CZECH TECHNICAL UNIVERSITY IN PRAGUE

FACULTY OF MECHANICAL ENGINEERING

DEPARTMENT OF INSTRUMENTATION AND CONTROL ENGINEERING

EXPERIMENTAL EVALUATION OF CAMERA BASED ADAS SYSTEM

BACHELOR THESIS

Supervisor: Ing. Václav Jirovský, Ph.D.

2021 Aslah Puliyath Hussain

BACHELOR‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Puliyath Hussain Aslah** Personal ID number: **474928**

Faculty / Institute: **Faculty of Mechanical Engineering**

Department / Institute: **Department of Instrumentation and Control Engineering**

Study program: **Bachelor of Mechanical Engineering**

Branch of study: **Information and Automation Technology**

II. Bachelor’s thesis details

Bachelor’s thesis title in English:

Experimental evaluation of camera based ADAS system

Bachelor’s thesis title in Czech:

Experimentální ověření funkcí ADAS systému založeném na rozpoznávání obrazu z kamery

Guidelines:

- Analyze the principles of currently used technologies for such object detection with primary focus on visual sensors (cameras).
- Design an experiment, which will identify minimum requirements for the picture quality while maintaining relevant object detection level. Focus on quantitative and qualitative evaluation of the experimental data.
- For the object recognition, use publicly available tool(s) as a benchmark.
- In the experiment include necessary variations of typical technical and environmental conditions.

Bibliography / sources:

[1] Cheng H.: Autonomous Intelligent Vehicles. London: Springer-Verlag London ltd., 2011. ISBN 978-1-4471-2279-1.
[2] The digital signal processing handbook. 2nd ed. Editor V. MADISETTI. Boca Raton: CRC Press, 2010. The electrical engineering handbook series. ISBN 978-1-4200-4604-5.

Name and workplace of bachelor’s thesis supervisor:

Ing. Václav Jirovský, Ph.D., 16123

Name and workplace of second bachelor’s thesis supervisor or consultant:

Date of bachelor’s thesis assignment: **30.04.2021** Deadline for bachelor thesis submission: **10.06.2021**

Assignment valid until: ______________

Ing. Václav Jirovský, Ph.D. Supervisor’s signature	Head of department’s signature	prof. Ing. Michael Valášek, DrSc. Dean’s signature

III. Assignment receipt

The student acknowledges that the bachelor’s thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor’s thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt	Student’s signature

Annotation List

Name: Aslah

Surname: Puliyath Hussain

Title Czech: Experimentální ověření funkcí ADAS systému založeném na rozpoznávání obrazu z kamery

Title English: Experimental Evaluation of Camera Based ADAS System

Scope of work:

number of pages: 86

number of figures: 37

number of tables: 17

number of appendices: 4

Academic year: 2020-2021

Language: English

Department: Department of Instrumentation and Control Engineering

Specialization: Information and Automation Technology

Supervisor: Ing. Václav Jirovský, Ph.D.

Reviewer:

Tutor:

Submitter:

Affidavit

I confirm that the bachelor's work was disposed by myself and independently, under the lead of my thesis supervisor. I stated all sources of the documents and literature.

In Prague ……………… ……………………

Aslah Puliyath Hussain

Acknowledgements

I would like to express my gratitude towards my supervisor Ing. Václav Jirovský, Ph.D. for his expert guidance throughout this thesis. I would also like to thank Uniqway for referring me to my supervisor, as well as my family for their unwavering support.

Abstract

This thesis is aimed at the experimental evaluation of camera based ADAS system. With a major focus on cameras, a comprehensive overview of the currently used sensor technologies is presented, along with their concepts, types, and limitations. For the purpose of evaluating the experiment, photographs are shot at eight different sites in Prague city, both during the day and night. These images are examined using a proposed object detection algorithm. The experiment is designed with an emphasis on quantitative and qualitative measurement of data analysis.

Keywords: *Object detection, Computer Vision, Autonomous Vehicles, Sensors*

Table of Contents

1. Introduction

The number of road traffic accidents is one of the world's major societal issues today. Accident reduction technologies are becoming increasingly important for automotive companies as consumers place a greater emphasis on safety. The development of driver assistance systems began with the introduction of the Anti-lock Braking System (ABS) into serial manufacturing in the late 1970s. Although Advanced Driver Assistance Systems (ADAS) cannot totally prevent accidents, they can better protect us from some of the human variables that cause most traffic incidents. Object detection is a critical issue for ADAS. Convolutional neural networks (CNN) have lately gained significant success in object detection, outperforming older algorithms that employ hand-engineered features. Popular CNN detectors, however, do not achieve very excellent object recognition accuracy because of the demanding driving environment e.g., huge object size variation, object occlusion, and poor lighting conditions.

In recent years, there has been a substantial surge in research interest supporting the development of the autonomous vehicle, which is an automotive platform capable of perceiving and reacting to its immediate surroundings in an attempt to navigate roadways without human involvement. Object detection is one of the most important prerequisites to autonomous navigation in many autonomous driving systems, as it allows the car controller to account for obstacles when considering possible future trajectories; as a result, we need object detection algorithms that are as accurate as possible.

The goal of this thesis is to analyze the fundamentals of currently utilized sensor technologies for such object detection, with a specific focus on visual sensors (cameras). This thesis also researches the environmental factors at work in the artificial intelligence industry, as well as their direct engagement in training big data models. The experiment model was then effectively designed to accommodate these findings by focusing on quantitative and qualitative analysis of the experimental data.

2. Sensor Overview

Sensors are advanced systems that sense and respond to some kind of feedback from the physical world, converting it into an electrical signal that can be measured. A sensor transforms a physical phenomenon into a digital signal (or, in some cases, an observable analog voltage), which is then displayed on a human-readable display or transmitted for further processing. It senses environmental changes and responds to any output on another system. The basic input may be light, heat, motion, humidity, pressure, or any of a variety of other environmental phenomena. The appropriate choice of a sensor is dependent on awareness of the application type, product variables, and operating environment conditions. Along with temperature, scale, safety class, and whether the sensor needs a discrete or analog input, sensor repetition accuracy, sensor reaction time, and sensing range are other factors taken into account during sensor selection. Choosing the right sensor for the appropriate application would aid in the most reliable and accurate optimization of the whole system.

Sensors are categorized as active, or passive based on their power requirements and mode of function. An active sensor is a sensing device that requires an external source of power to operate, whereas passive sensors simply detect and respond to some kind of feedback from the physical environment [9]. Active sensors emit energy to scan objects and locations, during which a sensor senses and analyses the radiation reflected or backscattered by the target. GPS, Radar and LiDAR are few examples of active sensor-based technologies, in which the time interval between emission and return is determined to evaluate an object's position, distance, and direction. Passive sensors produce power within themselves to run, which is why they are referred to as self-generating types. The quantity being calculated provides the energy required for operation. Passive sensors collect radiation produced or reflected by the target or its surroundings. The most frequent source of radiation detected by passive sensors is reflected sunlight. Film imaging, infrared, charge-coupled instruments, and radiometers are all examples of passive remote sensors.

2.1 **Radar Sensor**

Radar is an electromagnetic sensor that works by broadcasting radio waves out and then detecting reflections off of objects. The word RADAR stands for “Radio Detection and Ranging”. Radar is a detection device which uses radio waves to determine the range, angle, altitude or velocity of an object. These radio waves used in radar are equipped to travel well through air, fog, clouds, snow etc. The targets can be any moving objects such as automotive vehicles, people, animals, birds, insects or even rain. In addition to determining the presence, position, and velocity of such objects, radar can also obtain their size and shape.

Radar is an active sensor which has a transmitter that acts as its own source of illumination to detect objects. Usually, it resides in the microwave region of the electromagnetic spectrum measured in hertz (cycles per second) at frequencies ranging from around 400 MHz to 40 GHz [2]. However, for long-range applications it is used at lower frequencies i.e., HF (high frequency; 3 MHz – 30 MHz) and also at infrared and optical frequencies. Depending on the range of frequency it uses, the physical size of a radar system can vary from the size of a palm to the size of a soccer field [1].

Radar waves travel through the air at almost the speed of light which is roughly 300,000 km per hour. The most commonly used radar releases a chain of intermittent pulses in order to detect the object and is often called the pulse radar. This power focused, high radio pulses propagate at a speed of light and are directed in one direction with the help of an antenna. The antenna works both as a transmitter and receiver with the use of a vital equipment called "duplexer" which is a part of the radar apparatus. The duplexer performs the duty of swapping the antenna back and forth between transmitter and receiver. While the antenna is transmitting, it cannot receive and vice-versa. A radar antenna serves the purpose of concentrating or focusing, the radiated power in a small angular sector of space. The antenna is one of the most critical parts of the radar system. It transfers the transmitter energy from the transmitter to the environment with the necessary distribution and efficiency, while ensuring the signal has required pattern in space. And it provides the target position updates while on reception. Figure 1 shows the internal structure of a typical radar system.

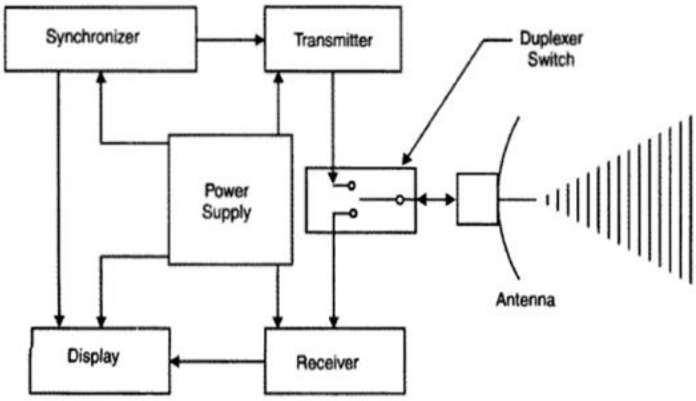


Figure 1. Block diagram of a typical radar system [8]

Upon receiving the transmitted signal, radar then evaluates the distance of the object depending on the information received using the formula:

$$d = \frac{cT}{2} \tag{1}$$

where, d is the distance to the target, c is the speed of propagation of waves and T is the time taken for the waves to complete the round-trip. With the potential to detect a moving or stationary object, radar's major advantage over other sensors like LiDAR is its ability to function in adverse weather and lighting conditions. It also takes low power to radiate signals that are capable of penetrating insulators. However, its inability to tell a target's color, internal aspects, or to recognize objects behind certain conducting sheets are its major downsides.

2.1.1 Detection Principles

The fundamental principle of radar operation is simple. The radar device transmits electromagnetic energy and analyzes the energy that is transmitted back to it by an object. It is a principle similar to that of an echo using short-wave microwaves instead of sound waves. When in contact with an object, the waves resound and thus, the distance and direction to the target can be accurately measured.

The measurement of an object's range from a radar antenna can be determined by these properties of an electromagnetic waves:

a) Reflection of Electromagnetic Waves
   The electromagnetic waves return as they land on the electrically conductive surface. If these waves are obtained back at the site of origin, this means that there is an obstruction in the direction of propagation.
b) Constant Speed
   The electromagnetic waves travel through air at a constant at approximately the speed of light (300,000 km/s). This constant speed allows the distance between the reflected target and the radar site to be evaluated by calculating the running time of the emitted pulses.
c) Direction of Travel
   The energy typically travels through a straight line, and only deviates due to atmospheric and weather conditions. By using radar antenna, this energy can be directed in a desired direction. This helps in knowing the azimuth and elevation of the target along with this its direction.

The fundamental concepts described above can be used in the design and application of a fully operational radar system, which then enables the distance, orientation, and elevation of the reflected target to be calculated with precision.

2.1.2 Performance

The maximum range of a radar system depends largely on the average power of its transmitter and its antenna size. In common cases, where transmitter and receiver are at the same location, the power returning to the receiving antenna can be defined by the equation:

$$P_r = \frac{P_t G_t A_r \sigma F^4}{(4\pi)^2 R^4} \tag{2}$$

where, P_t – transmitter power

G_t – gain of the transmitting antenna

A_r – effective aperture of the receiving antenna

R – is the range (total distance from the transmitter to target and target back to receiver)

σ – radar cross section, or scattering coefficient of the target

F – pattern propagation factor

The equation (2) indicates that the received power decreases as the fourth power of the range, which means that the received power from distant targets is comparatively weak. Some of the limiting factors that affect the performance of a radar in its environment include its beam path and range, signal noise or interference, clutter and jamming.

2.1.3 Power Limitations

Depending on the application, the radar system comes in different forms of shape, size and range of frequency. The frequency of a long-range surveillance radar can be somewhere from 50-1000 MHz, when systems used for moderate-range and marine purposes utilizes a range of 2-4 GHz to 8-12 GHz, respectively [1].

The radiofrequency (RF) used in some radar systems are limited due to its human-environment hazards as well as bad interferences caused to other equipment used in fields like radio astronomy. Frequent exposure to these kind of radar frequencies can cause harmful effect on human beings as well as other living organisms. If this radiofrequency radiation is absorbed by human body in excessive amounts, it can generate heat. This can lead to burns and body damage. For the same reason, the radar systems used in automobiles are regulated to a certain level of frequency and power by governments. Regulation specifies to decrease the power of radars when the vehicle on which radar is mounted to is stopped, or not moving. The power density should be below the threshold limit of 1 mW/cm adopted to human exposure level to RF radiation [15]. In the EU, the automotive radar system is limited to 77GHz to 81 GHz (79 GHz).

2.1.4 Signal Attenuation

The reduction or lack of signal strength is generally known as signal attenuation. Attenuation happens as the signal is transmitted through the medium which may be affected by different factors, such as atmospheric conditions and propagation route barriers, resulting in a smaller detection range.

Radar loses some of its strength while it travels through the atmosphere. The atmosphere induces losses in radar signal propagation due to atmospheric attenuation, and spread of beams [11]. The analysis shows, the greatest influence of all the causes in attenuation is atmospheric gases and rain [12]. This attenuation generally occurs due to atmospheric gases like oxygen and water vapor including fog and rain. The attenuation of radio waves in the atmosphere, L_{atm}, needs to be calculated in order to measure the detecting wavelength. This attenuation is defined by the following formula:

$$L_{atm} = 2 \cdot D_{atm} \cdot (\gamma_g + \gamma_R) \tag{3}$$

where, γ_g: specific attenuation due to atmospheric gases (dB/km)

γ_R: specific attenuation due to rain (dB/km)

D_{atm}: target detection distance in the Earth's atmosphere

From equation (3), it is clear that the atmospheric attenuation is directly proportional to the intensity of rain and atmospheric gases.

Along with the attenuation caused by weather conditions, the radars used in automobiles are subjected to material attenuation. Automobile radars are usually integrated behind an emblem or bumper. The radiofrequency (RF) transmission loss of the radome material attenuates twice since the signal has to pass through the material on the way to target and on the way back, producing reduced detection range [13]. Radomes are large dome-shaped structures that shield radars from bad weather, but at the same time allow electromagnetic signals to be obtained by radar without any interference or attenuation [14]. The reflectivity and uniformity of the radome material is also an important factor that impairs radar performance. For instance, metallic particles in paint can create reflections, and an RF mismatch in the base material can produce interference signals within the radome, near the sensor [13]. These interference signals are received and downturned in the receiver chain, reducing the detection sensitivity of the radar. Many car manufactures aim to minimize this effect by tilting the radome so that the transmitted radar signal is mirrored elsewhere and not directly back to the front end of the receiver.

2.1.5 SNR

In signal processing, noise is a general term for unintended (and usually unknown) changes that the signal may suffer during recording, storage, delivery, processing or conversion [22]. Noise usually occurs as unpredictable deviations superimposed on the ideal echo signal received by the radar receiver. The lower the power of the desired signal, the more difficult it is to separate it from the noise.

SNR is a ratio that determines the difference in level between the signal and the noise within a desired signal, often expressed in decibels [dB]. The lower the noise produced by the receiver, the higher the ratio of signal to noise. In radars, signal to noise ratio, SNR or S/N is a method of measuring the sensitivity of the radio receiver [21]. SNR in general can be defined as:

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{4}$$

where, P is the average power in P_{signal} and P_{noise}. According to the equation, the higher a radar system's SNR, the better it is at distinguishing actual targets from noise signals. It is also important to ensure that all signal and noise are measured at the same or equivalent point in the device and within the same circuit bandwidth [21]. Noise floor is another measure of performance that affect range performance. It can be defined as a measure of the signal generated by the sum of all noise sources and unwanted signals inside the device. A target that’s too far away generates too little signal to surpass the noise floor and cannot be detected. Detection thus requires a signal that exceeds the noise floor by at least the signal to noise ratio.

2.2 **LiDAR Sensor**

LiDAR, an acronym for “Light Detection and Ranging” and “Laser imaging Detection and Ranging” is a type of sensor used to detect its surroundings. Typically, a LiDAR sensor emits pulsed light waves into the surrounding environment which bounce off from the objects and return to the sensor. It emits usually up to 150, 000 pulses of laser light of either visible ultra-violet or near infrared light at the targets. The sensor then uses the time it took for each pulse to return to the sensor to calculate the distance it travelled. Like radar, the distance is then computed using equation (1).

