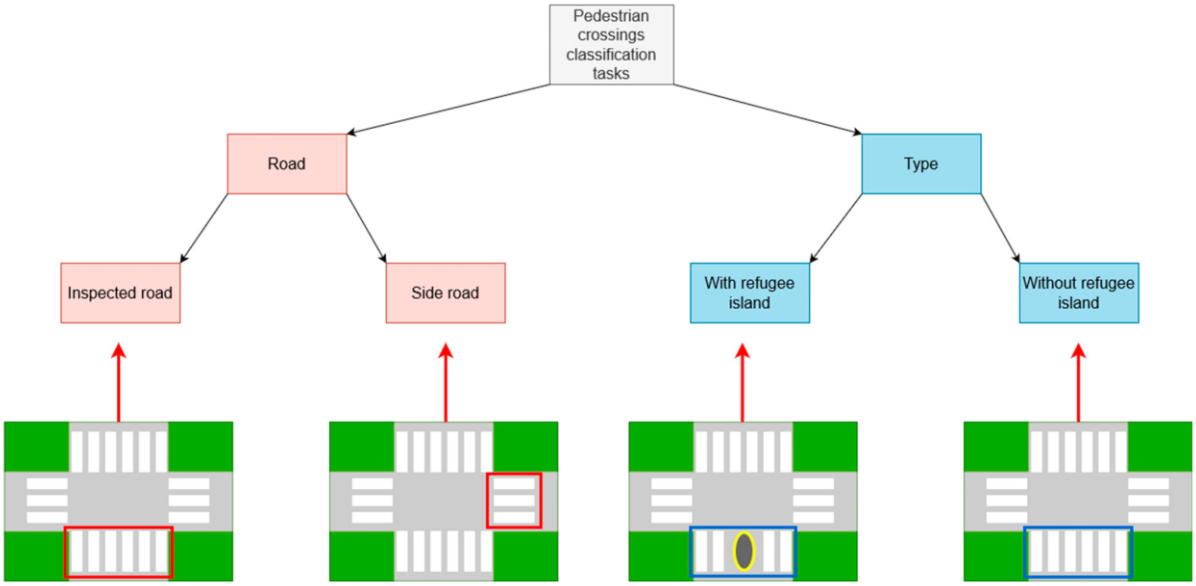


Figure 4.4 Diagram of pedestrian crossing classification tasks and classes annotated for this study with appropriate examples. Red arrow presents centerline of inspected road.

An annotation example of school zone road markings is shown in Figure 4.5a. By integrating two pedestrian crossing classification tasks, this paper ultimately focused on four classes of pedestrian crossings: pedestrian crossing on the inspected road (Figure 4.5b), pedestrian crossing on the inspected road with a refugee island (Figure 4.5c), pedestrian crossing on the side road (Figure 4.5d), and pedestrian crossing on the side road with a refugee island (Figure 4.5e).

The final attribute in this study was divided carriageways. While an undivided carriageway has no physical separation between opposing traffic flows, divided carriageways are those that physically separate opposing traffic flows by either a barrier or a wide physical median [26].

To tackle this attribute, divided objects (safety barriers, land areas, etc.) were annotated. An example of a road segment with divided carriageways is shown in Figure 4.5f.

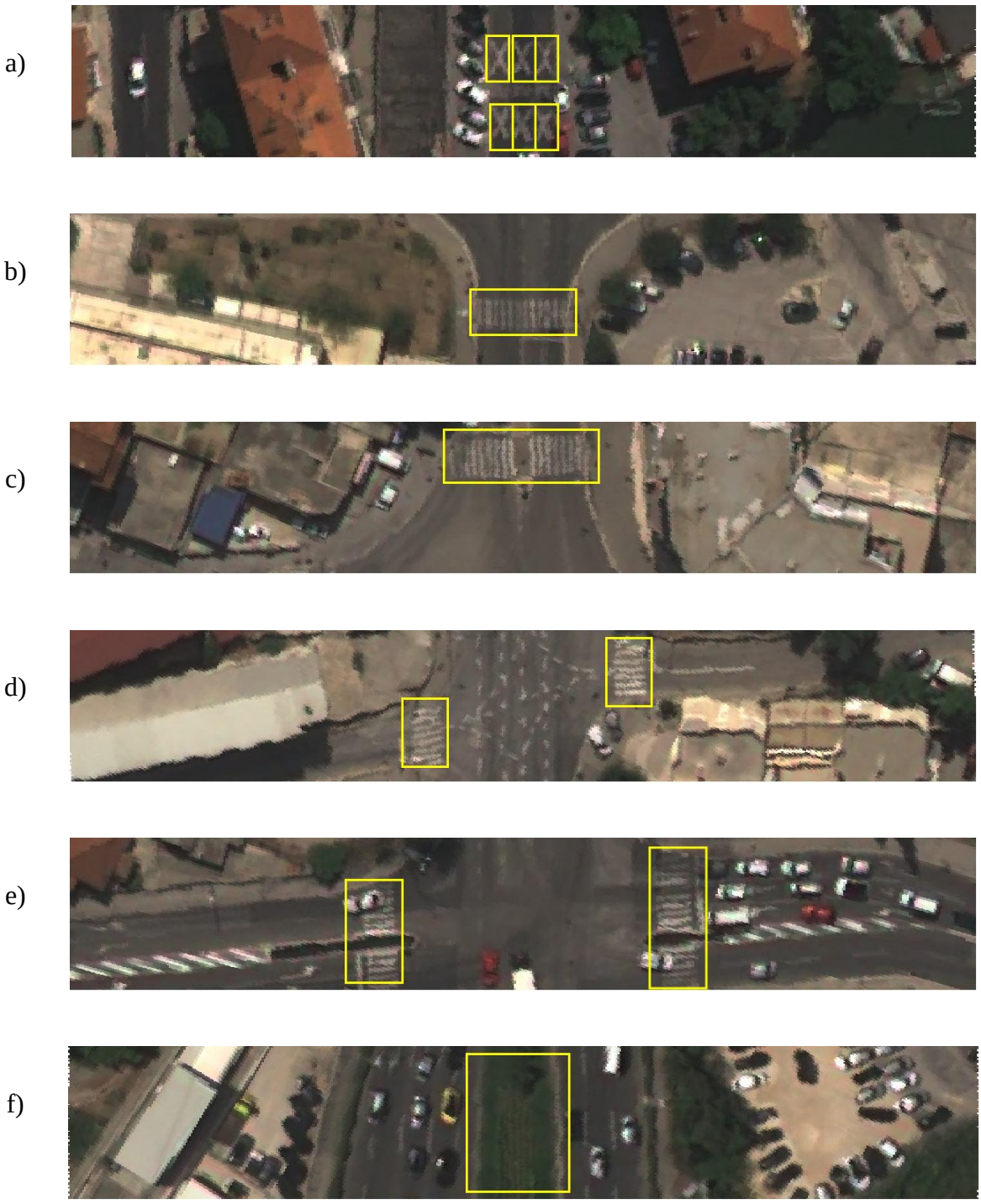


Figure 4.5 Examples of annotated classes on road segments. (a) school zone; (b) inspected road pedestrian crossing; (c) inspected road ossing with refugee island; (d) side road pedestrian crossing; (e) side road pedestrian crossing with refugee island; (c) inspected road pedestrian crossing with refugee island; (d) side road pedestrian crossing; (e) side road pedestrian crossing with refugee island; (f) divided carriageways.

4.2.3.1. Experiment Analysis

After the annotation process, the YOLO object detector was trained and tested. YOLO is a widely used algorithm. It has a small architecture size and a high inference speed [44]. It is also a single-stage detector with unique features such as small models with respectable inference times [161]. In this work, the fifth version of the YOLO detector was used [46] due to its high inference speed, which was significant for the processing time of a large number of road kilometers. To achieve more accurate models, dataset size can be crucial [162]. After a detailed analysis of existing freely available object detection datasets, no dataset containing iRAP-defined attributes was found. Therefore, a manual annotation was performed using the software LabelImg [163]. All 5846 road segments were annotated and only those that had one of the defined attributes were selected for the learning process. Therefore, 2602 images were selected for the learning process. All images were divided into training and test datasets in a 75:25 ratio of annotated samples. In terms of images, it amounted to 1951 training images and 651 test images. The training process was performed in 600 epochs and 12 h and 30 min on an NVIDIA GeForce RTX 2080 Ti GPU. In addition, the training process included image augmentation to increase the training dataset in order to achieve higher performances. The augmentation process included the transformation of images into Hue, Saturation, Value space (HSV) and left – right and up – down flipping and scaling. The prediction process provided the class of the detected object, the confidence rate (which indicated the probability that the detected object actually belonged to the detected class), and the image coordinates of the bounding boxes of the detected objects.

The evaluation process of the trained YOLO detector is expressed by the mean average precision (mAP), which is a standard for the evaluation of object detection models [123]. It is defined as the mean over classes of the interpolated Average Precision (AP). AP is given by the area under the precision – recall curve of the detected objects [122]. Definitions of precision and recall values are provided in [124]. In the mAP calculation, it was necessary to define what was a true prediction and what was a false prediction. For this purpose, the Intersection over Union (IoU) value had to be defined. The IoU value expressed the ratio between the intersection and union area of the true and predicted bounding box. The IoU value can be explained by the following equation:

$$IoU = \frac{GT \cap P}{GT \cup P} \tag{4.3}$$

where GT is a bounding box of the ground truth object, while P is a bounding box of the predicted object.

The IoU was set to 0.5. Therefore, predicted bounding boxes with an IoU greater than 0.5 were considered correct predictions, while others were considered incorrect. If there were multiple correct predictions for the same ground truth object, the predicted bounding box with the highest confidence rate was considered a correct prediction, while the other predictions were classified as incorrect. Therefore, IoU was the basic value for providing mAP, as well as the confusion matrix, of ground truth and predicted objects and further calculations of other statistical performance measures such as accuracy and f1 score. The above measures are described in detail in [89].

4.3. Results

The training dataset contained 2515 samples, while the test dataset contained 862 samples. The distribution of annotated training and test samples is shown in Figure 4.6. The figure shows more divided carriageway samples in the training and test datasets regarding the other classes.

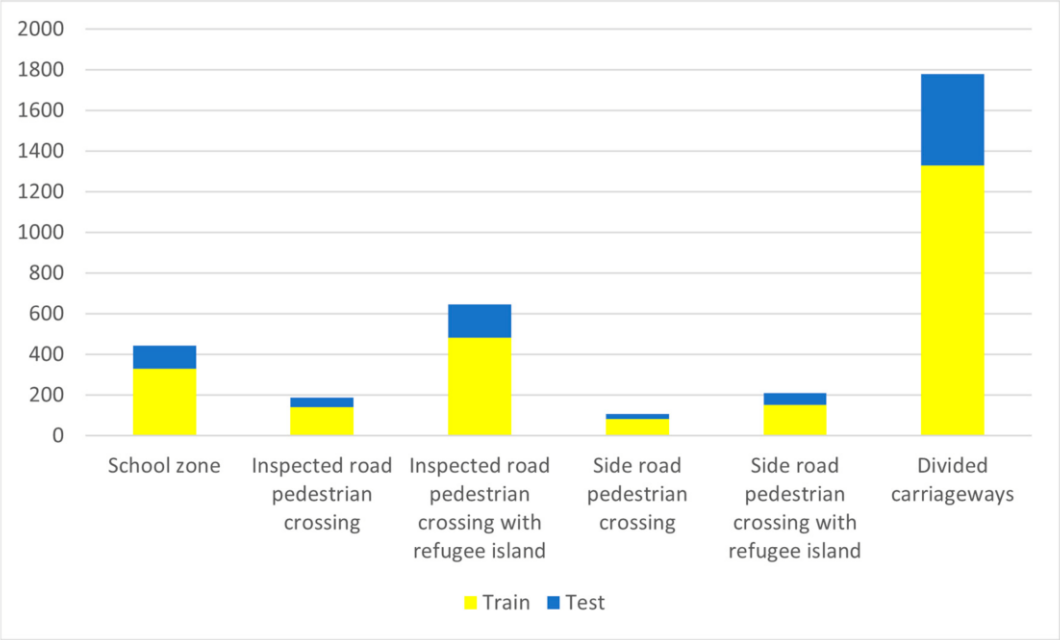


Figure 4.6 Distribution of annotated samples of school zones, divided carriageways, and pedestrian crossing classes. Number of training samples are shown in yellow, while test samples are shown in blue.

After conducting the training process, the evaluation process resulted in a confusion matrix, which is shown in Table 4.1 Confusion matrix of ground truth and predicted objects. In addition

to the detected objects, the number of Background False Positive (FP) and Background False Negative (BFN) detections is given. BFN is given on the predicted axis, while BFP is given on the ground truth axis. Bright shades of green present a lower number of matched classes between ground truth and predicted objects. Contrary, dark shades of green present higher number of matched classes between ground truth and predicted objects.. The confusion matrix provided data for the determination of performance measures such as accuracy, precision, recall, and f1 score. The mean values of the stated measures and the mAP value for all classes are shown in Table 4.2, as well as the same values per class. The precision – recall curve generated for the fixed IoU of 0.5 and confidence of 0.5 is shown in Figure 4.7. The visualization of correctly detected objects for each class is shown in Figure 4.8, while examples of false positive, false negative, and misleading detections are shown in Figure 4.9.

Table 4.1 Confusion matrix of ground truth and predicted objects. In addition to the detected objects, the number of Background False Positive (FP) and Background False Negative (BFN) detections is given. BFN is given on the predicted axis, while BFP is given on the ground truth axis. Bright shades of green present a lower number of matched classes between ground truth and predicted objects. Contrary, dark shades of green present higher number of matched classes between ground truth and predicted objects.

		Ground Truth						
		SZ	**IRPC**	**IRRI**	**SRPC**	**SRRI**	**DC**	**BFP**
Predicted	**SZ**	55						9
	IRPC		102	1				10
	IRRI		5	45				3
	SRPC				147	3		19
	SRRI			1	4	22		2
	DC						433	26
	BFN	4	10	1	14	1	21	

SZ – School Zone; IRPC – Inspected Road Pedestrian Crossing; IRRI – Inspected Road pedestrian crossing with Refugee Island; SRPC – Side Road Pedestrian Crossing; SRRI – Side Road pedestrian crossing with Refugee Island; DC – Divided Carriageways; BFP – Background False Positive; BFN – Background False Negative.

Table 4.2 Accuracy, recall, precision, F1 score, and AP for each class as well as mean values of every performance measure.

	Accuracy	**Recall**	**Precision**	**F1 Score**	**AP**
SZ	0.988	0.849	0.938	0.891	0.968
IRPC	0.957	0.870	0.891	0.880	0.939
IRRI	0.988	0.759	0.846	0.800	0.887
SRPC	0.986	0.859	0.932	0.894	0.923
SRRI	0.972	0.903	0.872	0.887	0.798
DC	0.950	0.943	0.954	0.949	0.979
Mean	0.974	0.864	0.905	0.884	0.91

SZ – School Zone; IRPC – Inspected Road Pedestrian Crossing; IRRI – Inspected Road pedestrian crossing with Refugee Island; SRPC – Side Road Pedestrian Crossing; SRRI – Side Road pedestrian crossing with Refugee Island; DC – Divided Carriageways.

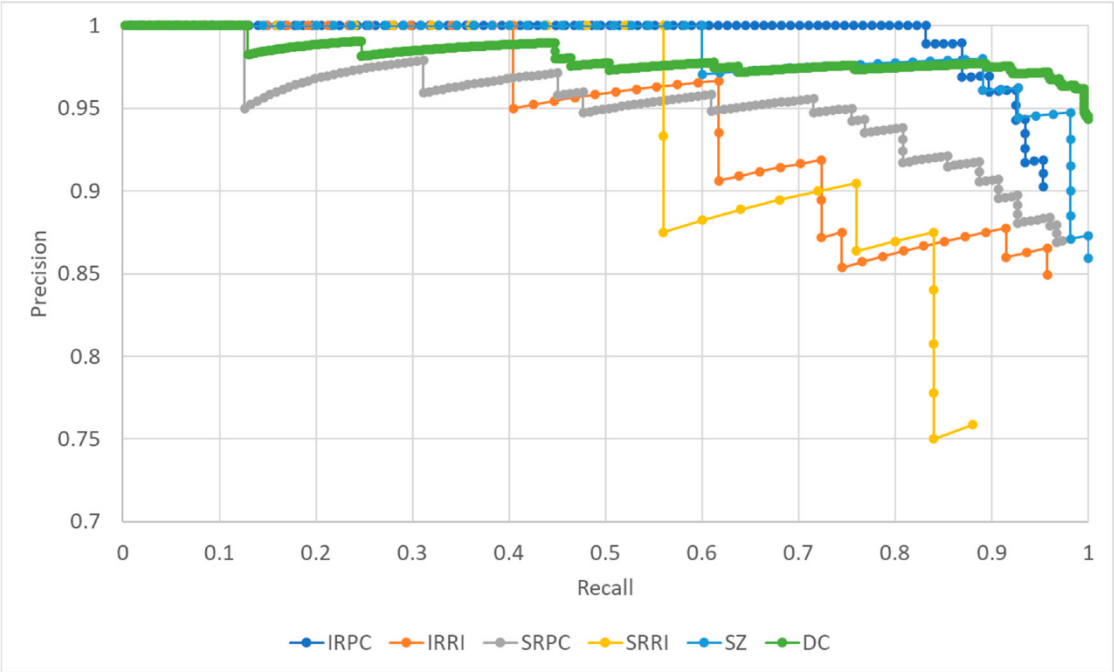


Figure 4.7 Precision – recall curve generated for each class with IoU and confidence thresholds at 0.5. It is visible that the object detector achieved a significantly lower precision – recall tradeoff for SRRI and IRRI classes.

a)

b)

c)

d)

Figure 4.8 Examples of correctly detected classes by trained YOLO detector. Green bounding boxes show ground truth objects, while g boxes present predicted objects; (a) school zone markings; (b) inspected and side road pedestrian crossings, (c) inspected and side road pedestrian crossings with refugee islands as well as side pedestrian crossing on right side; (d) divided carriageways.

a)

b)

c)

d)

*Figure 4.9 Examples of false positive, false negative, and misleading detections by trained YOLO detector. Green bounding boxes show ground truth objects, while red bounding boxes present predicted objects; (a) misleading detection of divided carriageways on non-inspected road; (b) false negative detection of side road pedestrian crossing; (**c**) false positive detection of inspected road crossing on roadside object; (d) false negative detection of side road pedestrian crossings.*

4.4. Discussion

This study offers significant improvements over related studies. According to EC, there is a need to collect clearly defined road attributes to evaluate road safety. This study focused on collecting school zones, pedestrian crossings, and divided carriageways. All of these road attributes are clearly defined by the iRAP program, which is used in many European countries to assess road safety. As for the detection of school zones, this has not yet been the subject of any research. In this work, it was shown that the integration of satellite imagery with deep learning object detection enabled the detection of iRAP-defined school zones with high

efficiency. In addition, divided carriageways have not been the subject of previous studies, although this is a critical attribute for both iRAP-defined and EC-defined KPI. Road attributes in the iRAP program are defined separately for divided and undivided carriageways. The preliminary KPI definition from the EC also includes divided carriageways. All these indicate that divided carriageway detection will soon be an important step in road safety assessment. In this study, divided carriageway detection was achieved with high efficiency and the performance measures provided could serve as benchmark values for future work.

With respect to pedestrian crossings, there are many works that focus on pedestrian crossing detection from various sources such as UAVs, vehicle-mounted lidars, or georeferenced videos. Considering that high-resolution satellite imagery is a more cost-effective solution, especially for linear objects such as roads very difficult and expensive to cover using the aforementioned technologies, this work focused on pedestrian detection using satellite imagery. There are several works that have used satellite imagery for the same task. While related studies have focused on the detection of one type of pedestrian crossings, this study focused on the detection of four iRAP-defined classes of pedestrian crossings; this was a much more detailed but also more challenging task. It was made possible by segmenting roads and transforming segments into road-oriented segments. Therefore, it was easier to distinguish pedestrian crossings on the inspected road from those on the side road.

Compared to related works, Prakash et al. (2015) [154] provided a similar approach in the vector processing of road segments, but they performed a pixel-based periodic analysis on satellite imagery as part of the detection process. They achieved a precision of 0.89 for one class of pedestrian crossings, while the precision in our study ranged from 0.846 to 0.932 for four different types. They also achieved a recall value of 0.63, while our Yolo detector had a recall value between 0.759 and 0.903. In contrast to Prakash et al. (2015) [154], Berriel et al. (2017) [156] performed binary image classification using ConvNet on a large dataset of pedestrian crossing tiles from Google Maps with a binary accuracy of 0.97, while the accuracy in our research ranged from 0.957 to 0.988 for four classes. In addition, Ghilardi et al. (2018) [155] performed classification with an SVM classifier and achieved an accuracy of 0.946, while Chen et al. (2021) [158] used the YOLO v3 deep learning-based detector for pedestrian crossing detection and achieved an accuracy of 0.946. All these performance measures show that our approach had similar efficiency to other deep learning-based approaches such as ConvNet and Yolo v3 and higher efficiency than approaches based on pixel-based periodic analysis and other machine learning algorithms such as SVM classifiers. The results were expected regarding the use of Yolo v5, which has already proven to be better than previous versions of Yolo [164].

Apart from the above advantages and high rate of performance measures, this approach had some limitations. The trained Yolo detector had some disadvantages such as misleading and false detections due to different reasons. With our approach, divided carriageways were detected with high performance measures, but the detected objects were not always on the inspected road. This was a significant problem, especially if the inspected road was an undivided carriageway. Although the detection was correct, it was misleading. This problem could be solved by cutting off less wide road segments that include only a narrow area around the inspected road. In this case, the roads must be divided into different classes to cut off different width road segments depending on the road class. Additionally, one of the major limitations of this approach was that the detection quality was based on the quality of road markings, which depended on road maintenance services. Therefore, it cannot be controlled. Although there are different laws on adequate road maintenance, unfortunately they are not always realized. This could be overcome with stricter law enforcement. Furthermore, another limitation could be the high rate of false detections due to similar patterns of school zones, pedestrian crossings, and divided carriageways with roadside objects. This is the case when a training dataset is not large enough for the detector to distinguish stated objects. This problem is generally a major obstacle in collecting road attributes based on deep learning approaches. There is no such large dataset that is harmonized with iRAP-defined road attributes. Therefore, it is necessary to build a larger dataset in the future to enable deep learning approaches for road attribute detection. Finally, the major obstacle in urban areas could be unrecognized pedestrian crossings due to shadows, but, with the development of satellite technologies and the annual increase in the number of satellites, this problem could be minimized. More satellites in space could make it possible to avoid shadows by choosing the time of day when the area is observed.