A LiDAR is an active system which generates its own energy – in this case, light – to measure things in its vicinity. Its rapidly firing light beams using visible, near infrared or ultra-violet light to map out the environment around it. It can then get both the sense of physical dimension and motion of the object it falls on to. Traditional LiDAR units use lasers at wavelength 905 nanometers [5]. The pulsed lasers track the time it takes at nano second speed for the signal to return to its source. This allows the LiDAR to produce a 3D model of the surface or object. A LiDAR system consists of four main components: a transmitter for transmitting laser pulses, a receiver for intercepting pulse echoes, an optical analysis system for processing input data, and a powerful computer for visualizing a live, three-dimensional image of the system environment. Photodetector and optics are elements that play a vital role in data collection and analysis in the LiDAR system [3]. A full LiDAR system can include other main components such as phased arrays and microelectromechanical devices. All of these elements work together to provide a 3D representation of the target.

Figure 2. LiDAR on a latest smartphone [17]

Based on the platform it used there are mainly two types of LiDAR systems: Airborne LiDAR and Terrestrial LiDAR. Airborne LiDAR is installed on drones and helicopters to collect data from the ground surface while the terrestrial LiDAR is the system implemented in moving vehicles or tripods to collect data points. Today, LiDAR is even installed in some modern smartphones which makes photography more efficient and precise and also, enhances the capabilities of augmented reality. In the case of self-driving cars, LiDAR is used to generate 3D maps in which the car can navigate. Using shorter wavelength laser lights, it is capable to precisely measure much smaller objects. Its major advantage is accuracy and precision.

2.2.1 Detection principles

The LiDAR sensor senses targets and measures some of the characteristics of the targets, such as distance, speed, reflectivity, angular location. The LiDAR device uses laser beams of chosen wavelength from the ultraviolet to the infrared spectrum. The laser-composed emitter sends light pulses and sets a timer. Objects in the LiDAR Field of View (FOV) reflects these light pulses back to the detector, which consists of an electro-optical system that converts the light signal into an electrical signal. The quantum efficiency of the detector relates to how effectively the photoelectric detector converts the received photons obtained from the event into power electronics. The optical efficiency of the receiver relates to the percentage of the light obtained that goes into the optical aperture, including the spectral filter [25]. In most LiDAR devices, a spectral filter is used to exclude incoming light outside a specific spectral band centered at the wavelength of the laser. The converted electrical signal is then interpreted by an electronic chain to acquire target information [16]. The observed target would then appear as a point cloud in the LiDAR monitor. When several laser transmitters are combined, monitoring capacities are massively expanded, acquiring millions of individual reflection points simultaneously.

2.2.2 Performance

Laser radar signal produced by the laser launches to the atmosphere. The target reflects back the signals and gets into the laser receiving system, after travelling back through the atmosphere. Laser radar power at that time can be defined as:

$$P_r = G_d \eta_s \eta_q \eta_r P_t A_\Delta / (R^2 \Omega_{\text{laser}}) \, A_{\text{r}} / (R^2 \Omega_{\text{t}}) T_{atm}^2 \tag{4}$$

where, P_r – is the instantaneous value of the echo-signal powered at wavelength λ, G_d – is the receiver gain, η_s – is the optical efficiency, η_q – is the detector quantum efficiency, η_r – is the reception efficiency, P_t – is the laser emission power, A_Δ – is the effective area of the target reflection aperture, A_{r} – is the area of receiving aperture, Ω_{laser} – is the solid angle of the laser beam, Ω_{r} – is the solid angle of the echo laser beam, R – is the current range, T_{atm} – is the atmospheric transmittance coefficient.

2.2.3 Power Limitations

Like all autonomous technologies, LiDAR also comes with its downfall. One key limitation of LiDAR sensors is that it cannot see beyond solid objects, which is true for any system that relies on signals travelling in a straight line [4]. If the system is obscured with anything in close range, a huge amount of data is lost. Likewise, adverse weather conditions and clashing signals from other systems are also not favorable for LiDAR's function. It is also unclear what so much laser activity would do to other biological and mechanical systems in the environment. For example, Luminar a tech company, works on a LiDAR system that operates at 1550 nm versus the traditional 905 nm, and there are claims that it could potentially damage the human eye cornea [5].

A tracking microwave (X-band) radar has a frequency of 10 GHz which corresponds to a wavelength of 3 cm and a typical search (L-band) radar has a frequency of 1 GHz and a wavelength of 30 cm [18]. A typical eye-safe LiDAR will have a frequency of 200 THz and a wavelength of 1.5 μm which is 20, 000 times smaller than the wavelength of a X-band tracking radar and 200, 000 smaller that the L-band search radar. Laser radiation can damage the eye by burning the retina after magnification, or by burning the surface of the eye. Lasers of greater than ~1.5 μm or less than than 0.4 μm are better because the water in the eye absorbs wavelengths in these areas, restricting light from concentrating on the retina [18]. It is common for LiDAR to operate at 1.5 μm or longer and it rarely operates below 0.4 μm. The traditional LiDAR used for ADAS system in automobiles utilizes a wavelength of 905 nm accounting for the human eye-safety threshold.

Depending on the application, cost can also be a consideration when selecting a LiDAR system. The major setback of implementing LiDAR system in modern self-driving technology is its cost. Google's system originally costs up to $75,000 [5]. Even though companies like Luminar and Velodyne are bringing down the price range from $100 to $1000, the real question lies on how many of these sensors each car or system needs in order to get the desired result. Its inability to read words, recognize colors, and its relatively large physical size also adds up to its downfalls, where typical cameras usually excel. The major benefit and distinction of LiDAR over radar is that the beam divergence or how fast the beam spreads when the distance is much smaller.

2.2.4 Signal Attenuation

Similar to radar, LiDAR drops its signal strength as the signal makes its way back to the sensor. Reflection, diffraction, absorption in various climatic conditions are the few causes for this reduction in signal power.

Target reflectivity, the difference in the material of a target, reflects the laser light in varying intensities. For instance, a car has a windshield made of glass, body made of metal and bumpers with plastic. It is experienced that the signal to noise values for the car body is greater for a set distance relative to the windshield and bumper [16]. Objects like metal can be seen at a longer distance compared to less reflective material. Weather effects are other reasons that impairs the LiDAR detection range. Moist air acts as a screen for the infrared radiation. Both fog and rain minimize the laser intensity by absorption and diffusion of the laser beam by tiny water droplets. Fog and rain then serve as a screen on LiDAR sensors that restrict their capabilities and range of detection. Glaring sun that dazzles the LiDAR during the daytime can also factor in laser energy attenuation. The signal-to-noise ratio (SNR) of LiDAR equation backscattering is often attenuated by noise and interference such as nonlinear turbulence, background noise, dark current, electronic noise readout and atmospheric turbulence [24]. Target signals get polluted with the noise and affect the effective working range and target precision.

2.3 **Flash LiDAR**

LiDAR can be mainly divided into two based on the illumination method, scanning LiDAR and Flash LiDAR. Flash LiDAR is a method of implementation under Solid-state Lidars [7]. While convectional scanner LiDAR uses mechanical rotation to spin the sensor for 360-degree detection, Flash LiDAR does not move the laser or light all. It functions like a camera, delivering a flash light to detect the entire surrounding area at once, and processing the details using an image sensor. Figure 3 shows multiple 3D flash LiDAR sensors used around a car for its 360^o coverage.

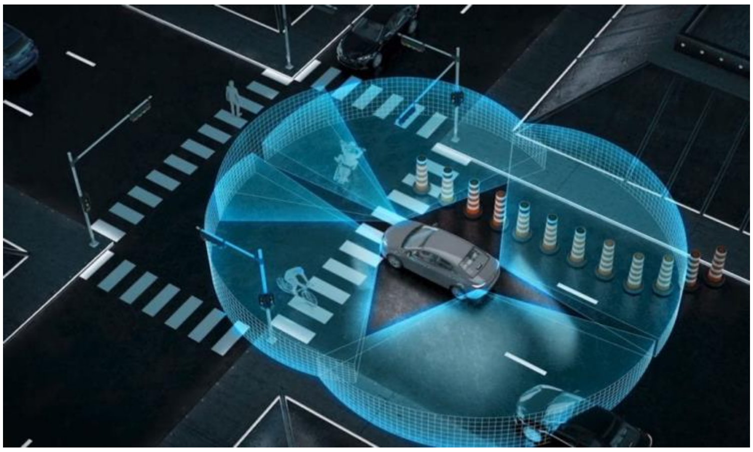


Figure 3. 3D Flash Lidar units providing 360° coverage [10]

As this method captures the entire scene in a single image as opposed to mechanical scanners, data acquisition is much faster. Also, it utilizes only a single flash to capture the entire image. Thus, making the images immune to distortion caused by vibration effects. A downside to this method is retroreflectors. Retroreflectors reflect most of the light in different directions and have minimal back scatter, hence blinding the entire sensor and rendering it useless [7]. Even though the light source of Flash Lidar is more powerful, the detecting distance and field of view is much lower compared to normal scanning LiDAR.

3. Camera

This chapter focuses on analyzing the principles of camera, which is one of the main objectives of this paper. A camera is an optical instrument or device that has the ability to capture and record both pictures and videos. Essentially, light rays bounce in different directions, and when all these light rays come together on a digital camera sensor, they create an image [6]. The lens of the camera takes all the light rays that bounce around and uses a glass to redirect them to a single point, producing a sharp image.

Today, cameras are available in all kinds of forms ranging from a button size to professional hand-held camera. They're utilized in various applications from surveillance to autonomous driving. The main internal components of a camera include multiple sensors, a shutter, mirror, pentaprism, diaphragm and a CPU to process the image. Cameras with advanced capabilities can be seen in almost every smartphone today. Similar to human vision, cameras in autonomous cars utilizes the same feature available in modern cameras. Using multiple cameras, the surrounding of the car is visualized and processed back to its CPU providing a better understanding of the environment around it and the information necessary to assist in autonomous driving.

Resembling a solar panel, a modern digital camera's sensor is divided up to millions of red, green and blue pixels i.e., megapixels. The sensor converts it into energy when light hits the pixel and a built-in computer reads just how much energy is being generated. Measuring how much energy each pixel has, enables the sensor to determine which areas of the picture are light or dark [6]. Using each pixel's color value, a camera's computer is able to assess the colors in the scene by looking what other nearby pixels are recorded. Gathering all the pixels together, the computer is able to estimate the approximate color and shape in the scene. Since each pixel is gathering light information, having a larger sensor helps in packing numerous megapixels and thus, making high resolution low-light images possible.

Cameras are much less expensive compared to LiDAR-like systems and essentially help bring down the cost of self-driving cars for the end-consumers. The availability of the cameras in the market in different forms makes it easier to incorporate it into the design of the car making it more appealing to the customers. Unlike both radar and LiDAR systems, it can also interpret the color, words, and street signs on the road. Just like human eyes, the main drawback of cameras is the change in lighting conditions where the subject matter becomes obfuscated. Situations like strong shadows, bright lights from the sun or oncoming cars can cause confusion. Its strong dependency on powerful machine or deep learning to interpret the exact distance, location or position of an object only using its raw image data makes it difficult to implement, as opposed to sensors like radar and LiDAR. It is one of the reasons why automotive companies like Tesla use a combination of both cameras and radars to make self-driving possible.

3.1 Detection Principles

A digital image, which is simply an array of numbers with each number representing a brightness value, or grey-level value, for each picture element, or pixel, is created by a chain of physical events. This physical chain of events is called an "imaging chain". Understanding the physical process that produces an image helps in clarifying many questions about the quality of the image and its limitations. The physical process of producing an image can be broken down into the individual steps that bind together to create the imaging chain. By modeling the links mathematically in the image chain and analyzing the device as a whole, the relationships between the links and the consistency of the final image product can be known, thereby reducing the probability that the camera will not meet standards when it is operational. Modeling and analyzing the end-to-end image creation process from scene radiometry to image display is crucial to understanding the device parameters required to achieve the optimal image quality.

The imaging chain, the method by which the image is created and viewed can be defined as a sequence of physical events, i.e., beginning with the light source and ending with the display of the image produced. The key links of the imaging chain are the radiometry, the camera, the processing, the display, and the image perception [33]. A block diagram of the image chain is shown in Figure 5.

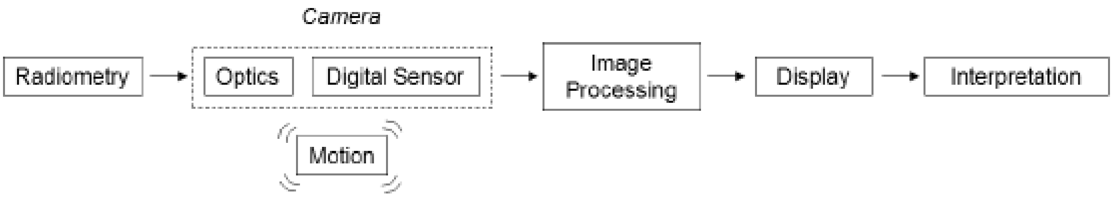


Figure 4. Imaging chain model [33]

Mathematical models representing the image chain can be used to simulate the real images that the camera would generate when it is built. This is a very helpful and valuable application of the imaging chain, since it helps the image content to be visualized during the design process and can detect design flaws before the hardware development costs are incurred. It can also help identify the image quality differences between different designs to help us understand how the images will be processed, displayed and interpreted. Many objects, such as waves, points, and circles, have basic mathematical representations that will prove very useful for mathematically modeling the image chain. A simple one-dimensional wave stationary in time, can be represented by the cosine function with amplitude A, wavelength λ, and phase ϕ:

$$f(x) = Acos(2\pi \frac{x}{\lambda} - \phi) = Acos(2\pi \xi_0 x - \phi) \qquad (5)$$

where, ξ_0 is the spatial frequency of the wave, i.e., the number of cycles that occur per unit distance.

i. Radiometry

Radiometry is the science of measuring electromagnetic radiation with a set of techniques including visible light. These techniques in optics characterize the propagation of radiation power in space, as opposed to photometric techniques that characterize the contact of light with the human eye [32]. The radiometry of the imaging chain is very important since this radiometry defines the energy that the camera "senses" to generate the final image that we see and determines the strength of the signal that will be generated by the sensor. It describes the light that enters the camera in the imaging chain. The energy recorded by the camera is in the form of electromagnetic radiation, a self-propagating wave composed of oscillating electrical and magnetic fields produced by the acceleration of charged particles. For electromagnetic waves, the relationship between the wavelength and frequency is given by:

$$c = \lambda v \qquad (6)$$

where, c $= 2.9979 \text{ x } 10^8$ m/s, the speed of electromagnetic waves in vacuum. Digital cameras designed to form images falls under the visible region of the spectrum within a range of 0.4 - 0.8 μm.

In the scope of electromagnetic waves in the visible spectrum, the amplitude determines the brightness and the frequency determines the colour. It is then much more straightforward to represent a propagating wave mathematically:

$$E(x,t) = Ae^{2\pi i\left(\frac{x}{\lambda} - vt\right) - \phi} = Ae^{i(kx - \omega t) - \phi} \qquad (7)$$

Where, $k = \frac{2\pi}{x}$ and $\omega = 2\pi v$. This function is related to cosine and sine waves by the Euler relation:

$$e^{2\pi ix} = \cos(2\pi x) + isin(2\pi x) \qquad (8)$$

ii. Optics

The optical components of the camera shape the electromagnetic radiation of the image generated by the sensor. Modeling the propagation of electrometric waves through optical elements is key to understanding the accuracy of the image that is produced. In the radiance of the image, photons are released in multiple directions from light sources or are dispersed in several directions. The lens absorbs these divergent rays in such a way that they converge to the irradiance image on the sensor surface. In radiometry, irradiance is the radiant flux (optical power) received by a surface per unit area whereas, radiance (brightness) is the radiant flux emitted, transmitted or received by a given surface, per unit solid angle, per unit projected area [32].

Optical irradiance, the irradiance image at the sensor surface prior to capture, can be computed by accounting for a number of factors like, the lens f-number, magnification, relative illumination, fall-off in intensity with lens field height and by blurring the optical irradiance image by different methods [35]. The camera equation specifies a basic model for translating the scene radiance function, L_{scene}, to the optical irradiance region of the sensor, I. The equation of the camera is:

$$I_{image}(x, y, \lambda) \cong \frac{\pi T(\lambda)}{4(f/\#)^2} L_{scene}(\frac{x}{m}, \frac{y}{m}, \lambda) \tag{9}$$

where, the term f/# is the effective f-number of the lens (focal length divided by the effective aperture), m is the magnification of the lens, and $T(\lambda)$ is the transmissivity of the lens. The camera equation maintains the center of the image with fair accuracy (i.e., on the optical axis).

iii. Digital Sensor

The camera sensor senses the light shaped by the optics to produce a record of the image. Image sensors convert the optical irradiance image into a two-dimensional array of voltage samples, one sample per pixel. Each sample is linked to the position in the image space. Generally, pixel locations are arranged in order to form a regular, two-dimensional sampling array to match the spatial sampling grids of common output devices.