4.5. Conclusions

From the above results and discussion, it is clear that the approach proposed in this paper has several important advantages. First, it focused on the detection of road attributes defined by iRAP, the main framework for road safety assessment in many European countries. Divided carriageway detection is also a significant step forward, as the EC temporarily defines a KPI that includes the length of physically separated roads in its definition. In terms of performance, this approach proved that the integration of satellite imagery and the Yolo object detector achieved a very good performance. The use of high-resolution satellite imagery is a more cost-effective solution, especially for linear objects such as roads. While school zones and divided

carriageways have not yet been explored, the performance of detecting pedestrian crossings in four classes could be compared to related work. With an accuracy ranging from 0.957 to 0.988, recall ranging from 0.759 to 0.903, and precision ranging from 0.846 to 0.932, our approach achieved similar or better performances to those in related works. Apart from the above advantages, the approach also had some limitations. The major one was the lack of control over the quality of the road markings. Another obstacle was the lack of an annotated dataset that was harmonized with iRAP attribute definitions. A larger dataset would also lead to fewer false detections. Finally, shadows on satellite imagery could be a serious obstacle for object detection, especially in urban areas.

According to the presented limitations of this research, future research on road attribute detection should include annotations of larger iRAP-harmonized datasets, which would allow greater efficiency of the Yolo detector. It would also be possible to include more spectral bands in the process and to evaluate potential improvements over the three visible bands used in this study. In addition, there are over 60 road attributes defined by iRAP that have not yet been studied for automatic detection. Therefore, there is still much room for the exploration of approaches for the automatic detection of these attributes.

Chapter 5

Joint Discussion

Among other things, road safety is influenced by a whole series of road attributes. Such attributes are clearly defined by different road assessment programs [20]. One of the most used road assessment programs in the world is iRAP, which defines 78 road attributes that affect road safety [26]. Most of the defined attributes have a spatial component, i.e., it is necessary to know their position in space exactly. Currently, such attributes are determined from georeferenced video, with trained experts simultaneously watching the video and coding noticed attributes [24]. This approach has several disadvantages. First, the coding process is not consistent. Although the attributes are clearly defined, many of the parameters defining the attribute refer to spatial quantities (distances, heights, angles of inclination, etc.). All attributes defined in this way are subject to coding error, i.e., two different coders will characterise the same object in different ways. This leads to a loss of consistency in attribute coding, even if two or more coders have worked on the same road segments. In particular, the problem of consistency at the level of the entire road network, on which more than one coder is certainly working, stands out.

Another disadvantage of coding road attributes from georeferenced video is that the georeferenced video is collected from a car-mounted camera [24]. Therefore, it is necessary for the coder to review all the video material and maintain the same level of concentration throughout the duration of the video. Such a process is not only associated with a loss of consistency due to coder fatigue but is also time consuming. The aforementioned disadvantages must be reduced and, if possible, completely eliminated in order to ensure a consistent and rapid road safety assessment. At the same time, a satisfactory level of accuracy should be maintained.

With the recent development of computer resources, the increasingly widespread use of complex machine learning models such as deep learning has become possible, opening up a whole spectrum of new applications. One of the most widely used deep learning models is CNN. The aforementioned networks achieve remarkable results in the processes of image, classification, image segmentation, but also object detection in images.

Within thesis, recent object detection methods have been applied to three different geospatial datasets (UAV video, Lidar and VHR satellite imagery) to address the shortcomings of the current road attributes coding approach described above.

It is important to note that not all road attributes can be recognised from a single geospatial data source. Rather, an individual approach to each of the road attributes is required to determine from which data source an attribute can be uniquely recognised. Furthermore, the integration of different spatial data sources with current object detection methods plays a key role in road attributes high accuracy in the encoding of road attributes. In view of this, this work focuses on six road attributes: traffic flow on a given road segment, RSS – O which refers to the 13 different classes of possible roadside objects, RSS – D which refers to the distance between detected roadside objects and the roadside, the presence of school zones on a given road segment, the detection of four different classes of pedestrian crossings and the detection of divided carriageways. The aforementioned attributes were selected to prove the applicability of UAV video, Lidar data and VHR satellite imagery for determining road attributes.

To tackle the traffic flow rate attribute, which is one of the road attributes defined in iRAP, a novel framework has been developed. This framework proposes an innovative approach for estimating traffic flow parameters by combining UAVs, high-precision GNSS technology, state-of-the-art object detection and spatial operations to produce highly accurate traffic flow analyses. The framework focuses on accuracy rather than inference time and uses the Faster R-CNN object detection network trained on the COCO dataset. With a modest input of 160 images for fine-tuning, the system achieved a high accuracy of 0.988 and a recall of 0.994. Compared to similar studies, Ke et al. (2019) [67] achieved a slightly better precision (0.995) but a lower recall (0.957) using a much larger dataset of 18 000 images for training. The higher recall rate in proposed framework is crucial as it represents the ratio between detected and actual vehicles on the road, which directly affects the accuracy of the determination of traffic flow parameters. The robustness of Faster R-CNN allows for fine-tuning and reduces training time by setting scales and aspect ratios of anchors based on available hardware capacity. For example, scales were set to 3, 7 and 11 and aspect ratios to 2, 4 and 6 to find all ground truth vehicles. In addition, a suitable image interval is crucial for the determination of TMS and SMS. Ke et al. (2019) [67] arbitrarily selected a frame interval of 5 in a 25 Hz video without detailed analysis. Proposed framework, on the other hand, bases the selection of the frame interval on the MAPE values of the vehicle speed and chooses a frame interval of 12, which corresponds to half a second, achieving a satisfactory MAPE rate of slightly less than 1%. Furthermore, this framework facilitates the analysis of individual vehicle trajectories and highlights the differences in speed and trajectory between vehicles under different traffic flow conditions. Furthermore, the incorporation of GNSS technology significantly improves the accuracy of

spatial resolution, which is crucial for estimating microscopic parameters. This is a remarkable advance over other studies that did not provide microscopic parameters [66]. A unique advantage of this framework is the possibility to analyse the parameters of traffic flow per lane, thanks to the spatial analysis enabled by the bounding boxes of detected vehicles and motorway lanes. This approach has not yet been used in related literature [66,67,69,70], highlighting the innovative aspect of proposed framework. However, the framework has its limitations. The flight time of the UAV is limited by battery capacity, which affects the observation time over the area of interest. This problem can be partially mitigated by multiple observations at different times of the day or by having multiple batteries. Although UAVs are a more cost-effective solution for traffic flow analysis compared to traditional methods such as inductive loops or pneumatic tube systems, the manual input required in the area of vehicle detection is a shortcoming.

Next two road attributes defined by iRAP, RSS – O and RSS – D, are addressed by a novel framework which is based on Yolo v5 object detector and Lidar point clouds. Unlike previous methods found in [104,105,111,112], which relied on cameras mounted on vehicles, this framework employs Lidar data i.e., point clouds, enhancing the spatial consideration in detecting roadside objects and their distances from the road edge. The limitation of camera-based approaches in determining the RSS – D attribute is well illustrated by the attempt of Jan Z. et al. (2019) [111], which lacked distance determination evaluation. The utilization of Lidar data enables multiple viewing angles of road segments, which is crucial for accurate road attributes classification. Few studies [108,109,113] have employed Lidar data for similar tasks. Notably, Zhong M. et al. (2019) [109] evaluated distance determination accuracy, specifically for two classes: poles and trees. Proposed framework significantly outperformed their findings, reducing the average error distance to 0.07 m and 0.38 m for poles and trees respectively. Moreover, this framework effortlessly and accurately detects iRAP RSS – O classes defined by spatial parameters; a task found challenging in camera-based methods as noted in [104,105,111,112]. Unlike the manual coding method [108,109], which is prone to errors, proposed fully automated framework ensures consistent classification of RSS – O and RSS – D classes, representing a substantial advancement in this field. One of the unique features of this framework is detection of all 13 RSS – O classes existing in Croatia, extending beyond the limited scope of related works [104,109,111,112] that focused on “easy-to-detect” classes. This broadened scope is due to a larger dataset, enhancing the overall detection and classification process. In the evaluation phase, despite dataset imbalances, proposed framework scored high in precision, recall, and AP metrics. Classes like “safety barrier concrete” and “safety barrier

metal" achieved an AP of >0.92, showcasing the framework's effectiveness. However, classes like rockface and rock, with unclear boundaries, posed detection challenges, similar to the ambiguously defined "pole" class in the iRAP Manual Guide [26]. A proposed solution is subclassification of the "pole" class, expected to boost the AP of this class as per iRAP standards, ultimately achieving a final classification accuracy of 85.1% for RSS – O classes. Spatial evaluation using RMSE revealed that larger classes tend to have larger error values, affecting the RSS – D attribute classification, which stood at an accuracy of 85.55% for road segments. However, the framework presents a few challenges. The cost of Lidar systems remains considerably higher compared to camera systems for video-based coding of road segments. Although the price is expected to drop with the increasing use of Lidar in autonomous driving [128–130], the cost remains a barrier. Moreover, the size of the dataset is another limitation, with the specificity of iRAP RSS – O classes requiring larger datasets for higher accuracy. This work, built on manually annotated objects, signifies the necessity of a larger, balanced dataset to overcome data imbalances and unrecognized objects, which in turn would enhance the accuracy and road safety measures, pushing the performance of the entire process a notch higher.

Finally, the school zone, four types of pedestrian crossings and divided carriageways were detected by combining satellite imagery with the Yolo v5 object detector. This approach demonstrated highly efficient detection of divided carriageways, establishing potential benchmarks for future research. Although divided carriageways have not been the subject of previous studies, they have a significant impact on road safety according to the KPIs defined by EC. In terms of pedestrian crossing detection, while much work has used various data sources such as UAVs, vehicle-mounted lidars or georeferenced videos, this approach is moving towards the more cost-effective high-resolution satellite imagery, which is particularly useful for linear objects such as roads. Unlike previous research which are focused on a single type of pedestrian crossing, this approach focuses on the four classes of pedestrian crossings defined by iRAP. This is done by converting the roads from the global CRS to local road-oriented CRS, with the Y-axis of the images corresponding to the road direction. Compared to similar studies, this approach achieves significant results in terms of accuracy, apart from the novelty of detecting school zones and divided carriageways. For example, Prakash et al. (2015) [154] had a similar way of processing road segments and managed to get a precision of 0.89 for identifying one type of pedestrian crossings, while proposed approach achieved precision between 0.846 to 0.932 but for four different types. They had a recall value of 0.63, while Yolo v5 detector used in this research got a recall value between 0.759 and 0.903. On the other side,

Berriel et al. (2017) [156] used a method called binary image classification with ConvNet on a bunch of pedestrian crossing images from Google Maps and got a binary accuracy of 0.97. In this research, the accuracy ranged from 0.957 to 0.988 for four classes. Also, Ghilardi et al. (2018) [156] used an SVM classifier and got an accuracy of 0.946, and Chen et al. (2021) [158] used YOLO v3 deep learning detector and got an accuracy of 0.946. All these numbers show that proposed approach is as good as other deep learning methods like ConvNet and Yolo v3. This was kind of expected because Yolo v5 detector was used, which is known to be better than its earlier versions [164]. However, proposed approach also had some drawbacks. The Yolo v5 detector trained in proposed research sometimes produced false or misleading results. For example, good results were achieved in detecting divided carriageways, but sometimes the detected objects were not on the inspected road. This is a big problem, especially when the inspected road was an undivided carriageway. Although the detection was correct, it was misleading. A possible solution could be to only look at smaller road segments around the inspected road and to divide the roads into different classes to choose the correct segment size. In addition, the quality of detection depended heavily on how good the road markings were, which in turn depended on road maintenance, which cannot be controlled. Although there are laws on road maintenance, they are not always followed. Stricter enforcement of the laws could help. Another problem were many false detections because of similar patterns between school zones, pedestrian crossings, and divided carriageways with other roadside objects. This was mainly because training dataset used in this research was not large enough for the detector to learn properly. This is a common problem when using deep learning for road attribute detection, as there is no large dataset that matches the road attributes defined by iRAP. So, there is a need to create a larger dataset in the future to make better use of deep learning for this task. Finally, in cities, pedestrian crossings may not be detected due to shadows. However, as satellite technology improves and more satellites are launched each year, this problem could be reduced. More satellites could help avoid shadows by choosing when to observe the area depending on the time of day.

While all detected road attributes are based on object detection methods, there are some differences specific to each attribute. First, determining traffic flow rate is a unique task compared to other recognised attributes. It is a complex process involving four different segments. While in the detection of other road attributes, object detection directly recognises the road attributes and their classes, in the determination of traffic flow rate it is only one of four segments in the workflow. Therefore, the object detection process is useless if it is not accompanied by a suitable tracking algorithm. The determination of the traffic flow rate

requires a high level of recall of the object detection. This enables the tracking algorithm to track the detected object in successive images. Therefore, Faster R-CNN is selected as a suitable object detector that offers the highest accuracy compared to other object detectors [165]. One of the shortcomings of Faster R-CNN is its slower inference time compared to other object detectors. This means that the detector cannot work in real time. Considering that traffic flow determination is a four-step process that requires video images alignment, the proposed framework is not suitable for a real-time process.

The determination of RSS – O and RSS – D in comparison to other detected road attributes is required for certain pre-processing techniques. While other detected attributes are detected from images, these two attributes have point clouds as input data set. Therefore, the point clouds need to be converted into images of road cross sections. There is a critical length of road sections defined by iRAP as 100 metres [26]. Considering that it is usually not possible to create clear images of 100-metre road cross-sections, a length of 10 metres is assumed in this framework. Furthermore, the roads studied are usually very long, which can be a time-consuming process if each image represents a 10-metre road segment. Therefore, the inference time is crucial to automate the detection process. Considering this, Yolo v5 is chosen as a suitable object detector because it offers a fast inference time [166].

The same object detector is used to detect the other road attributes from satellite imagery for the same reason. The inference time of Yolo v5 allows near real-time detection, which is much faster than manual coding by trained experts. Compared to the determination of traffic flow, RSS – O and RSS – D, satellite imagery also provides continuous observation of the earth, making it much easier to track changes in road attributes over time. The used Pleiades Neo 3 satellite observes every point on the earth twice a day. Compared to UAV videos and mobile Lidar datasets, which must be done manually for each observation, this is a major advantage. Particularly noteworthy is the acquisition of point clouds from mobile Lidar devices, which is expensive by default.

Chapter 6
Conclusion

The aim of this dissertation was to analyse the possibilities of recent object detection methods on three different sources of geospatial data road attributes process of road attribute detection. UAV video was tested in combination with the Faster R-CNN object detector to determine traffic flow parameters. The traffic flow parameters include the traffic flow rate which is a road attribute defined by iRAP. The Lidar point cloud was used as a geospatial dataset in combination with the Yolo v5 object detector for the determination of the RSS – O and RSS – D road attributes. Third geospatial dataset is satellite imagery used in combination with the Yolo v5 detector to automate the detection of school zones, four types of pedestrian crossings and divided carriageways. In this paper, a framework for automatic calculation of traffic flow parameters by applying the object detection method to UAV videos was proposed. Based on the theoretical knowledge and analysis of the research results, the application of the object detection method to mobile Lidar data for roadside feature detection was proposed. In addition, the application of the object detection method to high resolution satellite imagery was proposed for road attributes for road infrastructure safety assessment. This chapter summarises the conclusions and results of the research and aligns them with the objectives set out in section 1.8. The conclusions and responses to the hypotheses are drawn from the results of this thesis.

- **Object detection method on UAV aerial video can be utilized for calculation of traffic flow parameters.**

This research hypothesis is addressed in Chapter 2, where a new low-cost framework for accurately computing traffic flow parameters from UAV aerial videos was proposed. The framework is based on the state-of-the-art Faster R-CNN object detector. Three macroscopic and four microscopic traffic flow parameters were calculated using the proposed system. One of the three macroscopic parameters calculated is the traffic flow rate, which is also a road attribute defined by iRAP. The combination of a high-precision object detector with a suitable tracking algorithm on UAV video images can lead to an accurate calculation of the traffic flow parameters.

- **Object detection method on mobile Lidar data can be used for roadside feature detection and distance measurements.**

The second research hypothesis is addressed in Chapter 3, where a novel framework is proposed for detection of RSS – O and RSS – D iRAP road attributes. The system is based on road cross-section images generated from mobile Lidar point clouds. In addition, the framework includes the Yolo v5 object detector with fast inference time, which aims to automate the process of roadside feature detection efficiently and quickly. The framework enables the detection of 13 classes of roadside objects defined by iRAP. Point clouds allow a spatial view of the detected roadside features, leading to an accurate determination of the RSS – D attribute.

- **Combination of high-resolution satellite imagery with object detection method can be utilized for automated detection of road attributes to support road infrastructure safety assessment.**

The third research hypothesis is addressed in Chapter 4, which proposes the combination of VHR satellite imagery with the Yolo v5 object detector for automatic road attributes detection. The proposed approach achieves high accuracy in detecting four different types of pedestrian crossings, school zones and divided carriageways. The combination of satellite imagery with the Yolo v5 object detector enables automatic and fast detection of the specified road attributes with high accuracy. All mentioned road attributes are clearly defined by the iRAP, which is used in many European countries to assess road safety. Also, the detection of divided carriageways is important for the EC KPIs indicating the level of road safety until a uniform road assessment framework is developed in all EU countries. Therefore, the automated detection of these road attributes directly contributes to supporting the road infrastructure safety assessment process.

This work represents a contribution to the development and implementation of AiRAP, which aims to combine advanced technologies with various data sources to improve the road assessment process. The combination of recent object detectors with UAV, Lidar or VHR satellite imagery enables the automatic and efficient detection of a wide range of road attributes impact on road safety. It is important to note that each of the attributes detected in this work requires a specific approach. Therefore, appropriate data source has to be chosen for specific attribute detection. For those road attributes that have a spatial component, expertise in

geospatial data and traffic and transport science is required to decide which geospatial data source is appropriate for a particular road attribute.

There is a lot of potential future work to improve the efficiency of road attributes collection. In addition to the six road attributes processed in this work, there are 72 other road attributes defined by iRAP. Each of these attributes can be studied in detail and based on this, a suitable data source can be proposed. It is also important to note that object detection is not the only deep learning task that can be performed to support road safety assessment. There are a variety of tasks, such as image classification and semantic segmentation, that can be used in road safety assessment.

Bibliography

1. Tilburg, C.R. van (Cornelis) *Traffic and Congestion in the Roman Empire*; Routledge, 2007; ISBN 9780415512619.
2. Datta, A. Migration and Urban Living in Less Developed Countries. *International Encyclopedia of Housing and Home* **2012**, 294–297, doi:10.1016/B978-0-08-047163-1.00037-0.
3. National Highway Traffic Safety Administration *Traffic Safety Facts, State Traffic Data: 2021 Data*; Washington D.C., 2023.
4. Young, W.; Sobhani, A.; Lenné, M.G.; Sarvi, M. Simulation of Safety: A Review of the State of the Art in Road Safety Simulation Modelling. *Accid Anal Prev* **2014**, *66*, 89–103, doi: 10.1016/J.AAP.2014.01.008.
5. World Health Organization Road Traffic Injuries Available online: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed on 13 September 2023).
6. World Health Organization *STRENGTHENING ROAD SAFETY LEGISLATION: A Toolkit for Road Safety Legislation Workshops*; 2014.
7. United Nations (UN) Road Safety Strategy for the United Nations System and Its Personnel A Partnership for Safer Journeys Available online: https://www.un.org/sites/un2.un.org/files/2020/09/road_safety_strategy_booklet.pdf (accessed on 25 April 2022).
8. World Health Organization *GLOBAL STATUS REPORT ON ROAD SAFETY 2018 SUMMARY*; 2018.
9. European Road Safety Observatory Road Safety Statistics in the EU Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_accident_fatalities_-_statistics_by_type_of_vehicle&oldid=583880#Passengers_or_drivers_of_passenger_cars_accounted_for_44.6_.25_of_persons_killed_in_2021.2C_while_pedestrians_accounted_for_18.1_.25 (accessed on 13 September 2023).
10. European Road Safety Observatory *National Road Safety Profile-Croatia*; 2021.
11. American Association of State Highway and Transportation Officials *Highway Safety Manual (1st Edition.)*; Washington D.C., 2010.
12. Nason, N.R. National Highway Traffic Safety Administration Consumer Information; New Car Assessment Program. *Fed Regist* **2008**, *73*, 40016–40050.
13. European New Car Programme Assessment Protocol - Overall Rating. **2012**.
14. Wang, Y.; Zhang, W. Analysis of Roadway and Environmental Factors Affecting Traffic Crash Severities. *Transportation Research Procedia* **2017**, *25*, 2119–2125, doi: 10.1016/J.TRPRO.2017.05.407.

15. Wong, S.C.; Sze, N.N.; Li, Y.C. Contributory Factors to Traffic Crashes at Signalized Intersections in Hong Kong. *Accid Anal Prev* **2007**, *39*, 1107–1113, doi: 10.1016/J.AAP.2007.02.009.

16. European Commission *Europe on the Move - Sustainable Mobility for Europe: Safe, Connected, and Clean*; Brussels, 2018.

17. European Commission *Road Safety Study for the Interim Evaluation of Policy Orientations on Road Safety 2011-2020*; 2015.

18. Breen Consulting, J. *Preparatory Work for an EU Road Safety Strategy 2020-2030 Final Report*; 2018.

19. Hove, E. *The Handbook of Road Safety Measures*; Elvik, R., Høye, A., Vaa, T., Sørensen, M., Eds.; Emerald Group Publishing Limited, 2009; ISBN 978-1-84855-250-0.

20. European Commision *Next Steps towards "Vision Zero";* Brussels, 2021.

21. European Parliament *Directive 2008/96/EC*; 2008.

22. European Commission *Directive 2019/1936*; Official Journal of the European Union: European Union, 2019.

23. International Road Assessment Programme *IRAP Methodology Fact Sheet - Overview*; 2014.

24. International Road Assessment Programme *IRAP Survey Manual*; 2022.

25. International Road Assessment Programme *IRAP Methodology Fact Sheet - Road Attributes*; 2015.

26. International Road Assessment Programme IRAP Coding Manual Drive on the Right Edition Available online: www.irap.org/specifications. (accessed on 22 April 2022).

27. International Road Assessment Programme *Process for AiRAP Attribute Accreditation*; 2022.

28. Alpaydin, E. *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*; The MIT Press: Cambridge, Massachusetts, 2014; ISBN 9780262028189.

29. Deng, L.; Yu, D. Deep Learning Methods and Applications. *Foundations and Trends in Signal Processing* **2013**, *7*, doi:10.1561/2000000039.

30. Ding, B.; Qian, H.; Zhou, J. Activation Functions and Their Characteristics in Deep Neural Networks. *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018* **2018**, 1836–1841, doi:10.1109/CCDC.2018.8407425.

31. Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A Comprehensive Survey of Loss Functions in Machine Learning. *Annals of Data Science* **2022**, *9*, 187–212, doi:10.1007/S40745-020-00253-5/TABLES/8.

32. Dogo, E.M.; Afolabi, O.J.; Nwulu, N.I.; Twala, B.; Aigbavboa, C.O. A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural

Networks. *Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018* **2018**, 92–99, doi:10.1109/CTEMS.2018.8769211.

33. Zou, X. A Review of Object Detection Techniques. *Proceedings - 2019 International Conference on Smart Grid and Electrical Automation, ICSGEA 2019* **2019**, 251–254, doi:10.1109/ICSGEA.2019.00065.