In most digital image sensors, the transmission of photons to electrons is linear: precisely, the photodetector (either CCD or CMOS) reaction increases linearly with the number of incident photons. The photodetector wavelength sensitivity can differ depending on the material properties of the silicon substrate, such as its thickness. But even so, the response is linear in that the detector adds up the response across wavelengths. Ignoring system imperfections and noise, the number of electrons can be rounded up around the aperture and wavelength spectrum for the ith photodetector and can be written as:

$$\iint_{\lambda,x} S_i(\lambda)\ A_i(x)\ I(\lambda, x)\ d\lambda dx \tag{10}$$

where, the mean reaction of the photodetector to the irradiance image ($I(\lambda, x)$, photons/sec/nm/m2) is determined by the quantum spectral efficiency of the sensor ($S(\lambda)$, e$^-$/photon), the aperture function over space $A_i(x)$, and the exposure period (T, sec).

The key part of a digital camera is its sensor. The sensor is crucial in deciding the image size, resolution, low light performance, field depth, dynamic range, lenses, as well as the actual scale of the camera. The image sensor is a solid-state unit, part of the camera hardware that absorbs light and transforms what it sees to an image. The sensor consists of millions of cavities called photosites. The number of photosites is equal to the number of pixels the camera has. These photosites open when the shutter opens and shuts when the exposure is over. The photons that strike each photosite are perceived as electrical signals that differ in intensity depending on how many photons were actually recorded in the cavity. Simply said, the sensor operates as the shutter opens and absorbs the photons that strike it and transforms it to an electrical signal that the processor in the camera reads and interprets as colors. This detail is then stitched together to create an image.

A modern digital camera sensor is typically available in one of two types. It is either a Complementary Metal Oxide Semiconductor (CMOS) or a Charge Coupled Device (CCD) sensor [27]. Sensors of both types turn light into electric charge and then transform it into electronic signals. Every pixel's charge is transported through a relatively restricted number of output nodes (typically just one) in a CCD sensor before being converted to voltage, buffered, and delivered off-chip as an analog signal [74]. The entire pixel may be dedicated to light capture, and the output is consistent which is a key factor in image quality. In a CMOS sensor, each pixel has its own charge-to-voltage conversion, and the sensor generally contains amplifiers, noise-correction, and digitization circuits, allowing the chip to produce digital bits. These additional functionalities complicate the design and diminish the space available for light collection. With each pixel performing its own conversion, uniformity suffers, but it is also massively parallel, allowing for great overall bandwidth and speed. CMOS are widely used in today’s modern digital cameras. Each sensor has distinct strengths and limitations that provide advantages in certain applications.

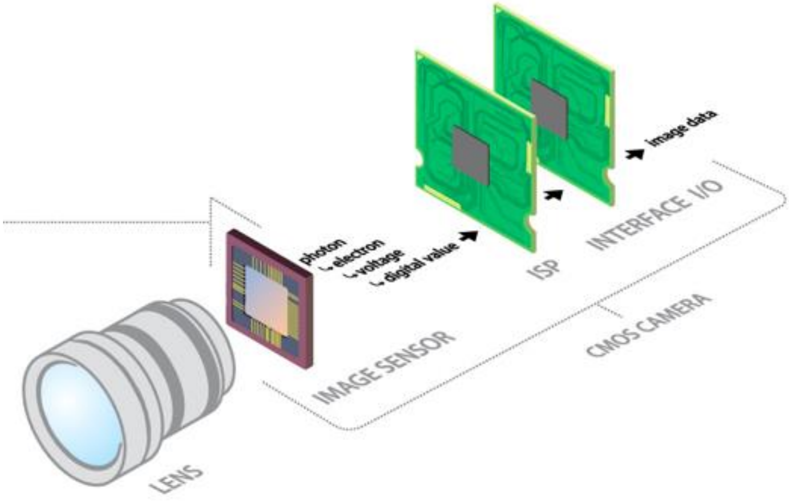


Figure 5. CMOS camera layout [28]

In a camera, as the image sensor receives incident light (photons) which are focused through the lens or other optics, depending on if the sensor is CCD or CMOS, the information would be passed to the next level as either a voltage or a digital signal [28]. Figure 5 is a schematic of a CMOS sensor which transforms photons to electrons, then to a voltage, and then to a digital value using an on-chip Analog-to-Digital converter (A/D).

iv. Image processing

The digital sensor output is a "raw" digital image composed of an array of digital count values reflecting the brightness, or gray level, of a pixel in the image for each value. Image processing is commonly used in the image chain to increase the quality of image data. It is a broad field that comprises of feature detection, compression and classification [36].

The camera acquires knowledge about the visual scene by first focusing and transmitting light through the optical device and then using an image sensor and an analog-to-digital (A/D) converter to sample the visual information. The exposure control mechanism adjusts the aperture size and the shutter speed based on the measured energy in the sensor by communicating with the gain controller to collect sensor values using a CCD or a CMOS sensor [37]. After an A/D conversion, various preprocessing operations are conducted on the acquired image data such as linearization, dark current compensation, flare compensation and white balance [38]. The aim of preprocessing is to remove noise and artifacts, eliminate flawed pixels, and create a precise representation of the scene captured. The image processing is used to perform estimation and interpolation operations on the sensor values after the sensor image data is preprocessed, in order to recreate the image's complete color representation and/or change its spatial resolution. Conventional digital cameras can be differentiated as three-sensor and single-sensor devices, based on the number of sensors used in the camera hardware [40]. Imaging pipeline of a single sensor device is shown in Figure 6.

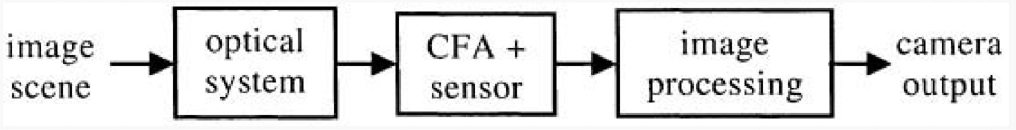


Figure 6. A single sensor imaging device [40]

The form of the CFA used in the imaging chain depends on the complexity and actual form of image processing operations. A color filter array (CFA) or color filter mosaic (CFM) in digital imaging is a mosaic of tiny color filters mounted over an image sensor's pixel sensors to capture color detail [39].

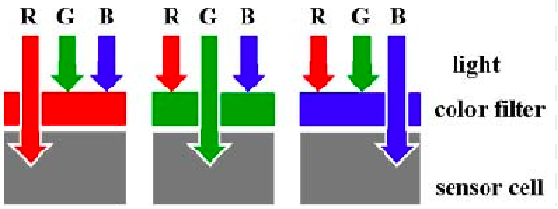


Figure 7. CFA based image acquisition [40]

Each pixel of the raw CFA sensor image has its own spectrally selective filter in the single-sensor imaging pipeline. The most commonly used color filters are RGB CFAs with alternative solutions including arrays constructed using Cyan-Magenta-Yellow (CMY) and other complementary colors. Among these, the Bayer pattern is widely used because of the ease of subsequent processing steps. Compared to R or B parts, this pattern comprises twice as many G parts, reflecting the fact that the spectral response of Green filters is similar to the luminance response of the human visual system [41].

Numerous image processing operations is performed in the camera pipeline after the CFA image is obtained. A technique called demosaicking or CFA interpolation is the most important step in a single-sensor imaging pipeline [40]. Usually, each pixel in the image of the sensor is red, green or blue. To view an image, each pixel must have a red, green and blue value. By interpolating the missing values, the display image from the sensor pixel mosaic can be build. This method of interpolation is called "demosaicking" [35]. In one dimension, the interpolation of a missing value is given by the function:

$$f(x) = \sum_{n=-\infty}^{\infty} f(n\Delta x) h_{interp}(x - n\Delta x) = f(x) * h_{interp}(x) \qquad (11)$$

where, Δx is the sampling interval and $h_{interp}(x)$ is the interpolation function [33].

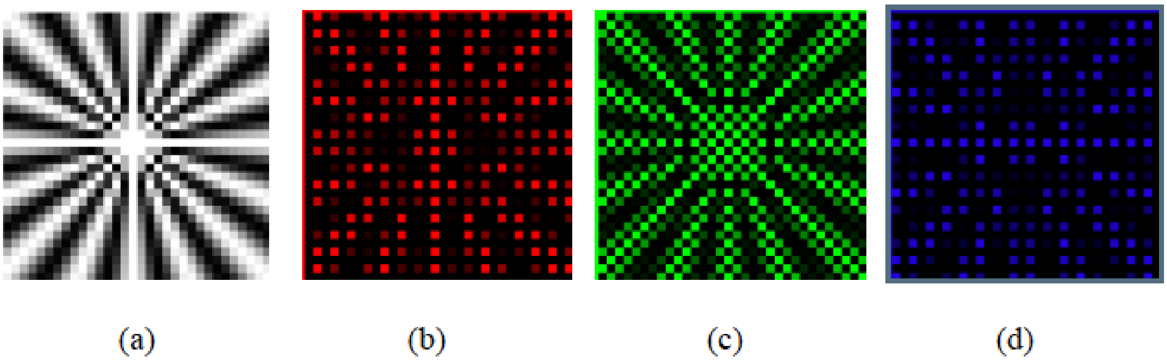


Figure 8. An illustration of color filter array (CFA) sampling [35]

Each pixel captures information about only one colour band. Figure 8 shows, (a) A cropped image from a Mackay ray chart, (b-d) the red, green, and blue CFA samples, respectively, from a Bayer CFA. Demosaicking algorithms rely on a wide variety of techniques for signal processing. The similarity of all these camera image processing techniques along with limited resources for single-sensor imaging devices, suggests that the objective is to unify these processing steps in order to provide the end-user with an integrated, cost-effective, imaging solution.

v. Display

The display media will modify the content of the depicted image, while the original data recorded by the camera remains unchanged. Generally, the user has control over the image quality associated with viewing the images on the display and has the ability to optimize the quality with adequate lighting and calibration. Modeling the display component of the image chain involves knowledge of the display device that will be used, i.e. encoding, video card, and monitor parameters, in order to accurately model the blurring, contrast, and brightness effects that will be placed on the image.

A great deal of time and cost can be invested in a camera to capture high-resolution pictures, but if the quality of the display device is low, then all the effort may be in vain. The primary image-quality considerations for the show are resolution, contrast, and brightness. The transfer function of the cathode ray tube (CRT) monitor can be modeled as the Fourier transform of the Gaussian spot that approximates the brightness profile of the pixel shown [42]. Assuming radial symmetry, the display transfer function is given by:

$$H_{display-CRT}(\rho) = e^{-2\pi^2 \sigma_{spot}^2 \rho^2} \qquad (12)$$

where, σ_{spot} is the standard deviation of the Gaussian spot. Flat panel displays, such as a liquid crystal display (LCD), have rectangular profiles, so the transfer function is given by:

$$H_{display-flat\,panel}\,(\xi,\eta) = sinc\big(d_x\xi, d_y\eta\big) = \frac{\sin{(\pi d_x\xi)}}{\pi d_x\xi}\,\frac{\sin{(\pi d_y\eta)}}{\pi d_y\eta} \tag{13}$$

where, $d_x$and d_y are the widths of the pixel elements in the x and y directions, respectively.

In reality, each pixel on a color display consists of a cluster of three separate color pixels (red, green, and blue) that our eye physically combines to see the color it desires. Color displays usually have reduced resolution, i.e., transfer functions that blur the image more due to the spatial distribution of three pixels relative to a single pixel on a monochrome display.

vi. Image Interpretation

Understanding how the image will be perceived and interpreted is the final stage of the image chain. But this understanding affects the design of the other elements of the image chain. The visual interpretation of an image can be performed both by human and a computer. For example, the intended use of the image could be for automatic detection algorithms like the ones employed in autonomous vehicle. Here, the image is not for viewing at all, in which case the optimum configuration of the image chain is likely to be different from that designed for viewing the images.

The Human Visual System (HVS) can be modeled and treated as an imaging chain to get a better understanding of the image interpretation by a viewer. Starting with the radiometry from the image monitor, then replacing the eye with the camera, the brain with the image processor, and the cognitive visualization of the image with the display. The eye pupil functions as the camera opening; thus, the optical transfer function (OTF) for the eye can be modeled as a Gaussian function that depends on the size of the pupil, i.e.,:

$$H_{eye-optics}\,(\rho) = e^{-2\pi^2\sigma_{eye}^2\rho^2} \tag{14}$$

where,

$$\sigma_{eye} = \sqrt{\sigma_0^2 + (C_{ab}\,d_{pupil})^2} \tag{15}$$

with, ρ in units of cycles/deg. The parameters σ_0 and C_{ab} are constants, and d_{pupil} is the diameter of the pupil.

3.2 Signal Attenuation

A standard camera image loses its clarity and contrast along the periphery due to optical attenuation. Bad weather - particularly heavy rain and snow are mainly the reason for poor image or weak signal in a camera system. Cameras have similar limitations as human eye. In other word their “vision” is impaired by poor lighting or adverse weather conditions like heavy snowfall/rain, swirling dust/snow, dense fog etc. Strong sunshine, road surface reflections, ice or snow covering the road, a dirty road surface, or obscure lane markings can dramatically reduce the ability of the camera to detect the side of a lane, a pedestrian, a bicycle, a large animal or another vehicle. These conditions can reduce the operation of camera-dependent systems or cause these systems to temporarily stop working.

As the light passes through the lens and reaches the image sensor, the light waves undergo diffraction and interference which also ultimately influence the quality of the image. Diffraction refers to the spreading of waves around obstacles. Diffraction is a result of interference, which in physics, is the net effect of the convergence of two or more wave trains on the intersection or coincidental path. Diffraction happens to all which have wavelike properties like sound, electric radiation, such as light, x-ray, and gamma rays; and with extremely small moving particles, such as atoms, neutrons, and electrons [29]. Diffraction of light happens as a light wave travels around a corner or through an aperture or a slit that is physically approximate in size or much less than the wavelength of the light. Lens diffraction in camera occurs as the light starts to scatter or diffract when going through a tiny opening such as the camera's aperture. Light rays entering through the narrow aperture will begin to diverge and interfere with each other. These divergent rays then travel various lengths, others shift out of phase and tend to interact with each other— adding in some areas partly or totally and cancelling out in others. This interference results in a diffraction pattern with peak intensities where the amplitude of the light waves adds, and less light where they deduct. Resolution is the smallest measurement that can be accurately distinguished by a sensor. In any electronic device that measures minor voltage changes, electrical noise is the overriding factor that restricts the smallest possible measurement [20]. Electrical noise creates graininess in images captured by the camera, and it becomes impossible to see small objects if the objects are the same as the noise induced granularity.

3.3 SNR

Signal to noise ratio is used to determine the sensitivity of a camera and how they perform at different light regimes. A number of photons P falling on a camera pixel with a quantum efficiency D_{QE} will generate a signal of N_E electrons and can be defined as:

$$N_E = D_{QE} \cdot P \tag{16}$$

The incoming photons have an intrinsic difference in the noise or ambiguity of the signal itself. This is known as "Shot" photon noise and can be represented as $\delta_{signal} = \sqrt{N_E}$. Considering the noise generated during the internal process, sensor implementation and package of a camera design, SNR can be written as:

$$\frac{S}{N} = \frac{D_{QE} \cdot P}{\sqrt{(\delta_{signal}^2 + \delta_{dark}^2 + \delta_{readout}^2)}} \tag{17}$$

where, $\delta_{readout}$ is noise generated during the readout process and δ_{dark} is the noise created by thermally induced electrons and often referred to as a dark signal since its produced in the absence of light [26].

The detected signals that reach the image sensor contains the actual signal and background signal (background noise). In order to detect the target by identifying it from the background noise it requires a high signal-to-noise ratio. Aiming for higher SNR, results in better image quality and quantitative analyses. Three main undesired signal components (noise) usually included in the measurement of the total signal-to-noise ratio of image sensor are described below.

i. Photon noise:

Photon noise results from the underlying statistical fluctuation in the image sensor incident photon arrival rate. The photoelectrons produced within the semiconductor system constitute a signal, the magnitude of which fluctuates spontaneously with photon incidence at each pixel on the image sensor [31]. The interval between photon arrivals is governed by the Poisson statistics and can be represented as:

$$photon\, noise = \sqrt{signal} \tag{18}$$

ii. Dark noise:

Dark noise is the result of statistical variation in the amount of electrons thermally produced within the silicon structure of the image sensor, which is independent of the photon-induced signal but strongly dependent on the temperature of the device. The rate of generation of thermal electrons at a given image sensor temperature is referred to as dark current [30]. Similar to photon noise, dark noise follows Poisson's relationship to dark current, which is equal to the square-root of the number of thermal electrons produced.

iii. Read noise:

Read noise or readout noise is a combination of noise from the pixel and the A/C. The Read Noise (RN) sensor is the corresponding noise level in RMS electrons at the camera output in the dark and at zero integration time. The main contribution to noise reading normally comes from the on-chip preamplifier, and this noise is applied equally to each image pixel [30]. This buildup is different for a CMOS sensor and a CCD sensor.

3.4 Limitations

Optical cameras can provide high-definition images. However, they can get costly, require considerable data processing, and are unable to provide range detail. Depending on the application, extreme weather conditions, the need for substantial data processing capacity, and expense will all hinder the use of cameras as vision sensors. The following chapter discusses some of the camera's limitations when used in an autonomous driving environment.