34. Ballard, D.H. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognit* **1981**, *13*, 111–122, doi:10.1016/0031-3203(81)90009-1.

35. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int J Comput Vis* **2004**, *60*, 91–110, doi:10.1023/B: VISI.0000029664.99615.94.

36. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2015**, *2016-December*, 779–788, doi:10.48550/arxiv.1506.02640.

37. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2016**, *2017-January*, 6517–6525, doi:10.1109/CVPR.2017.690.

38. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. **2018**.

39. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. **2020**.

40. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. **2021**, *5*, 12.

41. Xu, S.; Wang, X.; Lv, W.; Chang, Q.; Cui, C.; Deng, K.; Wang, G.; Dang, Q.; Wei, S.; Du, Y.; et al. PP-YOLOE: An Evolved Version of YOLO. **2022**.

42. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2015**, *9905 LNCS*, 21–37, doi:10.1007/978-3-319-46448-0_2.

43. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans Pattern Anal Mach Intell* **2017**, *42*, 318–327, doi:10.1109/TPAMI.2018.2858826.

44. Law, H.; Deng, J. CornerNet: Detecting Objects as Paired Keypoints. *Int J Comput Vis* **2018**, *128*, 642–656, doi:10.1007/s11263-019-01204-1.

45. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. *Proceedings of the IEEE International Conference on Computer Vision* **2019**, *2019-October*, 9626–9635, doi:10.1109/ICCV.2019.00972.

46. Zhang, H.; Cloutier, R.S. Review on One-Stage Object Detection Based on Deep Learning. *EAI Endorsed Transactions* **2022**, *7*, doi:10.4108/eai.9-6-2022.174181.

47. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2013**, 580–587, doi:10.1109/CVPR.2014.81.

48. Girshick, R. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision* **2015**, 1440–1448, doi:10.1109/ICCV.2015.169.

49. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell* **2017**, *39*, 1137–1149, doi:10.1109/TPAMI.2016.2577031.

50. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2014**, *8691*, 346–361, doi:10.1007/978-3-319-10578-9_23.

51. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* **2017**, *42*, 386–397, doi:10.1109/TPAMI.2018.2844175.

52. Uijlings, J.R.R.; Van De Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective Search for Object Recognition. *Int J Comput Vis* **2013**, *104*, 154–171, doi:10.1007/S11263-013-0620-5.

53. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; NanoCode012; Kwon, Y.; Michael, K.; TaoXie; Fang, J.; imyhxy; et al. Ultralytics/Yolov5: V7.0 - YOLOv5 SOTA Realtime Instance Segmentation. **2022**, doi:10.5281/ZENODO.7347926.

54. European Automobile Manufacturers Association *Vehicles in Use Europe 2019*; 2019.

55. Mathew, T. V.; Rao, K.V.K. Fundamental Relations of Traffic Flow. *Introduction to Transportation Engineering* **2006**, *1*, 1–8.

56. Hoogendoorn, S.; Knoop, V. Traffic Flow Theory and Modelling. *The Transport System and Transport Policy: An Introduction* **2012**, 125–159.

57. Leduc, G. Road Traffic Data: Collection Methods and Applications. *JRC Tehnical Notes* **2008**.

58. Handscombe, J.; Yu, H.Q. Low-Cost and Data Anonymised City Traffic Flow. *Sensors* **2019**, doi:10.3390/s19020347.

59. Martinez, A.P. Freight Traffic Data in the City of Eindhoven Available online: https://pure.tue.nl/ws/portalfiles/portal/47039665/801382-1.pdf (accessed on 15 July 2022).

60. Kanistras, K.; Martins, G.; Rutherford, M.J.; Valavanis, K.P. A Survey of Unmanned Aerial Vehicles (UAVs) for Traffic Monitoring. *2013 International Conference on Unmanned Aircraft Systems, ICUAS 2013 - Conference Proceedings* **2013**, 221–234, doi:10.1109/ICUAS.2013.6564694.

61. SESAR Joint Undertaking European Drones Outlook Study. *SESAR* **2016**.

62. El-Sayed, H.; Chaqfa, M.; Zeadally, S.; Puthal, A.D. A Traffic-Aware Approach for Enabling Unmanned Aerial Vehicles (UAVs) in Smart City Scenarios. *IEEE Access* **2019**, *7*, 86297–86305.

63. Menouar, H.; Güvenc, I.; Akkaya, K.; Uluagac, A.S.; Kadri, A.; Tuncer, A. UAV-Enabled Intelligent Transportation Systems for the Smart City: Applications and Challenges. *IEEE Communications Magazine* **2017**, 22–28.

64. Ghazzai, H.; Menouar, H.; Kadri, A. On the Placement of UAV Docking Stations for Future Intelligent Transportation Systems. In Proceedings of the IEEE 85th Vehicular Technology Conference (VTC Spring); 2017.

65. Beg, A.; Qureshi, A.R.; Sheltami, T.; Yasar, A. UAV-Enabled Intelligent Traffic Policing and Emergency Response Handling System for the Smart City. **2020**.

66. Khan, M.A.; Ectors, W.; Bellemans, T.; Janssens, D.; Wets, G. Unmanned Aerial Vehicle-Based Traffic Analysis: A Case Study for Shockwave Identification and Flow Parameters Estimation at Signalized Intersections. *Remote Sensing 2018, Vol. 10, Page 458* **2018**, *10*, 458, doi:10.3390/RS10030458.

67. Ke, R.; Li, Z.; Tang, J.; Pan, Z.; Wang, Y. Real-Time Traffic Flow Parameter Estimation from UAV Video Based on Ensemble Classifier and Optical Flow. *IEEE Transactions on Intelligent Transportation Systems* **2019**, *20*, 54–64, doi:10.1109/TITS.2018.2797697.

68. Chen, X.; Li, Z.; Yang, Y.; Qi, L.; Ke, R. High-Resolution Vehicle Trajectory Extraction and Denoising from Aerial Videos. *IEEE Transactions on Intelligent Transportation Systems* **2021**, *22*, 3190–3202, doi:10.1109/TITS.2020.3003782.

69. Fedorov, A.; Nikolskaia, K.; Ivanov, S.; Shepelev, V.; Minbaleev, A. Traffic Flow Estimation with Data from a Video Surveillance Camera. *J Big Data* **2019**, doi:10.1186/s40537-019-0234-z.

70. Wang, L.; Chen, F.; Yin, H. Detecting and Tracking Vehicles in Traffic by Unmanned Aerial Vehicles. *Autom Constr* **2016**, *72*, 294–308, doi: 10.1016/j.autcon.2016.05.008.

71. Ministry of the Sea; Transport and Infrastructure *Transport Development Strategy of the Republic of Croatia (2017 - 2030) - Development of Sectoral Transport Strategies*; 2013; ISBN 9788578110796.

72. Transportation Research Board *Highway Capacity Manual: A Guide for Multimodal Mobility Analysis.*; 2016; ISBN 978-0-309-36997-8.

73. Awad, A.I.; Hassaballah, M. *Studies in Computational Intelligence 630 Image Feature Detectors and Descriptors Foundations and Applications*; 2016; ISBN 9783319288529.

74. Pieropan, A.; Björkman, M.; Bergström, N.; Kragic, D. Feature Descriptors for Tracking by Detection: A Benchmark. **2016**.

75. Aglave, P.; Kolkure, V.S. Implementation of High-Performance Feature Extraction Method Using Oriented Fast and Rotated Brief Algorithm. *Int J Res Eng Technol* **2015**, *04*, 394–397, doi:10.15623/ijret.2015.0402052.

76. Zhou, Q.; Li, X. STN-Homography: Estimate Homography Parameters Directly. **2019**, 0–8.

77. Chum, O.; Matas, J.; Kittler, J. Locally Optimized RANSAC. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2003**, *2781*, 236–243, doi:10.1007/978-3-540-45243-0_31.

78. Fischler, M.A.; Bolles, R.C. *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*; Morgan Kaufmann Publishers, Inc., 1987.

79. Ma, Y.; Soatto, S.; Košecká, J.; Sastry, S.S. Reconstruction from Two Uncalibrated Views. In *An Invitation to 3-D Vision*; 2004; pp. 171–227.

80. Yague-Martinez, N.; De Zan, F.; Prats-Iraola, P. Coregistration of Interferometric Stacks of Sentinel-1 TOPS Data. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 1002–1006, doi:10.1109/LGRS.2017.2691398.

81. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection with Deep Learning: A Review. *IEEE Trans Neural Netw Learn Syst* **2019**, 1–21, doi:10.1109/TNNLS.2018.2876865.

82. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A Survey of Deep Learning-Based Object Detection. *IEEE Access* **2019**, *7*, 128837–128868, doi:10.1109/access.2019.2939201.

83. Li, B.; Liu, Y.; Wang, X. Gradient Harmonized Single-Stage Detector. *Proceedings of the AAAI Conference on Artificial Intelligence* **2019**, *33*, 8577–8584, doi:10.1609/aaai.v33i01.33018577.

84. Sun, B.; Xu, Y.; Li, C.; Yu, J. Analysis of the Impact of Google Maps' Level on Object Detection. In Proceedings of the IEEE Geoscience and Remote Sensing Society; Yokohama, Japan, 2019; pp. 1248–1251.

85. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep Learning for Generic Object Detection: A Survey. *Int J Comput Vis* **2019**, doi:10.1007/s11263-019-01247-4.

86. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2014**, *8693 LNCS*, 740–755, doi:10.1007/978-3-319-10602-1_48.

87. Wang, J.; Chen, K.; Yang, S.; Loy, C.C.; Lin, D. Region Proposal by Guided Anchoring. **2019**.

88. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016**, *2016-Decem*, 770–778, doi:10.1109/CVPR.2016.90.

89. Mohammad, H.; Md Nasair, S. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process* **2015**, *5*, 01–11, doi:10.5121/ijdkp.2015.5201.

90. Federal Geographical Data Committee Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy. *National Spatial Data Infrastructure* **1998**, 28.

91. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple Online and Realtime Tracking. *Proceedings - International Conference on Image Processing, ICIP* **2016**, *2016-Augus*, 3464–3468, doi:10.1109/ICIP.2016.7533003.

92. Jordahl, K. GeoPandas Documentation 2016.

93. Guo, J.; Huang, W.; Williams, B.M. Adaptive Kalman Filter Approach for Stochastic Short-Term Traffic Flow Rate Prediction and Uncertainty Quantification. *Transp Res Part C Emerg Technol* **2014**, *43*, 50–64, doi: 10.1016/j.trc.2014.02.006.

94. Turner, S.M.; Eisele, W.L.; Benz, R.J.; Douglas, J. Travel Time Data Collection Handbook. *Federal Highway Administration, USA.* **1998**, *3*, 293.

95. Knoop, V.; Hoogendoorn, S.P.; Van Zuylen, H. Empirical Differences between Time Mean Speed and Space Mean Speed. *Traffic and Granular Flow 2007* **2009**, 351–356, doi:10.1007/978-3-540-77074-9-36.

96. Dadić, I.; Kos, G.; Ševrović, M. *Traffic Flow Theory*; 2014.

97. Luttinen, R.T. Statistical Analysis of Vehicle Time Headways. PhD thesis, University of Technology Lahti Center, 1996.

98. Passmore, J.; Yon, Y.; Mikkelsen, B. Progress in Reducing Road-Traffic Injuries in the WHO European Region. *Lancet Public Health* **2019**, *4*, e272–e273, doi:10.1016/S2468-2667(19)30074-X.

99. International Road Assesment Programme (iRAP) IRAP Star Rating and Investment Plan Implementation Support Guide Available online: https://irap.org/2021/06/irap-star-rating-and-investment-plan-manual-version-1-0-is-now-available-for-download/ (accessed on 22 April 2022).