4. Traffic Sign Detection and Recognition

Traffic Sign Detection and recognition (TSDR) is an essential part of the ADAS. It is specifically designed to work in a real-time environment through the quick acquisition and analysis of track signs to increase driver safety. Traffic sign detection is conventionally classified into colour-based methods, shape-based methods and hybrid methods (colour-shaped methods) [50]. In the case of unmanned vehicles and the driving assistance systems, the safety issue is often the highest priority relative to the comfort or practicality of them. The key aim of driving assistance system (DAS) is to gather valuable insights for drivers in order to minimize their effort in safe driving. Drivers must pay attention to different factors, including vehicle speed and orientation, distance between vehicles, moving traffic, and potentially dangerous or unexpected accidents ahead. If these systems are able to gather such information beforehand, it can substantially reduce the driving pressure on drivers and make driving safer and simpler.

Road signs are placed to direct, warn and control traffic. They offer guidance to help drivers run their cars in a manner that assures the traffic safety. The difficulty in recognizing these signs can be largely due to fading of colors, outdoor lighting conditions, obstacles or weather conditions like rain, fog etc. A vision-based system for the detection and recognition of road signs is therefore desirable to attract the attention of the driver in order to avoid traffic hazards. Computer vision devices with the advantage of high resolution can be used to identify and distinguish road boundaries, barriers and signs. Vision technologies using visual sensing devices such as cameras have been used in a wide range of applications, such as identification, classification, navigation, monitoring and control. For the purpose of driver assistance, vision systems have been used to detect, distinguish and record items such as road signs and road signals. Generally, in a camera-based system, spatial and temporal knowledge of dynamic scenes is derived from video input sequences, and noise is then filtered out [43].

In road sign recognition, color is a local feature that can be derived from a single pixel. On the other hand, shape is a global feature, and must be determined by a neighborhood of pixels. Detection of road signs is very challenging in bad weather conditions due to the effect of constantly varying outside illumination. While the real colors of the road signs are initially very well regulated, the apparent colors are influenced by the lighting of different colors in their natural settings. Moreover, with the effects of sunshine, coloring on signs also fades away with time. The hue component of the HSI (hue, saturation, and intensity) model is invariant to light and shadow [44]. The hue aspect is also appropriate for the extraction of color characteristics, considering the volatility of the weather and the natural and artificial damage to road signs.

In a camera-based system, the most conventional detection method uses color and shape features to locate the positions of traffic signs in a single frame. The shape feature is the character of contour, which shows the contrast between the object and the background. The shape feature is also more robust compared to the color information since its invariant to changing light conditions. In addition, when the resolution of the traffic signs is minimal, the connected region of homogeneous colors is divided up by noise. Therefore, the shape feature is introduced as the initial step in detecting the traffic sign. Then using the color feature to review the detection results of the first stage.

4.1 Shape Detector

The most common approach used for shape-based identification is Hough Transformation (HT) and its derivatives [46]. Hough transform is a method of extraction of features used in image recognition, computer vision, and digital image processing. The purpose of the technique is to locate imperfect instances of objects within a certain class of shapes by means of a voting process [45]. The shape detector locates the area of the spherical object using the center and radius of the object. Any other circular objects are also observed, such as a car tire, which is considered a “false” positive candidate. For instance, if the sign is circular, it operates on the gradient of the image and uses the existence of the shapes that vote the center point for the circular sign. The center of the circular object is identified by a threshold of the total of all the voting outcomes at various radii [46]. And all the voting values of the detected center are tested at various radii, and the resulting radius of the maximum vote is the radius of the circular object.

4.2 Color Detector

The color detector generally consists of a segmentation stage by setting the threshold for a given color space to extract the color from the image [47]. The state of the lighting varies with different time and weather outside, so the color detector must be invariant to the change in illumination. Color information is useful in minimising the number of early mentioned false positive candidates. Traditionally, digital color cameras use a Bayer filter on its sensors. Color information for one pixel is expressed by the intensity of the Red, Green and Blue (RGB) elements. In reality, objects may be assumed to have the color of the light leaving their surfaces.

Considering the change of illumination influences the intensity at each wavelength, but does not affect the ratio of the intensity at each wavelength, the color value of the camera sensor varies linearly with the change of the illumination in the RGB color space [48]. This characteristic can be used to build a color space based on RGB and can be expressed by a set of equations:

$$Angle(R) = R/\sqrt{R^2 + G^2 + B^2} \tag{19}$$

$$Angle(G) = G/\sqrt{R^2 + G^2 + B^2} \tag{20}$$

$$Angle(B) = B/\sqrt{R^2 + G^2 + B^2} \tag{21}$$

$$Angle(R)^2 + Angle(G)^2 + Angle(B)^2 = 1 \tag{22}$$

4.3 Challenges in Recognition

While the traffic signals have been designed for fast and simple comprehension by humans, they are not so readily identifiable by the computer. Traffic signals are flat objects with simple shapes, colors and pictograms. They might seem easy to solve even from the point of recognition area. However, there are numerous challenges that make it impossible to identify road signs. Few of the most common ones are discussed below.

(a) Video Source (Camera)

Recognition depends on the quality of the image sensor (CMOS/CCD) and the image output format. Color or gray cameras may be used for different resolutions, configurations, compression speeds, etc. Issues can occur not only from setting the camera, but also if the camera is not correctly mounted in the car, so that vibration and blur may appear in the video sequences. The focus of the camera should also be set to infinity with the autofocus turned off to avoid negative adjustments of the focus.

(b) Lighting and Weather Conditions

There are variations in the acquisition of images by daytime and darkness, even by the effect of the light source. Thus the shade of the colors of the objects can be seen distinctly from the variations of the lighting. Issues often inevitably lead to reflection from some light source, such as sunshine in the daytime or street lights in the night. The captured image is also influenced by rain, snow or fog. For example, road signs can be covered in snow or poorly visible in the fog as seen in Figure 9.

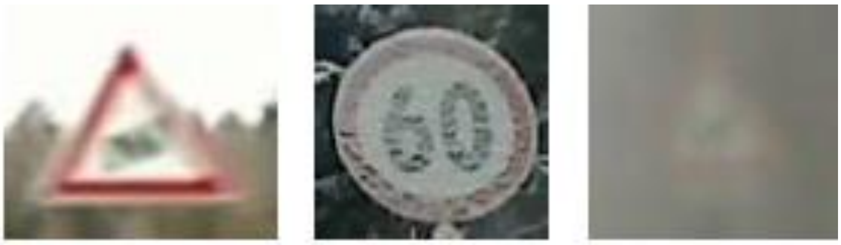


Figure 9. Traffics signs in different weather conditions [49]

(c) Occlusion and Damage

All kind of objects that obstruct the surface of the road signs, such as trees, cars, pedestrians, poles or objects on the road. Shadows may cause another particular occlusion. The traffic sign will then alter its meaning, e.g. the shadow from the power line to the priority road sign can be observed as the end of the priority road. Traffic signals can be affected not only by sunshine, but also by graffiti or weather over time (strong breeze, storm, raining). They can be dusty, scribbled, tilted, rusty, etc.

(d) Scene Complexity:

Multiple traffic signals may appear on the traffic scene to be identified in the image impacts, resulting in an increase in computational complexity and thus a decrease in real-time processing. Cascading of traffic signs having multiple signs placed side by side will also increase the scene complexity by appearing as one.

(e) Unavailability of Public Database:

A database is a key requirement for developing any TSDR system. It is used for training and testing the detection and recognition methods. The lack of big, well structured, and free public image databases is one of the challenges facing this area of research. For instance, only 600 training images and 300 assessment images are included in the most widely used database (GTSDB database). Of the seven categories classified in the Vienna Convention, only three categories of track signs for identification are protected by the GTSDB: prohibitive, obligatory and harmful [51]. To resolve the database scarcity problem, one of the ideas is to create a unified global database containing a large number of images and videos for road scenes in various countries around the world.

Detection and recognition of traffic signals are caught by the performance of the system in real-time. Precision and speed are certainly the two major criteria required for practical applications. A system with elegant algorithms and powerful hardware is needed to achieve these requirements. Convolutional Neural Network (CNN) based learning approaches with GPGPU technologies are a good alternative [52]. It is difficult to tackle the issue of traffic signal detection very well in terms of different lighting, motion blur, occlusion, and so on. More efficient and stable methods therefore need to be established.

5. Experiments

This chapter deals with the methodology, equipment and data processing used for the experiments performed on the images captured in various traffic and light conditions. The photos investigated were captured using a semi-professional DSLR camera with a maximum resolution of 5184 x 3456 pixels, with an effective resolution of 18,1 megapixels. An EF (Electro-focus) prime lens with 50 mm focal length and maximum aperture of f/1.8 was used on the device during the experiment. The lens also comes with a focusing distance of 0,35 meter with a maximum magnification of 0,21x. These images were then analyzed using a deep learning based object recognition algorithm known as YOLO, which stands for "You Only Look Once".

5.1 Experiment Design

The experiment was designed to determine the minimum criteria for image quality while preserving the relevant object detection standard, with an emphasis on quantitative and qualitative measurement data analysis. Two different approaches were employed within various ways attempt to discover the algorithm's limit. The first approach was to limit the images to three different resolutions: 1280 x 960 pixels, 640 x 480 pixels and 320 x 240 pixels before running the detector. The second approach was to convert the smallest sized image to black and white after decreasing the saturation of all common colors except green and then, validate the algorithm's recognition variations. These two approaches were proposed while keeping environmental hazards in mind.

The artificial-intelligence industry is often compared to the oil industry: data, like oil, can be a highly profitable asset once mined and refined. Deep learning, like its fossil-fuel analog, has a massive environmental effect. The model training method for natural-language processing (NLP), a branch of artificial intelligence that focuses on teaching computers to understand human language has achieved many notable success achievements in machine translation, sentence completion, and other common benchmarking activities over the last two years [67]. However, such advancements have compelled training ever larger models on massive data sets of sentences scraped from the internet. The method is computationally expensive as well as extremely energy consuming.

The first explanation on how this machine learning models consume energy is that the datasets used to train these models are becoming increasingly large. After being trained on a dataset of 3 billion words, the BERT (Bidirectional Encoder Representations from Transformers) model, a Transformer-based machine learning technique created by Google for natural language processing pre-training achieved best-in-class NLP results in 2018. Later in 2020, Generative Pre-trained Transformer 3 (GPT-3) an autoregressive language model that uses deep learning to generate human-like text was trained and published using a weighted dataset of approximately 500 billion words, which dwarfed all the previous attempts [66]. On each piece of data, they are fed during preparation, neural networks perform a lengthy series of mathematical operations (both forward and back propagation), updating their parameters in complex ways. As a result, larger datasets translate to increased compute and energy requirements.

The intense experimentation and tuning needed to build a model is another aspect behind AI's significant energy consumption. Today, machine learning is largely an experiment in trial and error. During training, practitioners will often create hundreds of iterations of a given model, experimenting with various neural architectures and hyperparameters before finding an optimal configuration. Researchers from the University of Massachusetts, Amherst, conducted a life cycle evaluation for training many common big AI models in a paper released on June 5th, 2019. The study found that training big data models like BERT, produces significant carbon emissions, roughly equivalent to a trans-American flight for one person [64]. Table 1 shows a typical carbon footprint benchmark of CO2 emission, demonstrating that training a single 213 million parameters NLP deep-learning model with an architecture search currently generates the same carbon footprint as five American cars, including gasoline.

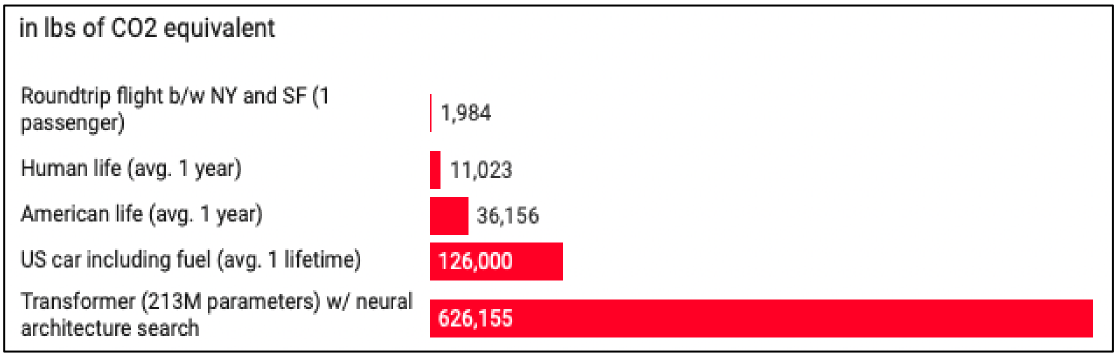


Table 1. Common Carbon footprint benchmarks [65]

The process of training these machine learning models is just the beginning of a model's lifecycle. Once a model has been trained, it is used in the real world. The method of deploying AI models to take action in real-world environments, known as inference, requires much more energy than training does. Nvidia predicts that inference, rather than training, accounts for 80% to 90% of the cost of a neural network [66]. Considering the artificial intelligence (AI) at the heart of a self-driving car, to learn to drive, neural networks must first be trained. Once training is completed and the autonomous vehicle is deployed, the model then continuously infers in order to control the environment—nonstop, day after day, for as long as the vehicle is in operation. And greater the number of parameters in these models, the greater the energy needs for this ongoing inference.

The second approach, which consisted of converting the image to black and white while decreasing all of the prominent colors except green, was also a step towards reducing the image size before running the algorithm. Digital images are computer-stored electronic copies of images. They are simply a group of numbers on a computer's hard drive that identify the individual elements of an image and how they are organized. Pixels (short for “picture elements”) are these elements, and they are arranged in a grid pattern, with each pixel providing details about its color or intensity. When an image is viewed on a computer screen or printed on paper, its actual dimension is determined by two factors: image size and image resolution. The image size corresponds to the number of pixels of an image, which is measured by the number of pixels between the image's horizontal and vertical sides, such as 640 x 480 pixels. The density at which pixels are displayed is referred to as image resolution, i.e., how many pixels are displayed per inch of screen or paper. This is often expressed as dots per inch, or dpi, pixels per inch, or ppi, which is a more precise expression [68]. Figure 10 shows this concept.

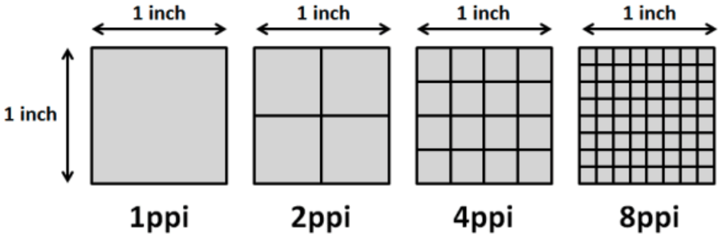


Figure 10. Pixels per inch [71]

If it is in BW (black and white), a pixel contains 8 bits (1 byte). For colored pictures, it employs the RGB (Red, Green, Blue) color scheme, which is represented as 1 byte each or 24 bits (3 bytes) per pixel. This is often referred to as an image's bit depth [69]. The bit depth is determined by the number of bits used to identify each pixel. Each color has a differing degree calculated by exponential values ranging from 256 colors for 8-bit images and 16,777,216 colors for 24-bit (3 bytes) images [69, 70]. As a result, a bit depth of 24 bits reflects 16.7 million tonal color representations. Image quality is simply the width (W) and height (H) of an image measured in pixels. The image size from the resolution can then be determined as:

$$\frac{(\mathrm{W\ x\ H\ x\ BitDepth})}{8\ \mathrm{bits/\ bytes}} = (\mathrm{W\ x\ H\ x\ BitDepth})\ \mathrm{x\ 1\ byte/8bits} \qquad (23)$$

According to the aforementioned formula, a color image with a resolution of 1280 x 960 pixels would be roughly 3,68 MB, while the same image with resolution of 640 x 480 pixels and 320 x 240 pixels would be only equivalent to less than one MB, with sizes 0,92 MB and 0,23 MB, respectively. These same images of equivalent resolution in black and white will be even smaller with 1,22 MB, 0,30 MB, and 0,076 MB in size, respectively. These calculations are included in the Appendix B. Based on the dimensions, these image sizes are an approximation. They can differ from image to image based on the color, depth, and luminance characteristics. Even if the bit depth is 24, for example, not all of those bits will exhibit a uniform tone or color, but rather gradients of the RGB color spectrum's range.

5.2 Method of Measurement

For the experiment, the camera was set to exposure bracketing mode, which captured a sequence of images at various exposures with a single click. This function aids in the making of photographs with limited adjustments to the context in various light settings. A total of eight separate locations inside Prague city were selected, each with its own set of traffic complications. These locations were photographed both during the day and night. All the photographs were shot handheld at standard car dashboard height, including the nighttime images of the eight-stop route. These eight sites featured a variety of traffic situations such as trams, motorbikes, pedestrian crossings, parked cars, and dead ends. The detection evaluation was carried out on all the eight locations, however, the photographs from only three of the eight locations were used in the results and discussion section of the experiment, and the others, along with the nighttime images, are included in the Appendix A.

5.3 Method of Data Processing

Object detection is a computer vision activity that involves both localizing and classifying one or more objects within an image [53]. It is a complex computer vision process that involves both efficient object localization (finding and drawing a bounding box around each object in an image) and object classification (predicting the correct type of object that was localized). Table 2 demonstrates the distinction between localization and classification.