100. European Union REGULATION (EU) No 1315/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - Consolidated Text Available online: https://eur-lex.europa.eu/eli/reg/2013/1315/oj (accessed on 26 April 2022).

101. Brkić, I.; Miler, M.; Ševrović, M.; Medak, D. An Analytical Framework for Accurate Traffic Flow Parameter Calculation from UAV Aerial Videos. *Remote Sensing 2020, Vol. 12, Page 3844* **2020**, *12*, 3844, doi:10.3390/RS12223844.

102. Kacan, M.; Oršic, M.; Šegvic, S.; Ševrovic, M. Multi-Task Learning for IRAP Attribute Classification and Road Safety Assessment. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020* **2020**, doi:10.1109/ITSC45102.2020.9294305.

103. Graf, S.; Pagany, R.; Dorner, W.; Weigold, A. Georeferencing of Road Infrastructure from Photographs Using Computer Vision and Deep Learning for Road Safety Applications. *GISTAM 2019 - Proceedings of the 5th International Conference on Geographical Information Systems Theory, Applications and Management* **2019**, 71–76, doi:10.5220/0007706800710076.

104. Sanjeewani, P.; Verma, B. Optimization of Fully Convolutional Network for Road Safety Attribute Detection. *IEEE Access* **2021**, *9*, 120525–120536, doi:10.1109/ACCESS.2021.3108543.

105. Sanjeewani, P.; Verma, B. Single Class Detection-Based Deep Learning Approach for Identification of Road Safety Attributes. *Neural Comput Appl* **2021**, *33*, 9691–9702, doi:10.1007/S00521-021-05734-Z/TABLES/4.

106. Björnstig, U.; Björnstig, J. Flying Roadside Stones—a Deadly Risk in a Crash. *Traffic Safety Research* **2021**, *1*, 000002, doi:10.55329/tcfh3140.

107. Song, W. Image-Based Roadway Assessment Using Convolutional Neural Image-Based Roadway Assessment Using Convolutional Neural Networks Networks. **2019**, 13, doi: https://doi.org/10.13023/etd.2019.136.

108. Martín-Jiménez, J.A.; Zazo, S.; Arranz Justel, J.J.; Rodríguez-Gonzálvez, P.; González-Aguilera, D. Road Safety Evaluation through Automatic Extraction of Road Horizontal Alignments from Mobile LiDAR System and Inductive Reasoning Based on a Decision Tree. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *146*, 334–346, doi: 10.1016/J.ISPRSJPRS.2018.10.004.

109. Zhong, M.; Verma, B.; Affum, J. Neural Information Processing.; Gedeon, T., Wong, K.W., Lee, M., Eds.; Springer International Publishing: Cham, 2019; Vol. 11954.

110. Ziakopoulos, A.; Yannis, G. A Review of Spatial Approaches in Road Safety. *Accid Anal Prev* **2020**, *135*, doi: 10.1016/j.aap.2019.105323.

111. Jan, Z.; Verma, B.; Affum, J.; Atabak, S.; Moir, L. A Convolutional Neural Network Based Deep Learning Technique for Identifying Road Attributes. *International Conference Image and Vision Computing New Zealand* **2019**, *2018-November*, doi:10.1109/IVCNZ.2018.8634743.

112. Sanjeewani, P.; Verma, B. An Optimisation Technique for the Detection of Safety Attributes Using Roadside Video Data. *International Conference Image and Vision Computing New Zealand* **2020**, *2020-November*, doi:10.1109/IVCNZ51579.2020.9290590.

113. Zhong, M.; Verma, B.; Affirm, J. Point Cloud Classification for Detecting Roadside Safety Attributes and Distances. *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019* **2019**, 1078–1084, doi:10.1109/SSCI44817.2019.9002813.

114. Pubudu Sanjeewani, T.G.; Verma, B. Learning and Analysis of AusRAP Attributes from Digital Video Recording for Road Safety. *International Conference Image and Vision Computing New Zealand* **2019**, *2019-December*, doi:10.1109/IVCNZ48456.2019.8960997.

115. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); New Orleans, LA, June 19, 2017.

116. Ural, S.; Shan, J.; Romero, M.A.; Tarko, A. Road and Roadside Feature Extraction Using Imagery and Lidar Data for Transportation Operation. *ISPRS Annals of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences **2015**, *2*, 239–246, doi:10.5194/ISPRSANNALS-II-3-W4-239-2015.

117. Han, X.; Wang, H.; Lu, J.; Zhao, C. Road Detection Based on the Fusion of Lidar and Image Data. *Int J Adv Robot Syst* **2017**, *14*, doi:10.1177/1729881417738102.

118. Zeybek, M. Extraction of Road Lane Markings from Mobile LiDAR Data: *https://doi.org/10.1177/0361198120981948* **2021**, *2675*, 30–47, doi:10.1177/0361198120981948.

119. Roodaki, H.; Bojnordi, M.N. Compressed Geometric Arrays for Point Cloud Processing. **2021**.

120. Wu, Y.; Wang, Y.; Zhang, S.; Ogai, H. Deep 3D Object Detection Networks Using LiDAR Data: A Review. *IEEE Sens J* **2021**, *21*, 1152–1171, doi:10.1109/JSEN.2020.3020626.

121. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput Sci* **2022**, *199*, 1066–1073, doi: 10.1016/J.PROCS.2022.01.135.

122. Henderson, P.; Ferrari, V. End-to-End Training of Object Class Detectors for Mean Average Precision. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2017**, *10115 LNCS*, 198–213, doi:10.1007/978-3-319-54193-8_13/COVER.

123. Oksuz, K.; Cam, B.C.; Akbas, E.; Kalkan, S. Localization Recall Precision (LRP): A New Performance Metric for Object Detection. In Springer, Cham, 2018 ISBN 978-3-030-01234-2.

124. Davis, J.; Goadrich, M. The Relationship Between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning* **2006**, doi: https://doi.org/10.1145/1143844.1143874.

125. Stitt, J.M.; Svancara, L.K.; Vierling, L.A.; Vierling, K.T. Smartphone LIDAR Can Measure Tree Cavity Dimensions for Wildlife Studies. *Wildl Soc Bull* **2019**, *43*, 159–166, doi:10.1002/WSB.949.

126. Chan, J.; Raghunath, A.; Michaelsen, K.E.; Gollakota, S. Testing a Drop of Liquid Using Smartphone LiDAR. *Proc ACM Interact Mob Wearable Ubiquitous Technol* **2022**, *6*, 27, doi:10.1145/3517256.

127. Tavani, S.; Billi, A.; Corradetti, A.; Mercuri, M.; Bosman, A.; Cuffaro, M.; Seers, T.; Carminati, E. Smartphone Assisted Fieldwork: Towards the Digital Transition of Geoscience Fieldwork Using LiDAR-Equipped IPhones. *Earth Sci Rev* **2022**, *227*, 103969, doi: 10.1016/J.EARSCIREV.2022.103969.

128. Wolcott, R.W.; Eustice, R.M. Robust LIDAR Localization Using Multiresolution Gaussian Mixture Maps for Autonomous Driving: *http://dx.doi.org/10.1177/0278364917696568* **2017**, *36*, 292–319, doi:10.1177/0278364917696568.

129. Li, Y.; Ma, L.; Zhong, Z.; Liu, F.; Chapman, M.A.; Cao, D.; Li, J. Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review. *IEEE Trans Neural Netw Learn Syst* **2021**, *32*, 3412–3432, doi:10.1109/TNNLS.2020.3015992.

130. Li, Y.; Ibanez-Guzman, J. Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems. *IEEE Signal Process Mag* **2020**, *37*, 50–61, doi:10.1109/MSP.2020.2973615.

131. European Commision *Annual Statistical Report on Road Safety in the EU, 2021.*; Brussels, 2022.

132. European Commision *Road Safety Thematic Report–Pedestrians*; Brussels, 2021.

133. Ke, R.; Feng, S.; Cui, Z.; Wang, Y. Advanced Framework for Microscopic and Lane-Level Macroscopic Traffic Parameters Estimation from UAV Video. *IET Intelligent Transport Systems* **2020**, *14*, 724–734, doi:10.1049/IET-ITS.2019.0463.

134. Guan, H.; Lei, X.; Yu, Y.; Zhao, H.; Peng, D.; Marcato Junior, J.; Li, J. Road Marking Extraction in UAV Imagery Using Attentive Capsule Feature Pyramid Network. *International Journal of Applied Earth Observation and Geoinformation* **2022**, *107*, 102677, doi: 10.1016/J.JAG.2022.102677.

135. Bu, T.; Zhu, J.; Ma, T. A UAV Photography–Based Detection Method for Defective Road Marking. *Journal of Performance of Constructed Facilities* **2022**, *36*, 04022035, doi:10.1061/(ASCE)CF.1943-5509.0001748.

136. Biçici, S.; Zeybek, M. An Approach for the Automated Extraction of Road Surface Distress from a UAV-Derived Point Cloud. *Autom Constr* **2021**, *122*, 103475, doi: 10.1016/J.AUTCON.2020.103475.

137. Yuan, Y.; Xiong, Z.; Wang, Q. An Incremental Framework for Video-Based Traffic Sign Detection, Tracking, and Recognition. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *18*, 1918–1929, doi:10.1109/TITS.2016.2614548.

138. Changzhen, X.; Cong, W.; Weixin, M.; Yanmei, S. A Traffic Sign Detection Algorithm Based on Deep Convolutional Neural Network. *2016 IEEE International Conference on Signal and Image Processing, ICSIP 2016* **2017**, 676–679, doi:10.1109/SIPROCESS.2016.7888348.

139. Han, C.; Gao, G.; Zhang, Y. Real-Time Small Traffic Sign Detection with Revised Faster-RCNN. *Multimed Tools Appl* **2019**, *78*, 13263–13278, doi:10.1007/S11042-018-6428-0/FIGURES/8.

140. Neven, D.; De Brabandere, B.; Georgoulis, S.; Proesmans, M.; Van Gool, L. Towards End-to-End Lane Detection: An Instance Segmentation Approach. *IEEE Intelligent Vehicles Symposium, Proceedings* **2018**, *2018-June*, 286–291, doi:10.1109/IVS.2018.8500547.

141. Lee, S.; Kim, J.; Yoon, J.S.; Shin, S.; Bailo, O.; Kim, N.; Lee, T.-H.; Hong, H.S.; Han, S.-H.; Kweon, I.S. VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition. **2017**.

142. Brkić, I.; Miler, M.; Ševrović, M.; Medak, D. Automatic Roadside Feature Detection Based on Lidar Road Cross Section Images. *Sensors 2022, Vol. 22, Page 5510* **2022**, *22*, 5510, doi:10.3390/S22155510.

143. Gargoum, S.; Karstenl, L.; El-Basyouny, K.; Chen, X. Enriching Roadside Safety Assessments Using LiDAR Technology: Disaggregate Collision-Level Data Fusion and Analysis. *Infrastructures 2022, Vol. 7, Page 7* **2022**, *7*, 7, doi:10.3390/INFRASTRUCTURES7010007.

144. De Blasiis, M.R.; Benedetto, A. Di; Fiani, M.; Garozzo, M. Assessing the Effect of Pavement Distresses by Means of LiDAR Technology. *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019* **2019**, 146–153, doi:10.1061/9780784482445.019.

145. Weng, S.; Li, J.; Chen, Y.; Wang, C. Road Traffic Sign Detection and Classification from Mobile LiDAR Point Clouds. *https://doi.org/10.1117/12.2234911* **2016**, *9901*, 55–61, doi:10.1117/12.2234911.