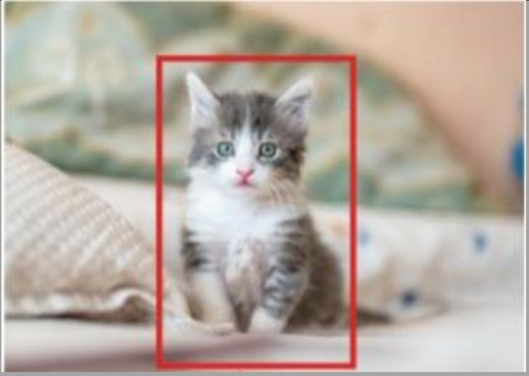
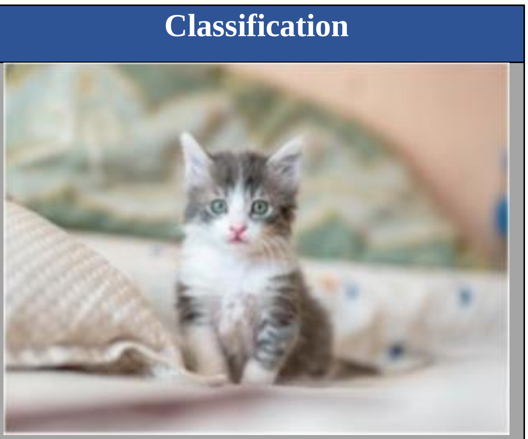
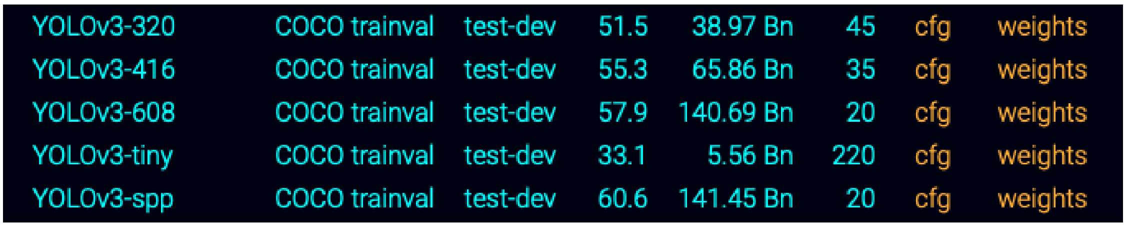
Localization	Classification
 *Figure 11. Here is the CAT [73]*	 *Figure 12. This is an image of CAT [73]*

Table 2. Localization vs Classification

5.3.1 YOLO for Object Detection

"You Only Look Once", or YOLO, family of models is a collection of end-to-end deep learning models created by Joseph Redmon for fast object detection. The method employs a single deep convolutional neural network (originally a version of GoogLeNet, later revised and named DarkNet) that divides the input into a grid of cells, each of which predicts a bounding box and object classification directly [53]. As a result, a large number of candidates bounding boxes are produced, which are then combined into a final prediction by a post-processing stage. At the time of this writing, there are three major versions of the approach: YOLOv1, YOLOv2, and YOLOv3. The first edition proposed the general architecture, the second improved the concept and used predefined anchor boxes to improve bounding box proposal, and the third refined the model architecture and training mechanism even further. The model's accuracy is similar, and they are popular for object detection due to their detection speed, which is often demonstrated in real-time on video or with camera feed input. When trained and evaluated on the same set of images, YOLO outperforms detection approaches like R-CNN (Region Based CNN) by a wide margin, and since it has a general idea of the objects detected, it is less likely to fail when exposed to new images or inputs.

All of the appropriate versions of YOLO, as well as its pre-trained weights and configurational files, can be found in J Redmon's original repository and on his website. This website compares each version based on its accuracy and frame per second (FPS) rate, and depending on the version, there are pre-trained files and guidance on how to access and run it on one's device. Since YOLOv3 outperforms its predecessors in terms of speed and precision, it was chosen as the solution for this experiment.



YOLOv3-320	COCO trainval	test-dev	51.5	38.97 Bn	45	cfg	weights
YOLOv3-416	COCO trainval	test-dev	55.3	65.86 Bn	35	cfg	weights
YOLOv3-608	COCO trainval	test-dev	57.9	140.69 Bn	20	cfg	weights
YOLOv3-tiny	COCO trainval	test-dev	33.1	5.56 Bn	220	cfg	weights
YOLOv3-spp	COCO trainval	test-dev	60.6	141.45 Bn	20	cfg	weights

Figure 13. Different versions of YOLO with their specifications [72]

❖ Network Architecture

The network architecture of YOLO can be explained from a high-level diagram as shown in Figure 14. The system is split into two main components: The Feature Extractor and the Detector, all of which are multi-scale. When a new image is sent, it is first processed by the feature extractor so that we can achieve feature embeddings at three (or more) different scales. These features are then fed into one of three (or more) branches of the detector to obtain bounding boxes and class information [54].

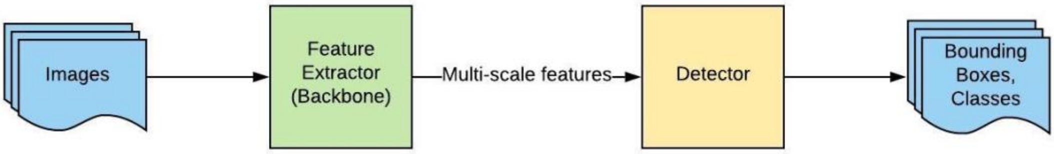
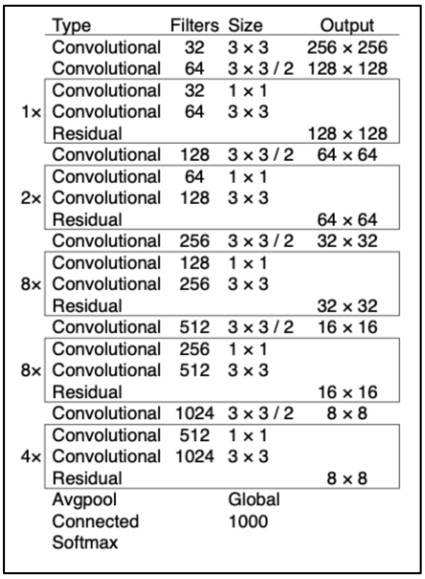


Figure 14. Network Architecture [54]

i. Feature extractor

The feature extractor YOLOv3 uses the early mentioned single deep CNN named Darknet-53. Darknet-53 employs the idea of skip connections to help activations propagate through deeper layers without gradient loss, successfully expanding the network from 19 to 53 layers, opposed to previous YOLO models.



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 3. Darknet-53 [56]

This structure is demonstrated in Table 3. Skip connections in deep architectures, as the name implies, skip several layers in the neural network and feed the output of one layer as the input to the subsequent layers (instead of only the next one) [55].

ii. Multi-scale Detector

To extract features, YOLOv3 employs successive 3x3 and 1x1 convolutional layers, as well as the Residual Networks concept [58]. Residual blocks are essentially a subset of highway networks that lack gates in their skip connections. In essence, residual blocks enable memory (or information) to flow from the first to the last layers [57]. YOLOv3 contains 5 residual blocks. Each residual block is made up of many residual units. Figure 15 depicts the structure of one residual unit.

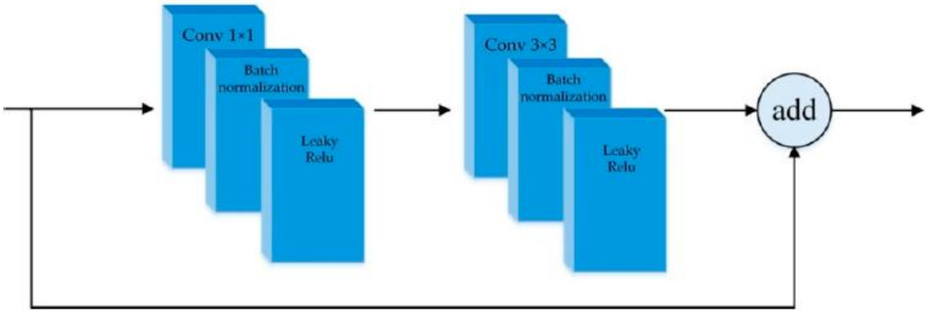


Figure 15. The structure of Residual unit [59]

The residual unit allows the network depth to be increased while avoiding gradient fading. Consider each rectangle's layers from Table 3 to be a residual block. To minimize dimension, the entire network is a chain of many blocks intermixed with several strides and two convolution layers. There is only a bottleneck structure (1x1 followed by 3x3) and a skip relation within the block. Since YOLOv3 is intended to be a multi-scaled detector, it needs features from multiple scales in addition to the detection head appended to the feature head. As a result, features from the last three remaining blocks are all included in the subsequent identification.

YOLOv3 detects targets on three different scales using feature maps. A CNN's feature maps capture the outcome of adding filters to an input image. In other words, the feature map is the output of each layer. The aim of visualizing a feature map for a given input picture is to achieve a better understanding of the features that the CNN detects [60]. The input image is down sampled five times. The targets in the last three down sampled layers are predicted by YOLOv3. To detect small targets at scale 3, the feature map is down sampled by 8. The feature map down sampled by 16 is used to detect medium-sized targets at scale 2. At scale 1, a function map that has been down sampled by 32 is used to detect large targets [59]. Since a small feature map provides deep semantic knowledge and a broad feature map provides finer-grained information about the targets, feature fusion is used to detect them. To perform feature fusion, YOLOv3 resizes the deeper layer's feature maps by up sampling. The feature maps would then be the same size at different scales. YOLOv3 concatenates the features from the earlier layer with the features from the deeper layer. As a result, YOLOv3 performs well in detecting both big and small targets.

iii. Anchor Box Prediction

Anchor boxes, which are used in Faster R-CNN (Region Based Convolution Neural Network), were added in YOLOv3. Anchor boxes are a set of initial candidate boxes that have a specified width and height. The initial anchor box selection has a direct impact on detection precision and detection time. Instead of manually selecting anchor boxes, YOLO v3 uses K-means clustering on the dataset to identify good priors [59].

K-means clustering is a basic and widely used unsupervised machine learning algorithm. The goal of K-means is simple: group related data points together and uncover underlying patterns. K-means seeks a fixed number (k) of clusters in a dataset to accomplish this objective [62]. A cluster is a group of data points that have been aggregated together due to such similarities. K-means clusters will represent the distribution of samples in each dataset, making it simpler for the network to make good predictions.

The goal of object detection is to obtain a bounding box as well as the class of the object. The bounding or anchor box is usually represented in t_x, t_y, t_w, t_h format. If the cell is offset from the top left corner of the image by (c_x, c_y) and the bounding box prior has width and height p_w, p_h, then the predictions correspond to [56]:

$$b_x = \sigma(t_x) + c_x \tag{24}$$

$$b_y = \sigma(t_y) + c_y \tag{25}$$

$$b_w = p_w e^{t_w} \tag{26}$$

$$b_h = p_h e^{t_h} \tag{27}$$

During training the sum of squared error loss is used. Ground truth is a term used in statistics and machine learning to refer to the process of comparing machine learning outcomes to the real world. If the ground truth for any coordinate prediction is $\widehat{t_*}$, then the gradient is the ground truth value (calculated for the ground truth box) minus the prediction: $\widehat{t_*} - t_*$ [56]. By inverting the equation above, this ground truth value can be effectively coupled.

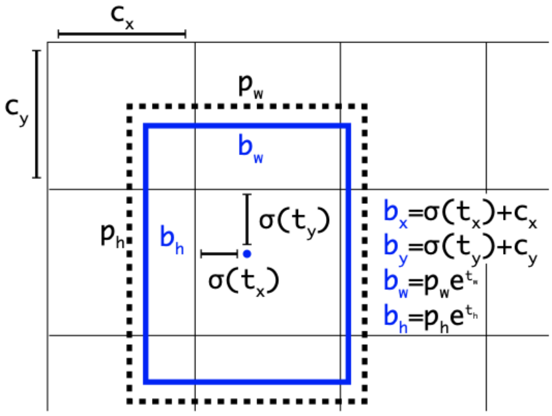


Figure 16. Bounding boxes with dimension priors and location [61]

YOLOv3 uses logistic regression to estimate an objectness score for each bounding box. This should be 1 if the bounding box prior overlaps a ground truth object by more than any other bounding box prior [56]. If the bounding box prior is not the strongest but does overlap a ground truth object by more than any threshold, it neglects the prediction. Generally, it uses a threshold of 0,5. The method only assigns one bounding box prior to each ground truth object. If a bounding box prior is not allocated to a ground truth object, there is no loss of coordinate or class predictions; only objectness is lost.

iv. Class prediction

Using multi label classification, each box predicts the groups that the bounding box will contain. Yolov3 employs logistics regression classifier in order to classify the objects. Logistic Regression is a ‘Statistical Learning' approach that falls in the category of ‘Supervised' Machine Learning (ML) techniques that are used for ‘Classification' tasks. Logistic regression is a classification algorithm that is used when the target variable's value is categorical in nature. It is most widely used where the data in question has a binary output, that is, when it belongs to either of two classes or is either a 0 or 1. The results of such an algorithm in a classification task fall into one of many pre-determined classes [63]. When given multiple input variables, the classification model tries to predict the output value, categorizing the case. Provided a particular dataset comprising various classes of objects, it may predict whether or not an object belongs to that specific group. At the time of experiment, there are two pre-set weight files, yolov3 and yolov3-tiny, ready for use. Since, yolov3-tiny is generally used for detection of smaller objects, for the experiment, only the yolov3.weight file was used. yolov3 file is the most recent and provides the necessary weights, as well as a model that has already been trained on the COCO dataset, which contains 80 classes, which includes the requisite objects to be observed and more.

5.4 Results and Discussion

The findings of the experiments on the investigated images are discussed in this section using the two image processing methods described in Section 5.1. The findings are divided into two sections: daytime and nighttime, with both approaches (reducing image size and rendering them black and white) applied to each. Only the normally exposed photographs were converted to black and white during the analysis.

A qualitative and quantitative analysis of the photographs was then carried out to determine the results of the recognition of both methods. The analysis was conducted based on an ego-vehicle, which is a vehicle whose action is of primary interest in each respective scenario. From the photographer's Point of View (POV), who can be considered as a potential car in a real-world situation, my POV was regarded as the ego-vehicle’s POV for the experiment analysis.

Quantitative analysis only considers the number of observed elements, while qualitative analysis will have semantic information - for example, a weight ranging from 0 to 1 is assigned to each object in the image based on its "worth of recognition," i.e., a car in front of an ego-vehicle has weight 1, whereas a tree on the sidewalk with less meaning to the ego-vehicle's path has weight 0. Prior to conducting this qualitative study, each object is given a predefined general weight, independent of its position or importance, but based on the likely fatality risk if the object is involved in an accident. Table 4 lists the predefined weight assigned to each item in this manner.

Object	Weight
Cars	0,9
Trams	0,7
Cyclist	1
People	1
Trees	0,8

Table 4. General Weight of the Objects

These predefined weights are then adjusted based on the previously specified semantic information. During this study, variables such as driving direction, distance, and object position are also considered. In addition, when a target's evasive maneuver is taken into account, it is given a higher weight. The overall final weight of a target is then calculated based on the potential seriousness of the contact/crash, the distance from the ego-vehicle, and the time-to-collision etc.

5.4.1 Daytime

i) Location 1

The collages for the study were designed in a way that the first image is of resolution 1280 x 960 pixels, and the second image (top-right) is with a resolution of 640 x 480 pixels. The grid of four images on the bottom-right corner, including the black and white photograph, are 320 x 240 pixels in resolution.

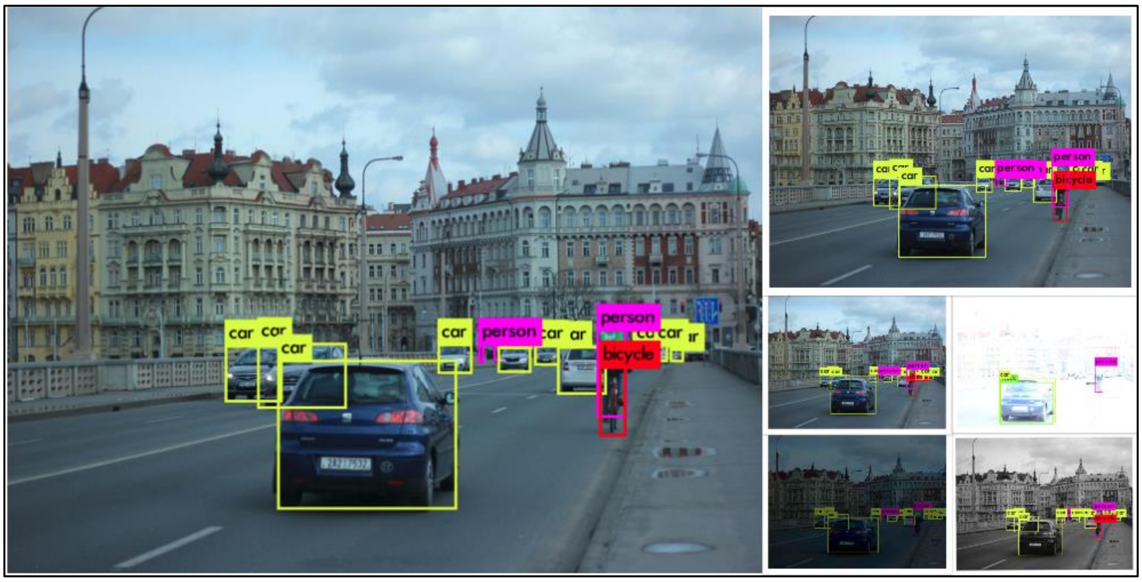


Figure 17. Jirásek Bridge [Location 1]

Figure 17 shows a location on the Jirásek bridge on a relatively cloudy day, which has four lanes, two in each direction. This location is further studied in Figure 18 by segmenting it into distinct zones, each with its own weight. We also imagine ourselves to be behind the blue car and to assess the situation as the ego-vehicle.