146. Kilani, O.; Gouda, M.; Weiß, J.; El-Basyouny, K. Safety Assessment of Urban Intersection Sight Distance Using Mobile LiDAR Data. *Sustainability 2021, Vol. 13, Page 9259* **2021**, *13*, 9259, doi:10.3390/SU13169259.

147. Gargoum, S.A.; El-Basyouny, K.; Froese, K.; Gadowski, A. A Fully Automated Approach to Extract and Assess Road Cross Sections from Mobile LiDAR Data. *IEEE Transactions on Intelligent Transportation Systems* **2018**, *19*, 3507–3516, doi:10.1109/TITS.2017.2784623.

148. Holgado-Barco, A.; González-Aguilera, D.; Arias-Sanchez, P.; Martinez-Sanchez, J. Semiautomatic Extraction of Road Horizontal Alignment from a Mobile LiDAR System. *Computer-Aided Civil and Infrastructure Engineering* **2015**, *30*, 217–228, doi:10.1111/MICE.12087.

149. Gargoum, S.; El-Basyouny, K.; Sabbagh, J. Automated Extraction of Horizontal Curve Attributes Using LiDAR Data. *Transp Res Rec* **2018**, *2672*, 98–106, doi: 10.1177/0361198118758685/ASSET/IMAGES/LARGE/10.1177_0361198118758685-FIG2.JPEG.

150. Alshehhi, R.; Marpu, P.R. Hierarchical Graph-Based Segmentation for Extracting Road Networks from High-Resolution Satellite Images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *126*, 245–260, doi: 10.1016/J.ISPRSJPRS.2017.02.008.

151. Henry, C.; Azimi, S.M.; Merkle, N. Road Segmentation in SAR Satellite Images with Deep Fully Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters* **2018**, *15*, 1867–1871, doi:10.1109/LGRS.2018.2864342.

152. Buslaev, A.; Seferbekov, S.; Iglovikov, V.; Shvets, A. Fully Convolutional Network for Automatic Road Extraction from Satellite Imagery. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); IEEE, June 2018; pp. 197–1973.

153. Wijnands, J.S.; Zhao, H.; Nice, K.A.; Thompson, J.; Scully, K.; Guo, J.; Stevenson, M. Identifying Safe Intersection Design through Unsupervised Feature Extraction from Satellite Imagery. *Computer-Aided Civil and Infrastructure Engineering* **2021**, *36*, 346–361, doi:10.1111/MICE.12623.

154. Prakash, T.; Comandur, B.; Chang, T.; Elfiky, N.; Kak, A. A Generic Road-Following Framework for Detecting Markings and Objects in Satellite Imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2015**, *8*, 4729–4741, doi:10.1109/JSTARS.2015.2495142.

155. Ghilardi, M.C.; Jacques Junior, J.; Manssour, I. Crosswalk Localization from Low Resolution Satellite Images to Assist Visually Impaired People. *IEEE Comput Graph Appl* **2018**, *38*, 30–46, doi:10.1109/MCG.2016.50.

156. Berriel, R.F.; Lopes, A.T.; De Souza, A.F.; Oliveira-Santos, T. Deep Learning-Based Large-Scale Automatic Satellite Crosswalk Classification. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 1513–1517, doi:10.1109/LGRS.2017.2719863.

157. Ahmetovic, D.; Manduchi, R.; Coughlan, J.M.; Mascetti, S. Mind Your Crossings. *ACM Transactions on Accessible Computing (TACCESS)* **2017**, *9*, doi:10.1145/3046790.

158. Chen, Z.; Luo, R.; Li, J.; Du, J.; Wang, C. U-Net Based Road Area Guidance for Crosswalks Detection from Remote Sensing Images. *https://doi.org/10.1080/07038992.2021.1894915* **2021**, *47*, 83–99, doi:10.1080/07038992.2021.1894915.

159. Croatian Bureau of Statistics *Census of Population, Households and Dwellings 2021 Results by Settlements*; Zagreb, 2022.

160. Airbus Defence and Space *Pléiades Neo*; 2021.

161. Carranza-García, M.; Torres-Mateo, J.; Lara-Benítez, P.; García-Gutiérrez, J. On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data. *Remote Sensing 2021, Vol. 13, Page 89* **2020**, *13*, 89, doi:10.3390/RS13010089.

162. Althnian, A.; AlSaeed, D.; Al-Baity, H.; Samha, A.; Dris, A. Bin; Alzakari, N.; Abou Elwafa, A.; Kurdi, H. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences* **2021**, *11*, 796, doi:10.3390/app11020796.

163. Tzutalin LabelImg.

164. Nepal, U.; Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors* **2022**, *22*, 464, doi:10.3390/S22020464/S1.

165. Kim, C.E.; Dar Oghaz, M.M.; Fajtl, J.; Argyriou, V.; Remagnino, P. A Comparison of Embedded Deep Learning Methods for Person Detection. *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* **2018**, *5*, 459–465, doi:10.5220/0007386304590465.

166. Gupta Pola, V.; Bhavya Vaishnavi, A.; Suraj Karra, S. Comparison of YOLOv3, YOLOv4 and YOLOv5 Performance for Detection of Blood Cells. *International Research Journal of Engineering and Technology* **2021**.

Appendix A

List of iRAP-defined road attributes with associated classes

Attribute	Class
Coder name	-
Coding date	-
Road survey date	-
Image reference	-
Road name	-
Section	-
Distance	-
Length	-
Latitude	-
Longitude	-
Landmark	-
Comments	-
Carriageway label	Carriageway A of a divided road Carriageway B of a divided road Undivided road Carriageway A of a motorcycle facility Carriageway B of a motorcycle facility
Motorcycle observed flow	8+ motorcycles 6 to 7 motorcycles 4 to 5 motorcycles 2 to 3 motorcycles 1 motorcycle None
Bicycle observed flow	8+ bicycles 6 to 7 bicycles 4 to 5 bicycles 2 to 3 bicycles 1 bicycle None
Ped observed flow across	8+ pedestrians 6 to 7 pedestrians 4 to 5 pedestrians 2 to 3 pedestrians 1 pedestrian None
Ped observed flow along – driver-side	8+ pedestrians 6 to 7 pedestrians 4 to 5 pedestrians

Attribute	Class
	2 to 3 pedestrians
	1 pedestrian
	None
Ped observed flow along – passenger-side	8+ pedestrians
	6 to 7 pedestrians
	4 to 5 pedestrians
	2 to 3 pedestrians
	1 pedestrian
	None
Speed limit	≥150km/h
	140km/h
	130km/h
	120km/h
	110km/h
	100km/h
	90km/h
	80km/h
	70km/h
	60km/h
	50km/h
	40km/h
	<30km/h
	≥90mph
	80mph
	70mph
	60mph
	50mph
	40mph
	30mph
	<20mph
Motorcycle speed limit	≥150km/h
	140km/h
	130km/h
	120km/h
	110km/h
	100km/h
	90km/h
	80km/h
	70km/h
	60km/h
	50km/h
	40km/h
	<30km/h
	≥90mph

Attribute	Class
	80mph
	70mph
	60mph
	50mph
	40mph
	30mph
	<20mph
Truck speed limit	≥150km/h
	140km/h
	130km/h
	120km/h
	110km/h
	100km/h
	90km/h
	80km/h
	70km/h
	60km/h
	50km/h
	40km/h
	<30km/h
	≥90mph
	80mph
	70mph
	60mph
	50mph
	40mph
	30mph
	<20mph
Speed differential	Present
	Not present
Speed management	Not present
	Present
Number of lanes	Four or more
	Three
	Three and two
	Two
	Two and one
	One
Lane width	Narrow 0m to <2.75m
	Medium 2.75m to <3.25m
	Wide ≥3.25m
Curvature	Very sharp
	Sharp
	Moderate

Attribute	Class
	Straight or gently curving
Quality of curve	Poor
	Not applicable
	Adequate
Upgrade cost	High
	Medium
	Low
Median type	Centre line
	Wide centre line 0.3m to 1m
	Central hatching >1m
	Continuous central turning lane
	Flexible posts
	Physical median width 0 to <1m
	Physical median width 1 to <5m
	Physical median width 5 to <10m
	Safety barrier – concrete
	Safety barrier – metal
	Safety barrier – motorcycle friendly
	Safety barrier – wire rope
	Physical median width 10 to <20m
	Physical median width ≥20m
	One way
Skid resistance	Unsealed – poor
	Unsealed – adequate
	Sealed – poor
	Sealed – medium
	Sealed – adequate
Road condition	Poor
	Medium
	Good
Vehicle parking	Two side
	One side
	None
Grade	≥10%
	7.5% to <10%
	0% to <7.5%
Roadworks	Major road works
	Minor road works
	No road works
Sight distance	Poor
	Adequate
Delineation	Poor
	Adequate
Street lighting	Not present

Attribute	Class
	Present
Service road	Not present
	Present
Centre line rumble strips	Not present
	Present
Roadside severity – driver- side distance	0 to <1m
	1 to <5m
	5 to <10m
	≥10m
Roadside severity – driver- side object	Cliff
	Tree ≥10cm
	Rigid sign, post or pole ≥10cm
	Unprotected safety barrier end
	Aggressive vertical face
	Upwards slope – roll over
	Deep drainage ditch
	Downwards slope
	Low rigid object ≥20cm high
	Rigid structure or building
	Upwards slope – no roll over
	Semi-rigid structure or building
	Safety barrier – concrete
	Safety barrier – metal
	Safety barrier – wire rope
	Safety barrier – motorcycle friendly
	No object
Roadside severity – passenger-side distance	0 to <1m
	1 to <5m
	5 to <10m
	≥ 10m
Roadside severity – passenger-side object	Cliff
	Tree ≥10cm
	Rigid sign, post or pole ≥10cm
	Unprotected safety barrier end
	Aggressive vertical face
	Upwards slope – roll over
	Deep drainage ditch
	Downwards slope
	Low rigid object ≥20cm high
	Rigid structure or building
	Upwards slope – no roll over
	Semi-rigid structure or building
	Safety barrier – concrete
	Safety barrier – metal

Attribute	Class
	Safety barrier – wire rope
	Safety barrier – motorcycle friendly
	No object
Shoulder rumble strips	Not present
	Present
Paved shoulder – driver-side	None
	Narrow 0m to <1m
	Medium 1m to <2.4m
	Wide ≥2.4m
Paved shoulder – passenger-side	None
	Narrow 0m to <1m
	Medium 1m to <2.4m
	Wide ≥2.4m
Intersection type	4-leg
	4-leg with protected turn lane
	4-leg signalised
	3-leg
	3-leg with protected turn lane
	Mini roundabout
	3-leg signalised
	4-leg signalised with protected turn lane
	3-leg signalised with protected turn lane
	Roundabout
	Railway Crossing – passive
	Merge lane
	Railway Crossing – active
	Median crossing point – informal
	Median crossing point – formal
	None
Intersection quality	Poor
	Adequate
	Not applicable
Intersection channelization	Present
	Not present
Property access points	Commercial access ≥1
	Residential access ≥3
	Residential access <3
	None
Intersecting road volume	≥15,000 vehicles
	10,000 to 15,000 vehicles
	5,000 to 10,000 vehicles
	1,000 to 5,000 vehicles
	100 to 1,000 vehicles
	1 to 100 vehicles