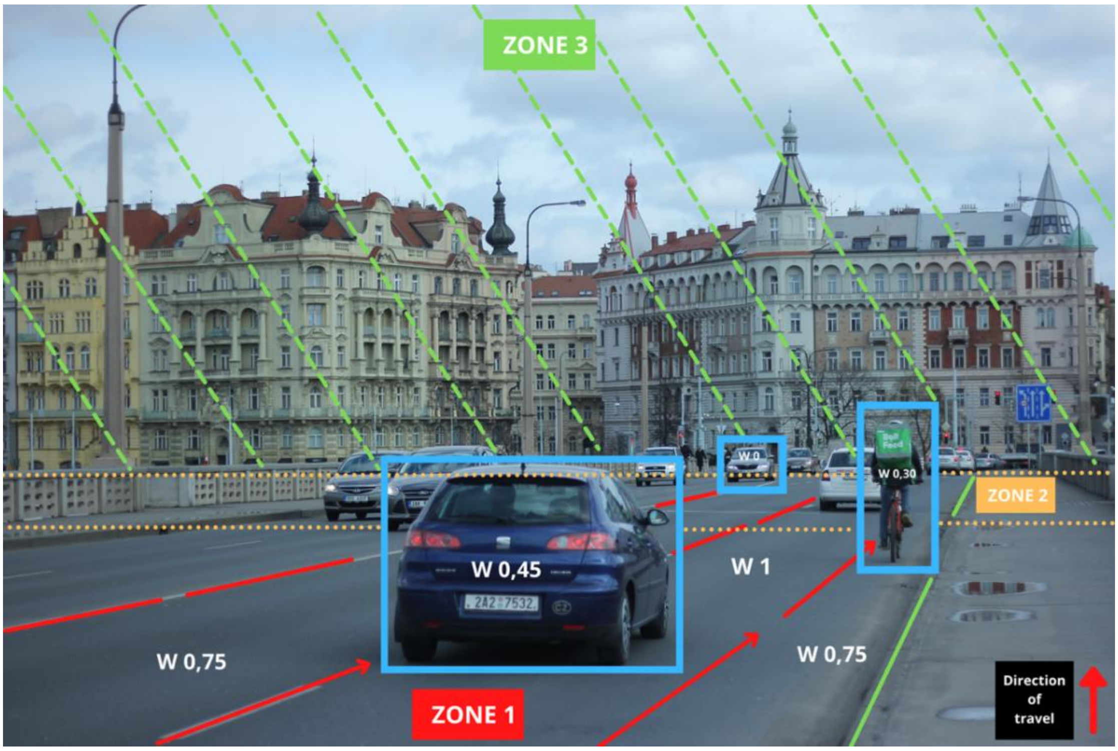


Figure 18. Weight distribution [Location 1]

Each zone is assigned a weight based on how much the targets inside these zones impact us as the ego-vehicle. Thus, in this particular scenario, zone 1 has the highest weight and is assigned a weight of 1 because any target inside this zone has a significant effect on ego-vehicle's path. Whereas zones 2 and 3 are assigned weights of 0,5 and 0,1, respectively, since they have less influence in our direction of travel. Similarly, before analyzing the detector results, a separate weight is allocated to each target based on its lane, direction, distance and time-to-collision respective to the path of the ego-vehicle, in order to compute the total weight or worth of a target's recognition. These individual weights are then multiplied together to determine the overall weight of each target. The formula would then look like:

$$\begin{aligned} Overall\ target\ weight &= Object\ weight\ \ x\ \ Lane\ weight\ \ x\ \ Zone\ weight \\ &\quad x\ Direction\ Weight\ x\ Distance\ Weight\ x\ Time\ Weight \qquad (28) \end{aligned}$$

Targets in the same direction of ego-vehicle have a direction weight of 0,5 and targets in the opposite direction have a weight of 1 since they are more prone to a dangerous collision. For example, the blue car in front of us has an object weight of 0,9 and a lane weight of 1 since it is on the same lane as our drive. Direction, distance and time-to-collision carries a product weight of 0,5 since it's the closest to us. The car is also in zone 1, so it has a zone weight of 1. As a result, the total weight of the blue car as a target will be:

$$Overall\ weight_{Blue\ car} = 0{,}9\ \cdot\ 1\ \cdot 1 \cdot 0{,}5 \cdot 1 \cdot 1 = 0{,}45$$

Similarly, for the cyclist, it is important to note that, despite the fact that cyclist is on a different lane, it has an object weight of 1 which is greater than that of the blue car, because the cyclist here is more vulnerable to a fatal accident. If the overall weight of the cyclist carrying a lane weight of 0,75 and direction – 0,5; distance – 0,9; time – 0,9 weights is then evaluated, the result is:

$$Overall\ weight_{cyclist} = 1 \cdot 0{,}75 \cdot 1 \cdot 0{,}5 \cdot 0{,}9 \cdot 0{,}9 = 0{,}30$$

To further determine the necessary weights of other targets, the location was segmentized to few more zones and a top view of this is illustrated in Figure 19. The weight of each zone is given in Table 5.

Zones	Weight
Zone 1	1
Zone 2	0,5
Zone 3	0,1
Zone 4	0,3
Zone 5	0,2
Zone 6	0

Table 5. Weights of Different Zones [Location 1]

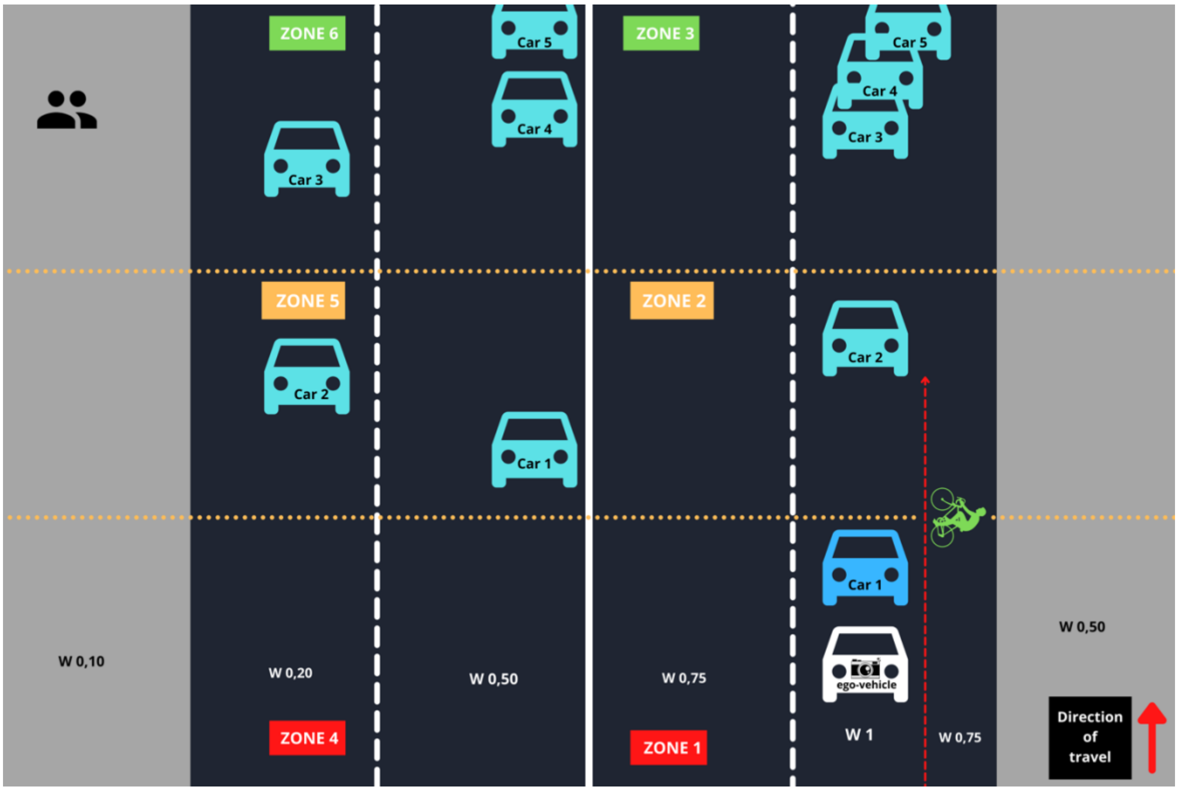


Figure 19. Top view [Location 1]

The targets in locations are split into groups of respective zones to compare the algorithm's detected effects and calculation simplicity. Table 6 shows the calculated weights of each target.

	Zone 1		Zone 2	Zone 3			
Weights	Car 1	Cyclist	Car 2	Car 3	Car 4	Car 5	
Object Weight	0,9	1	0,9	0,9	0,9	0,9	
Lane Weight	1	0,75	1	1	1	1	
Zone Weight	1	1	0,5	0,1	0,1	0,1	
Direction Weight	0,5	0,5	0,5	0,5	0,5	0,5	
Distance Weight	1	0,9	0,8	0,3	0,2	0,1	
Time Weight	1	0,9	0,8	0,3	0,2	0,1	Total Sum
Overall Weight	0,450	0,303	0,144	0,0040	0,0018	0,00045	0,9032
	Zone 5		Zone 6				
	Car 1	Car 2	Car 3	Car 4	Car 5	People on foot walk	
Object Weight	0,9	0,9	0,9	0,9	0,9	1	
Lane Weight	0,5	0,2	0,2	0,5	0,5	0,10	
Zone Weight	0,2	0,2	0	0	0	0	
Direction Weight	1	1	1	1	1	1	
Distance Weight	0,7	0,4	0,3	0,2	0,1	0,1	
Time Weight	0,7	0,4	0,3	0,2	0,1	0,1	Total Sum
Overall Weight	0,12	0,012	0	0	0	0	0,1320

Table 6. Calculated Weights [Location 1]

The final result is then calculated as the sum of weights of all the targets in each zone:

$$Total\ Weight_{location\ 1} = \sum_{i=1}^{6} Total\ Weight_{Zone\ i} \qquad (24)$$

$$Total\ Weight_{Location\ 1} = 0{,}753 + 0{,}144 + 0{,}00625 + 0 + 0{,}132 + 0 = 1{,}0352$$

❖ Detected results

When compared to the initial target count, the underexposed, naturally exposed, black and white 320 x 240 resolution images revealed no differences in detection. As a result, their total weight distribution remains unchanged. However, the overexposed image, Figure 20, showed some significant detection changes.

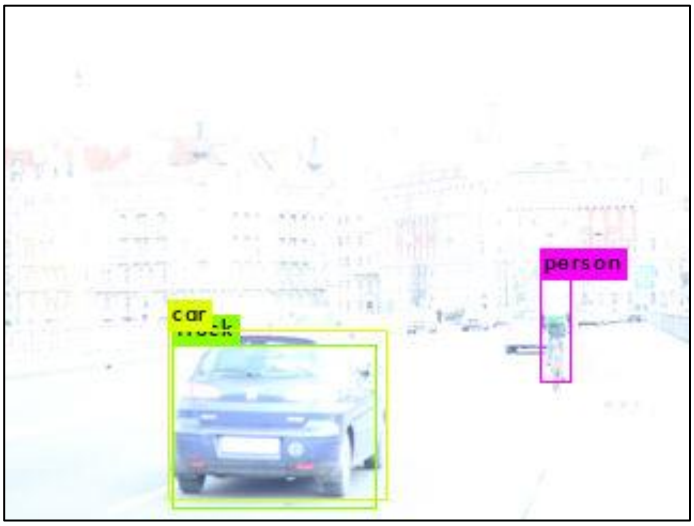


Figure 20. Overexposed 320x240 [Location 1]

The total weight distribution for this image was then calculated only using the observed target weights in the image. The overall weight distribution is depicted in Table 7.

Zone 1			
Weights	Car 1	Cyclist	
Object Weight	0,9	1	
Lane Weight	1	0,75	
Zone Weight	1	1	
Direction Weight	0,5	0,5	
Distance Weight	1	0,9	
Time Weight	1	0,9	Total Sum
Overall Weight	0,450	0,303	0,753

Table 7. Detected Weights - Overexposed [Location 1].

As seen in Figure 20, zero targets were detected in all the other zones, so the total sum of those zones is equal to zero. When the two outcomes are then compared:

a) Zone 1, Zone 2 and Zone 3:

$$Weight\ Difference_{zones\ 1-3} = Total\ Weight_{defined} - Total\ Weight_{detected}$$

$$Weight\ Difference_{zones\ 1-3} = 0{,}9032 - 0{,}753 = 0{,}1502$$

16.62% decrease ↓

b) Zone 4, Zone 5 and Zone 6:

$$Weight\ Difference_{zones\ 4-6}\ =\ 0{,}1320 - 0 =\ 0{,}1320$$

100% decrease ↓

c) Overall weight:

$$Weight\ Difference_{overall}\ =\ Total\ Sum_{defined} - Total\ Sum_{detected} \tag{25}$$

$$Weight\ Difference_{overall}\ =\ 1{,}0352 - 0{,}753 =\ 0{,}2822$$

27.26% decrease ↓

According to the final result, the margin of difference between the predefined and detected total target weight is considered high. When the overall weight of targets in the same direction decreased by 16,62 percent, the opposite direction decreased by 100 percent since no targets were detected. The overall weight of location 1 has a decrease of 27,26 percent, indicating that the detector failed to detect a considerable number of important targets in the path of ego-vehicle. It is worth noting that, the blue car and the cyclist are the most critical to the direction of the ego-vehicle, and they both were identified even though the scene complexity was compromised.

❖ Detection faults

When the detected pictures from Figure 17 are closely examined, it is clear that the algorithm did not perform well in identifying the appropriate classes for the targets under adverse lighting and image quality conditions. While the underexposed 320 x 240 resolution picture failed to recognize the bicycle, in the overexposed picture with similar resolution, it got confused whether the blue car belonged to the class car or truck. Even though the objective in both pictures is to recognize the target, detecting only the person without the bicycle or not knowing whether the target is a car or truck can lead to many confusions in a real-world fully autonomous system. The algorithm may detect it as a person who’s walking at a much slower pace, when the actual target is a cyclist who travels at a much faster rate. This might end up influencing the decision the autonomous system has to make in the event of a collision.

ii) Location 2

The location in Figure 21 has a dead zone at the end of the lane, which adds a new set of road complication for the detector. Also, there are cars parked on both sides of the lane, with a left turn being the right way of exit.

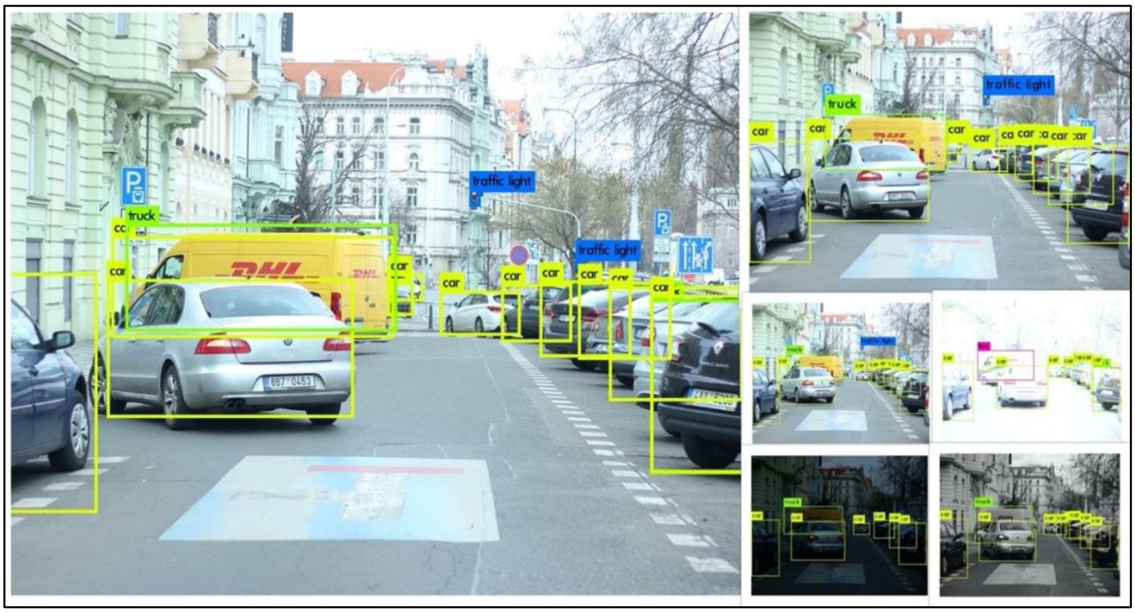


Figure 21. Janáčkovo nábř [Location 2]

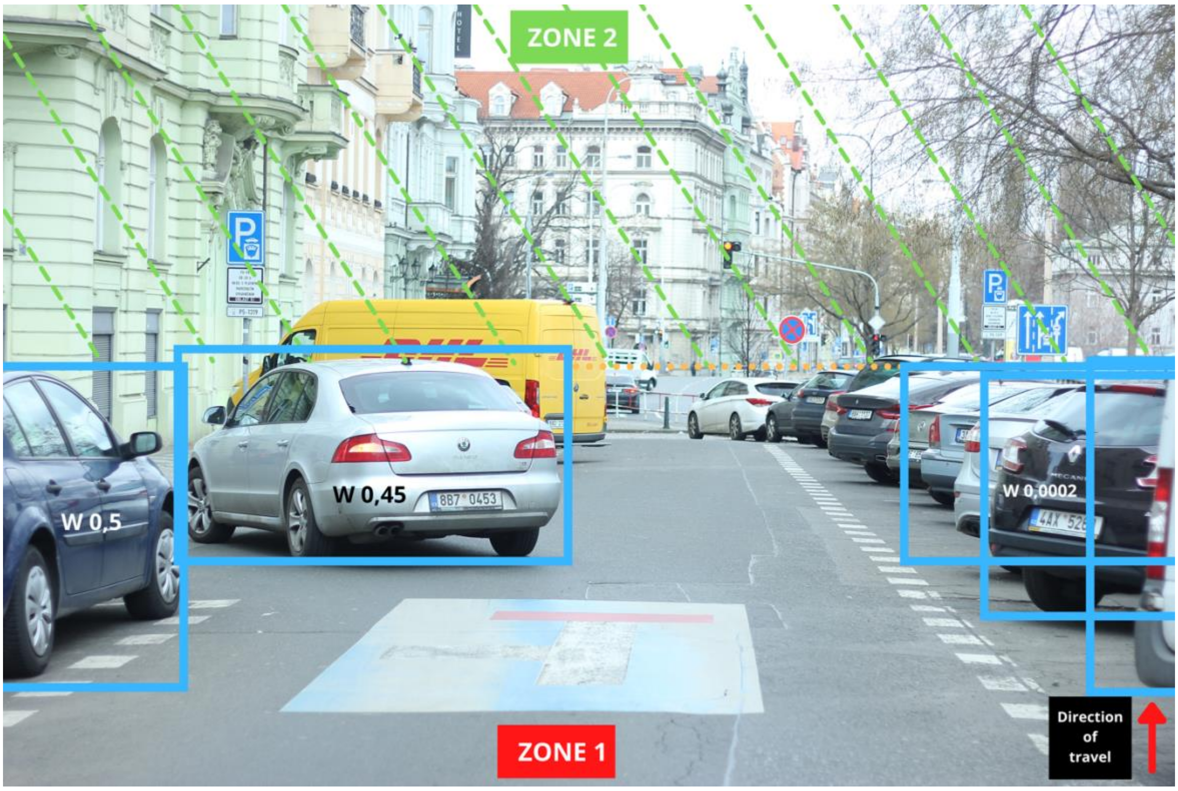


Figure 22. Weight Distribution [Location 2]

The photograph was taken on a clear day and the street has just one lane and is one way. There are no vehicles traveling from the opposite direction. We visualize ourselves as the ego-vehicle behind the grey car making the left turn. Figure 23 shows the top view of location 2.

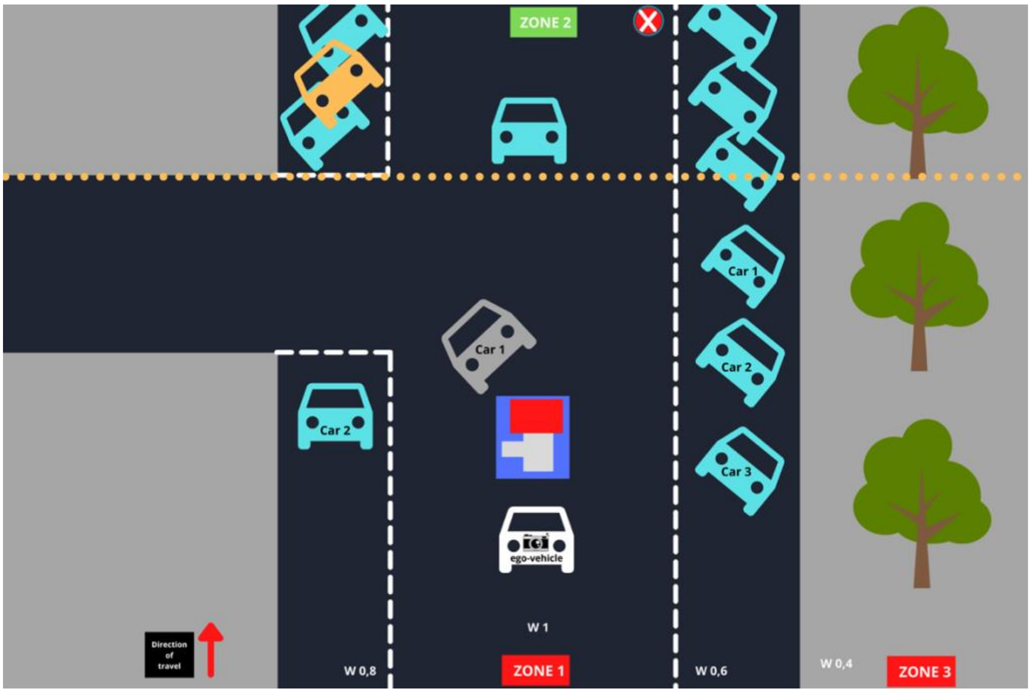


Figure 23. Top View [Location 2]

The location is split into three different zones and assigned different weights to them as shown in Table 8.

Zones	Weight
Zone 1	1
Zone 2	0
Zone 3	0,1

Table 8. Different Weight of Zones [Location 2]

Here, the zone 2 has a weight of zero since the targets in the zone have no impact on the ego-vehicle's movement, making all of the target weight in that specific zone equal to zero. The remaining targets are then split into two categories: Zone 1 and Zone 3. Table 9 shows the weight distribution of the targets in location 2.

	Zone 1		**Zone 3**				
	Car 1	Car 2	Car 1	Car 2	Car 3	Tree	
Object Weight	0,9	0,9	0,9	0,9	0,9	0,8	
Lane Weight	1	0,8	0,6	0,6	0,6	0,4	
Zone Weight	1	1	0,1	0,1	0,1	0,1	
Direction Weight	0,5	0,7	0,4	0,4	0,2	0	
Distance Weight	0,9	1	0,2	0,1	0,1	0,1	
Time Weight	1	1	0,2	0,1	0,1	0,1	Total Sum
Overall Weight	0,450	0,504	0,000864	0,000216	0,000108	0	0,9551

Table 9. Weight Distribution [Location 2]

It is important to note that how Car 2 (parked car) in zone 1 has a higher weight here than the Car 1 (moving car) in front. This is because the parked car is critical to the direction of ego-vehicle's left turn and there's a higher chance of collision than the moving car. The total weight of the targets in location 2 is then the sum of overall target weights of all the zones.

$$Total\ Weight_{location\ 2} = \sum_{i=1}^{3} Total\ Weight_{Zone\ i}$$

$$Total\ Weight_{Location\ 2} = 0{,}954 + 0 + 0{,}001188 = 0{,}9551$$

- Detected results

The detected results of location 2 did not show any significant changes in target count. The overexposed, naturally exposed, black and white 320 x 240 resolution images revealed almost no differences in detection. As a result, their total weight distribution remains unchanged. However, the underexposed image, Figure 24, failed to recognize one parked car, showing a very minor difference in overall target count. The trees are also not detected by the algorithm which is no different in case of all other exposure and size levels.

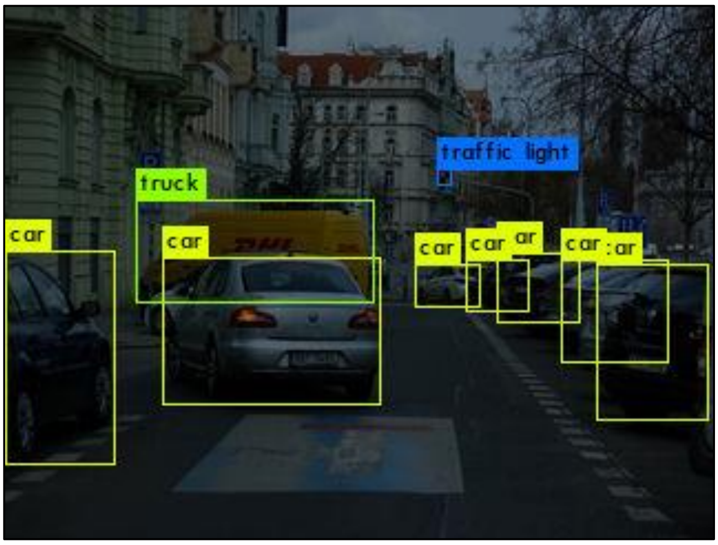


Figure 24. Underexposed 320x240 [Location 2]

It is a fair to say that the algorithm detects no trees, proving that the overall total weight calculated to zero for tree in Table 9 is appropriate. Table 10 shows the detected target results of Figure 24.

	Zone 1		**Zone 3**		
	Car 1	Car 2	Car 1	Car 2	
Object Weight	0,9	0,9	0,9	0,9	
Lane Weight	1	0,8	0,6	0,6	
Zone Weight	1	1	0,1	0,1	

Direction Weight	0,5	0,7	0,4	0,4	
Distance Weight	0,9	1	0,2	0,1	
Time Weight	1	1	0,2	0,1	Total Sum
Overall Weight	0,450	0,504	0,000864	0,000216	0,9550

Table 10. Detected Weights - Underexposed [Location 2]

$$Weight\ Difference_{overall} = Overall\ Weight_{defined} - Overall\ Weight_{detected}$$

$$Weight\ Difference_{overall} = 0{,}9551 - 0{,}9550 = 0{,}0001$$

0.01% decrease ↓

The weight gap between detected and defined is only 0.01 percent, which is almost negligible, proving that the target that wasn't recognized in Figure 24 is less significant to the ego-vehicle's course and thus, proving that the weight allocated to that target in location 2 is close to the ground truth.

❖ Detection faults

The street at Location 2 leads to a dead zone, which is indicated by two different traffic signs. One sign, as seen in Figure 25 from Google Maps, is obscured by trees, making it almost impossible for the detector or a new driver on the roadway to notice.

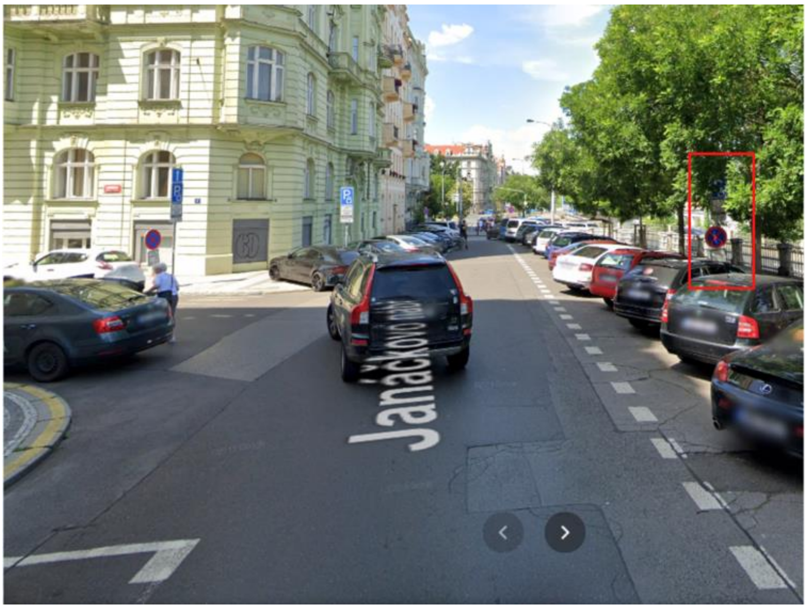


Figure 25. Traffic sign obscured by trees [19]

As a result, the second sign, which is painted on the road as seen in Figure 22, is critical for detection. The detection results of the location 2 show that the algorithm never identified the road sign. This can be for the reason that the algorithm hasn't been trained to look for traffic signs on the road, or simply because these specific signs aren't included in the dataset used for the experiment.

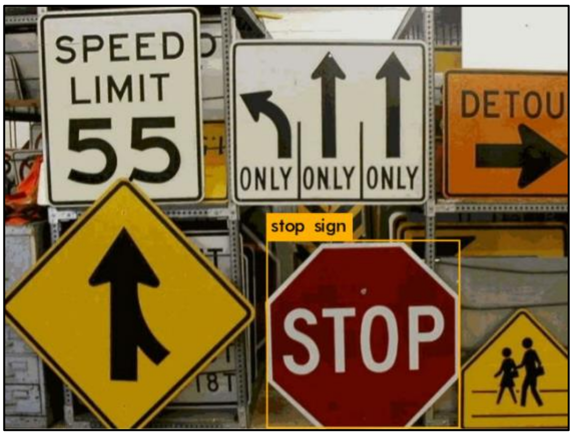


Figure 26. Detected traffic signs [23]

Figure 26 indicates that the dataset has not been trained for a wide range of road signs, with the exception of key road signs such as STOP, which are more general. This flaw or failure in traffic sign identification is critical when it comes to an autonomous vehicle system and the study of its future advancements.

iii) Location 3

This location in Karlovo námesti portrays a set of different traffic complications from location 1 and 2. The location includes a tram line next to the driving lane, parked cars, pedestrian crossing and traffic signals, providing a combination of different complexities.

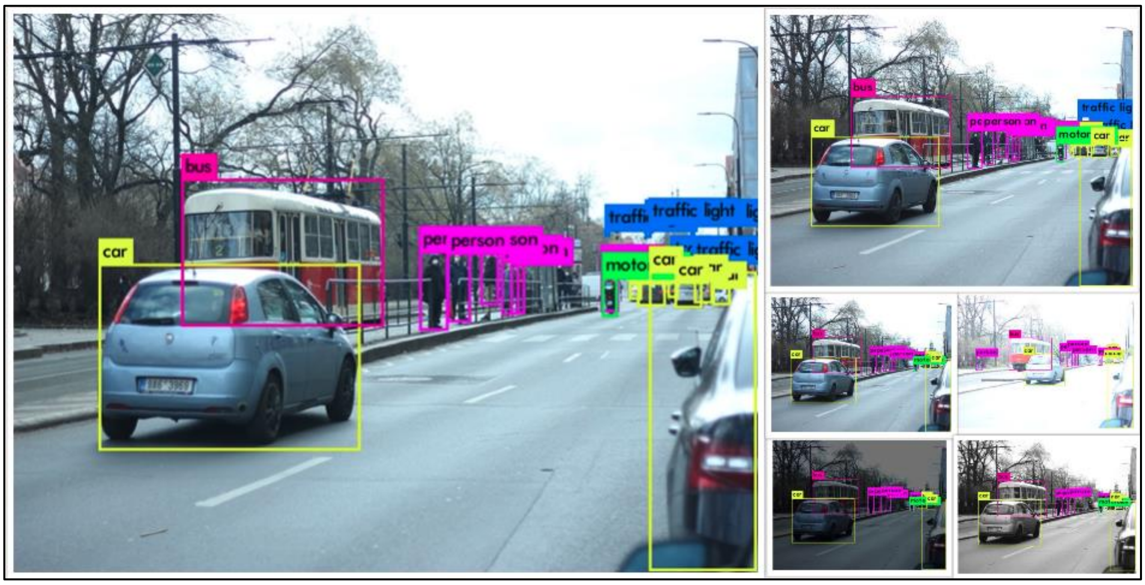


Figure 27. Karlovo námesti [Location 3]

The photo was taken on a clear day with moderate amount of traffic on the road. We visualize ourselves to be behind the light blue car driving straight the lane. Figure 28 depicts the zone segmentation from the perspective of ego-vehicle.