Attribute	Class
	Not applicable
Land use – driver-side	Educational
	Commercial
	Industrial and manufacturing
	Residential
	Farming and agricultural
	Undeveloped areas
	Not Recorded
Land use – passenger- side	Educational
	Commercial
	Industrial and manufacturing
	Residential
	Farming and agricultural
	Undeveloped areas
	Not Recorded
Area type	Urban
	Rural
Pedestrian crossing facilities – inspected road	No facility
	Refuge only
	Marked crossing only
	Raised unmarked crossing
	Marked crossing with refuge
	Raised unmarked crossing with refuge
	Raised marked crossing
	Raised marked crossing with refuge
	Signalised crossing
	Signalised crossing with refuge
	Grade separated facility
Pedestrian crossing facilities quality	Poor
	Adequate
	Not applicable
Pedestrian crossing facilities – side road	No facility
	Refuge only
	Marked crossing only
	Raised unmarked crossing
	Marked crossing with refuge
	Raised unmarked crossing with refuge
	Raised marked crossing
	Raised marked crossing with refuge
	Signalised crossing
	Signalised crossing with refuge
	Grade separated facility
Pedestrian fencing	Not present
	Present

Attribute	Class
Sidewalk – driver-side	None
	Informal path 0m to <1m
	Informal path ≥1m
	Sidewalk 0m to <1m from road
	Sidewalk 1m to <3m from road
	Sidewalk ≥3m from road
	Sidewalk with barrier
Sidewalk – passenger-side	None
	Informal path 0m to <1m
	Informal path ≥1m
	Sidewalk 0m to <1m from road
	Sidewalk 1m to <3m from road
	Sidewalk ≥3 from road
	Sidewalk with barrier
Facilities for motorcycles	None
	Motorcycle lane on roadway
	Motorcycle path – one way
	Motorcycle path – two ways
	Motorcycle path – one way with barrier
	Motorcycle path – two ways with barrier
Facilities for bicycles	None
	Signed shared roadway
	Extra wide outside ≥4.2m
	Dedicated bicycle lane on roadway
	Shared use path
	Segregated bicycle path
	Segregated bicycle path with barrier
School zone warning	No school zone warning (school present)
	School zone static signs or road markings
	School zone flashing beacons
	Not applicable (no school at the location)
School zone crossing supervisor	School zone crossing supervisor not present
	School zone crossing supervisor present at school start and finish times
	Not applicable (no school at the location)

List of Figures

Figure 1.1 Contributing factors to road traffic accidents shown in percentage [10]. 16
Figure 1.2 Example of coding system interface 24
Figure 1.3 Data sources and road attributes which can be collected by each source proposed by AiRAP [25] 26
Figure 1.4 A simple example of one convolutional layer with one 3x3 pixels filter applied on input image with one channel and size of 6x6 pixels. With stride 1, output array (feature map) is 4x4 pixels. 28
Figure 1.5 (a) max-pooling process applied on the feature map; (b) average pooling process applied on the feature map. 28
Figure 1.6 Example of four different computer vision tasks which can be solved by CNN; (a) Image classification; (b) Object detection; (c) Semantic segmentation; (d) Instance segmentation. 31
Figure 1.7 Architecture of Faster R-CNN. Yellow elements present CNNs, while green elements present output vectors of CNNs. Final regression layer provides bounding boxes of detected objects (red rectangles), while classification layer classifies detected object into one of pre-defined classes. 34
Figure 2.1 Location of the observed area on the Open Street Map and Croatian digital orthophoto from 2018 along with the image recorded from Unmanned Aerial Vehicles (UAV) whence Ground Control Points (GCPs) were marked. 44
Figure 2.2 Proposed framework for the determination of traffic flow parameters. 45
Figure 2.3 Image processing segment: firstly, frames were extracted from the video, then image alignment was applied, and finally, a narrow motorway area was cropped. 47
Figure 2.4 (a) Characteristic points of bounding boxes; **(b)** example of difference between ground truth and estimated bounding boxes. 49
Figure 2.5 Marked characteristic locations of the observed area used for calculating location-based traffic flow parameters. 51
Figure 2.6 Marked lane segments of the observed area used for calculating segment-based traffic flow parameters. 51
Figure 2.7 The difference between headways and gaps; red line marks the reference line for measuring location-based microscopic parameters such as time headways and gaps. 55

Figure 2.8 (a) Distribution of vehicles in the training set based on their scales and aspect ratios if anchor size is set to 16 × 16 pixels; **(b)** distribution of vehicle sizes based on height and width of their bounding boxes. 56
Figure 2.9 (a) MAPE – Frame interval diagram showing a decrease in MAPE as the number of frame intervals increases, the selected optimal frame interval is determined by the red line. (b) Time – speed diagram for ground truth and estimated data of vehicle 139 with N = 1; it is visible that speeds of estimated and ground truth bounding boxes have high rate of noise. (c) Time – speed diagram for ground truth and estimated data of vehicle 139 with N = 12; it is visible that the increase in frame interval causes a smoother speed curve. 58
Figure 2.10 (a) Trajectories of vehicles 133 and 136 and their change of speed during travel. (b) Speed – time diagram of vehicles 133 and 136. (c) Space – time diagram of vehicles 133 and 136. 59
Figure 3.1 Proposed framework for the determination of RSS – O and RSS – D attribute. 73
Figure 3.2 Example of one created road segment with appropriate dimensions. Red arrow represents driving direction, while black arrows represent dimensions of segment. 75
Figure 3.3 Example of upsampling 10 m road segments to one 100 m iRAP defined road segment. 75
Figure 3.4 (a) Process of translation in direction of x-axis and y-axis and rotation around z-axis for Rz angle; (b) process of translation in direction of z-axis. 77
Figure 3.5 (a) Example of road cross section on part of road with overpass; (b) example of road cross section on part of road with irregular rockface; (c) example of road cross section on part of road with tunnel; (d) example of road cross section on part of road with both safety barriers. 78
Figure 3.6 Example of road cross section with ground truth and detected road bounding boxes. 79
Figure 3.7 Example of road cross section with labeled iRAP defined objects. 80
Figure 3.8 Example of reference X coordinate of rigid object and reference X coordinate of upward slope rollover object. 81
Figure 3.9 Distribution of labelled RSS – O classes. 82
Figure 4.1 Study area on Open Street Map (OSM) and satellite imagery used in this study with observed roads plotted with red line (created by QuantumGIS software v3.22). 97
Figure 4.2 Three-stage workflow for detection of school zones, pedestrian crossings, and physically divided carriageways. 98

Figure 4.3 Transformation process of cropped road segments. Red arrow presents centerline of road, while Tx, Ty, and θ are transformation parameters; (a) translation by Tx and Ty distances and rotation for θ angle; (b) transformed road segment with centerline aligned with Y-axis (road-oriented). 100
Figure 4.4 Diagram of pedestrian crossing classification tasks and classes annotated for this study with appropriate examples. Red arrow presents centerline of inspected road. 101
Figure 4.5 Examples of annotated classes on road segments. (a) school zone; (b) inspected road pedestrian crossing; (c) inspected road ossing with refugee island; (d) side road pedestrian crossing; (e) side road pedestrian crossing with refugee island; (c) inspected road pedestrian crossing with refugee island; (d) side road pedestrian crossing; (e) side road pedestrian crossing with refugee island; (f) divided carriageways. 102
Figure 4.6 Distribution of annotated samples of school zones, divided carriageways, and pedestrian crossing classes. Number of training samples are shown in yellow, while test samples are shown in blue. 104
Figure 4.7 Precision – recall curve generated for each class with IoU and confidence thresholds at 0.5. It is visible that the object detector achieved a significantly lower precision – recall tradeoff for SRRI and IRRI classes. 107
Figure 4.8 Examples of correctly detected classes by trained YOLO detector. Green bounding boxes show ground truth objects, while g boxes present predicted objects; (a) school zone markings; (b) inspected and side road pedestrian crossings, (c) inspected and side road pedestrian crossings with refugee islands as well as side pedestrian crossing on right side; (d) divided carriageways. 108
Figure 4.9 Examples of false positive, false negative, and misleading detections by trained YOLO detector. Green bounding boxes show ground truth objects, while red bounding boxes present predicted objects; (a) misleading detection of divided carriageways on non-inspected road; (b) false negative detection of side road pedestrian crossing; (**c**) false positive detection of inspected road crossing on roadside object; (d) false negative detection of side road pedestrian crossings. 109

List of Tables

Table 2.1 Table of parameters and their values used in Faster R-CNN ResNet50 COCO network. 48
Table 2.2 Confusion matrix of test dataset. 57
Table 2.3 Evaluating metrics of test dataset. 57
Table 2.4 Root Mean Square Error (RMSE) value of characteristic points. 60
Table 2.5 Estimated location-based traffic flow parameters. 60
Table 2.6 Estimated segment-based traffic flow parameters. 61
Table 3.1 Technical specification of Trimble MX8 Land Mobile Mapping System. 74
Table 3.2 Confusion matrix for predicted and ground truth objects presented with percentage. 83
Table 3.3 Recall, precision, and AP for each class as well as mean recall, precision and AP. 84
Table 3.4 RMSE value for every RSS – O class. 85
Table 3.5 Confusion matrix for final road segments classification into one of RSS – O classes presented with percentage. 86
Table 3.6 Confusion matrix for final road segments classification into one of RSS – D classes presented with percentages. 87
Table 4.1 Confusion matrix of ground truth and predicted objects. In addition to the detected objects, the number of Background False Positive (FP) and Background False Negative (BFN) detections is given. BFN is given on the predicted axis, while BFP is given on the ground truth axis. Bright shades of green present a lower number of matched classes between ground truth and predicted objects. Contrary, dark shades of green present higher number of matched classes between ground truth and predicted objects. 106
Table 4.2 Accuracy, recall, precision, F1 score, and AP for each class as well as mean values of every performance measure. 106

Curriculum Vitae

Ivan Brkić was born on 24 June 1994 in Split, Croatia. He completed his primary and high school education in Split in 2013. Afterwards, he started his academic journey and enrolled in a Bachelor's degree programme at the Faculty of Civil Engineering, Architecture and Geodesy in Split in 2013, which he successfully completed in 2016. His commitment to education led him to pursue a Master's degree, for which he enrolled at the Faculty of Geodesy in Zagreb in 2016. He completed his Master's degree in 2018 by defending Master thesis entitled "Spatial-Temporal Analysis of Bat Plecotus Habitat Using Machine Learning Techniques".

After completing his Master's degree, he worked as a Master of Geodesy at Geographica d.o.o. from 2018 to 2019, where he gained practical experience and insight into the field. Since 2019, he has been working as a researcher and teaching assistant at the Department of Geoinformatics at the Faculty of Geodesy in Zagreb. In this role, Ivan actively contributes to the education of students by teaching in courses such as Databases, Geoinformation Modelling and Mobile Survey and GIS. He has also been involved in supervising several master's theses and has given numerous scientific and popular lectures.

He has authored and co-authored several scientific papers and research projects in the fields of geoinformatics, machine learning and transportation. In addition, he presented his work at numerous international and national conferences and workshops.

In terms of research projects, he has actively participated as a researcher in project entitled "Advanced Forest Environmental Services Assessment and Geospatial Monitoring of Green Infrastructure by Means of Terrestrial, Airborne, and Satellite Imagery" (GEMINI).

Ivan Brkić has also contributed to several professional projects. Particularly noteworthy is his participation in projects such as "Change Detection between Two Sets of Satellite Imagery for the Area of the Republic of Croatia", "Safer Bicycle Routes in Danube Area" (SABRINA) with a focus on license plate and human detection using machine learning algorithms. Furthermore, he worked on creation of road cross-section images from mobile Lidar point clouds as part of project "Determination of Priorities and Creation of Project Documentation for Harmonising the System of Protective Fences on HAC Highways with the Regulations". At project "Saving Lives, Assessing and Improving TEN -T Road Network Safety" (SLAIN) he was focused on license plate detection using machine learning algorithms.

In his scientific and professional work, he is an active user of Python programming language and series of packages such as Tensorflow, PyTorch, Pandas, Geopandas, GDAL, PDAL, etc.