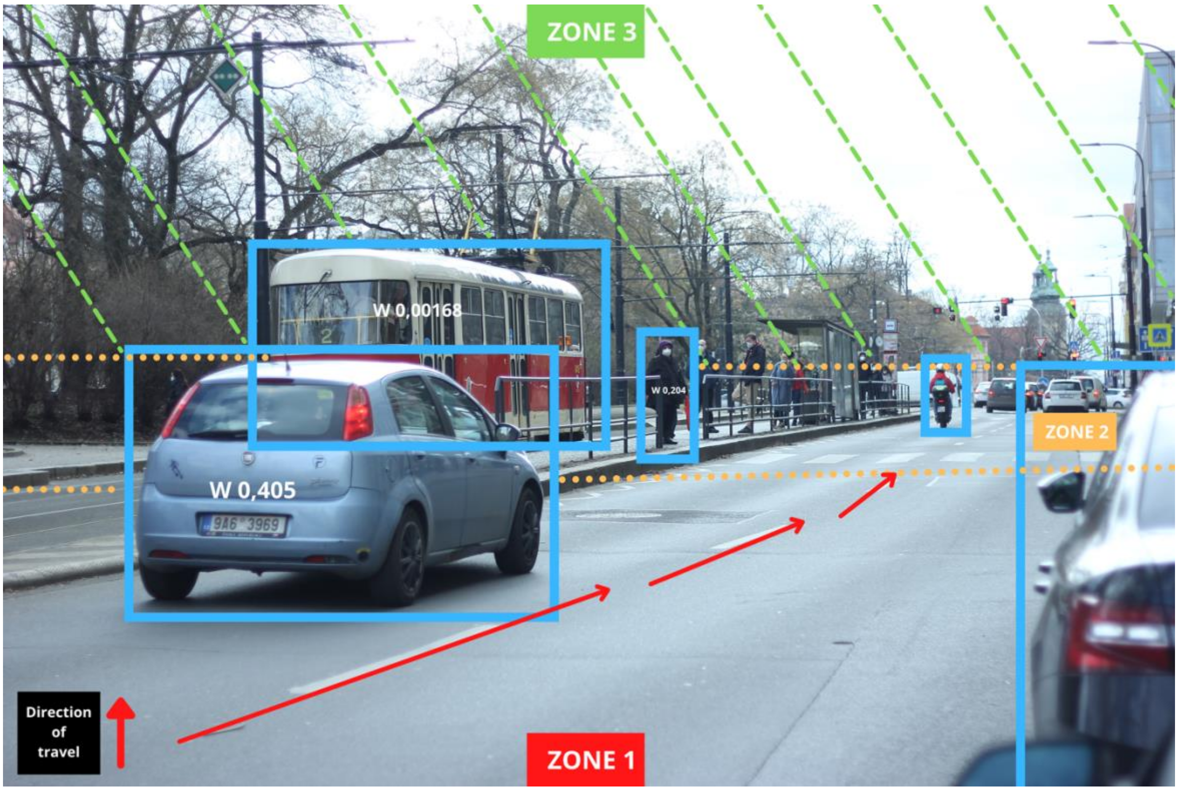


Figure 28. Weight Distribution [Location 3]

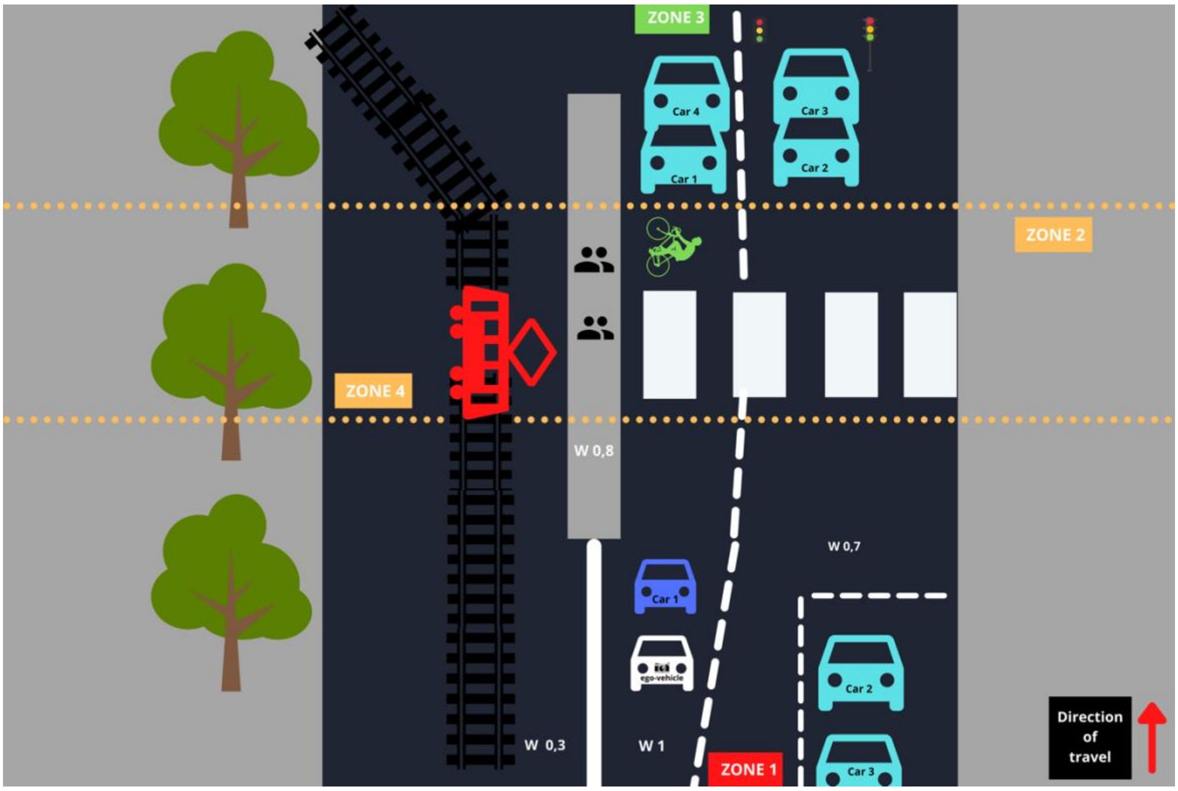


Figure 29. Top View [Location 3]

The location is split into four different zones and assigned different weights to them as shown in Table 11.

Zones	Weight
Zone 1	1
Zone 2	0,8
Zone 3	0,3
Zone 4	0,1

Table 11. Zone Weights [Location 3]

The weight distribution of the targets is shown in Table 12.

	Zone 1			**Zone 2**		
	Car 1	Car 2	Car 3	Person (Crossing)	Bike	
Object Weight	0,9	0,9	0,9	1	1	
Lane Weight	1	0,7	0,7	0,8	1	
Zone Weight	1	1	1	0,8	0,8	
Direction Weight	0,5	0,2	0	0,5	0,5	
Distance Weight	0,9	0,2	0,1	0,8	0,3	
Time Weight	1	0,2	0,1	0,8	0,3	Total Sum
Overall Weight	0,405	0,00504	0	0,204	0,036	0,4972

	Zone 4			Zone 3				
	Tram	Tree	People (waiting)	Car 1	Car 2	Car 3	Car 4	
Object Weight	0,7	0,9	1	0,9	0,9	0,9	0,9	
Lane Weight	0,3	0	0,8	1	0,7	0,7	1	
Zone Weight	0,1	0,1	0,1	0,3	0,3	0,3	0,3	

Direction Weight	0,5	0	0,3	0,4	0,5	0,5	0,5	
Distance Weight	0,4	0	0,2	0,2	0,1	0	0	
Time Weight	0,4	0	0,2	0,2	0,1	0	0	Total Sum
Overall Weight	0,00168	0	0,00096	0,00432	0,000945	0	0	0,0079

Table 12. Weight Distribution [Location 3]

It is worth noting that there are two kinds of pedestrians in this location: the pedestrian waiting to cross the street and those waiting for the tram. And they both have a different weight, with the pedestrian crossing bearing a higher weight. The total weight of the targets in location 3 is then the sum of overall weight of targets in all the zones.

$$Total\ Weight_{Location\ 3} = \sum_{i=1}^{4} Total\ Weight_{Zone\ i}$$

$$Total\ Weight_{Location\ 3} = 0{,}4972 + 0{,}0079 = 0{,}5051$$

❖ Detected results

Only the black and white image with the resolution of 320 x 240 had changes in the detection results of location 3, while the remainder of the images of varying sizes and exposure levels remained unchanged. Figure 30 shows that it failed to identify the majority of targets in Zone 3.

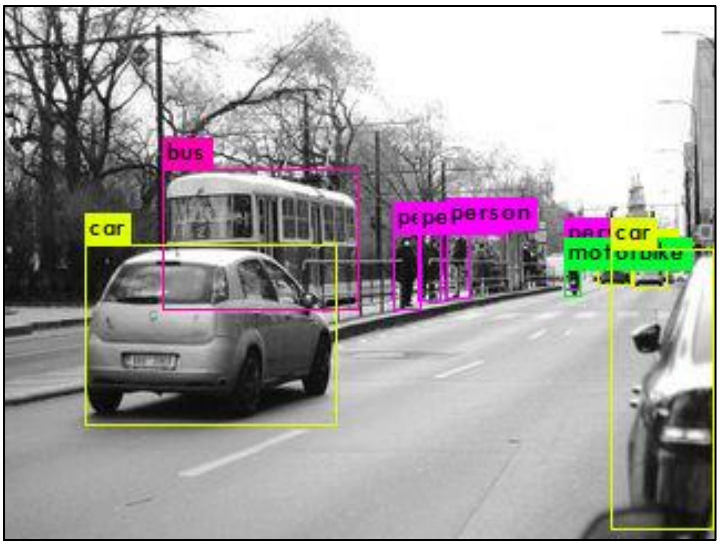


Figure 30. Black & White 320 x 240 [Location 3]

	Zone 1			Zone 2		
	Car 1	Car 2	Car 3	Person (Crossing)	Bike	
Object Weight	0,9	0,9	0,9	1	1	
Lane Weight	1	0,7	0,7	0,8	1	
Zone Weight	1	1	1	0,8	0,8	
Direction Weight	0,5	0,2	0	0,5	0,5	
Distance Weight	0,9	0,2	0,1	0,8	0,3	
Time Weight	1	0,2	0,1	0,8	0,3	Total Sum
Overall Weight	0,405	0,00504	0	0,204	0,036	0,4972

	Zone 4			Zone 3	
	Tram	Tree	People (waiting)	Car 1	
Object Weight	0,7	0,9	1	0,9	
Lane Weight	0,3	0	0,8	1	
Zone Weight	0,1	0,1	0,1	0,3	
Direction Weight	0,5	0	0,3	0,4	
Distance Weight	0,4	0	0,2	0,2	
Time Weight	0,4	0	0,2	0,2	Total Sum
Overall Weight	0,00168	0	0,00096	0,00432	0,00696

Table 13. Detected Weights - Black & White [Location 3]

$$Total\ Weight_{Location\ 3} = \sum_{i=1}^{4} Total\ Weight_{Zone\ i}$$

$$Total\ Weight_{Location\ 3} = 0{,}4972 + 0{,}00696 = 0{,}5041$$

The overall weight difference between defined and detected are then found:

$$Weight\ Difference_{overall} = Overall\ Weight_{defined} - Overall\ Weight_{detected}$$

$$Weight\ Difference_{overall} = 0{,}5051 - 0{,}5041 = 0{,}0010$$

0,197 % decrease ↓

Having weight drop of only 0.197 percent between detected and defined is again proving that, the weight allocated to the targets in location 3 is close to the ground truth and the undetected targets in Figure 28 are unimportant to the ego-vehicle's course.

❖ Detection faults

From the detected images of location 3, it is seen the algorithm classified the tram as a bus. In a real-world scenario, if an autonomous vehicle's detection system makes a similar mistake, the consequences can be severe. The tram always follows the path of its track, with a degree of freedom in one to four directions whereas, a bus can move in any direction. Because of this uncertainty, the detector can make incorrect conclusions or judgements in the case of a collision depending on the target's movement. Similarly, Figure 27 shows that the pedestrian crossing traffic sign or zebra-lines on the road was never identified. This is critical to the course of any vehicle, and it can be improved by training the algorithm with new data that includes a diverse collection of traffic signs.

5.4.2 Nighttime

i) Location 2

Location 2, Janáčkovo nábř street from the Section 5.4.1 is used for the nighttime study, and the remaining locations are added in Appendix A. The location has cars parked on both sides with a dead zone, and a left turn as the right of way exit.

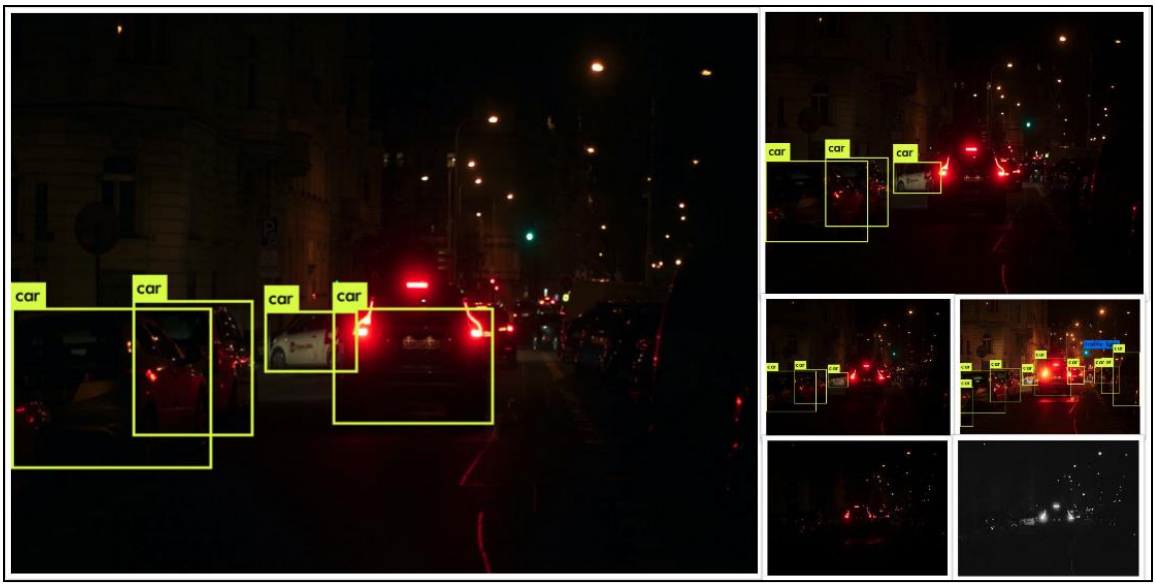


Figure 31. Janáčkovo nábř - Nighttime [Location 2]

Similarly, we consider ourselves to be the ego-vehicle behind the car making the left turn, with the similar zone weights given in Table 8. From Figure 31, it is evident that the overexposed picture shows more detail than the normally exposed image. For the same reason, to be more articulate, Figure 32 the overexposed image at nighttime was used to show the segmentized zones and weight distribution of the location, while Figure 33 represents the top perspective of Location 2 at night.

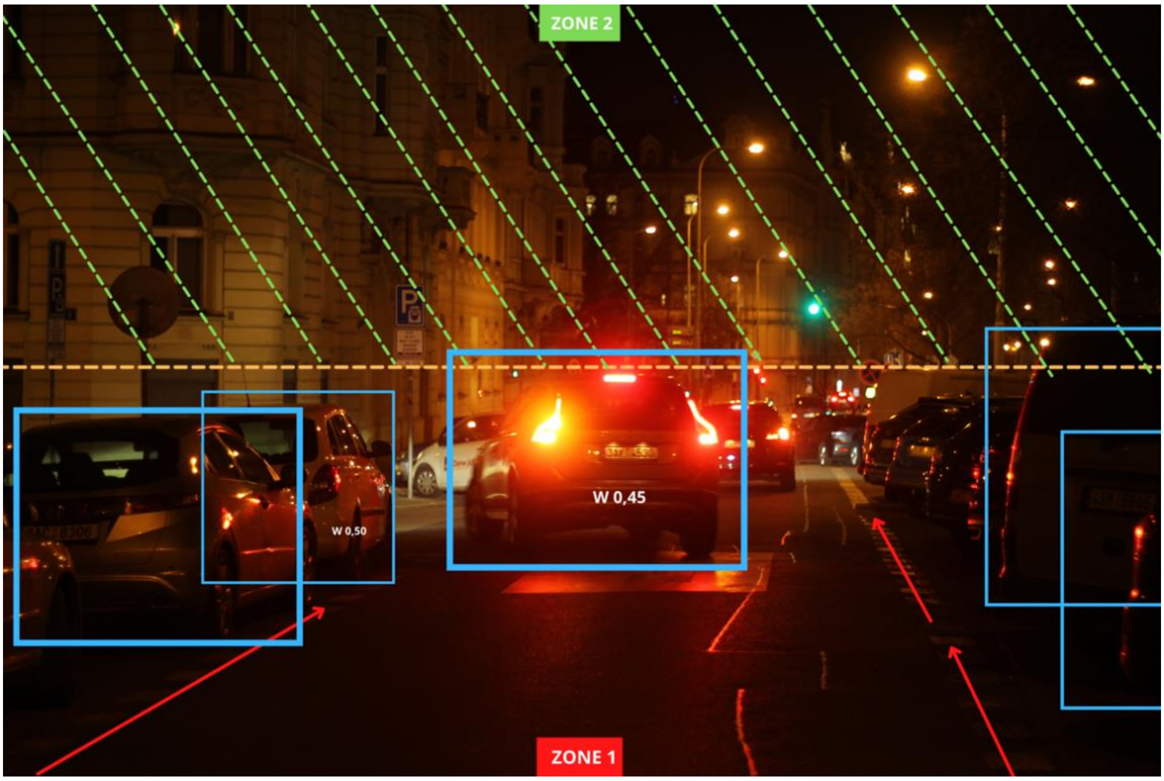


Figure 32. Weight distribution - Nighttime [Location 2]

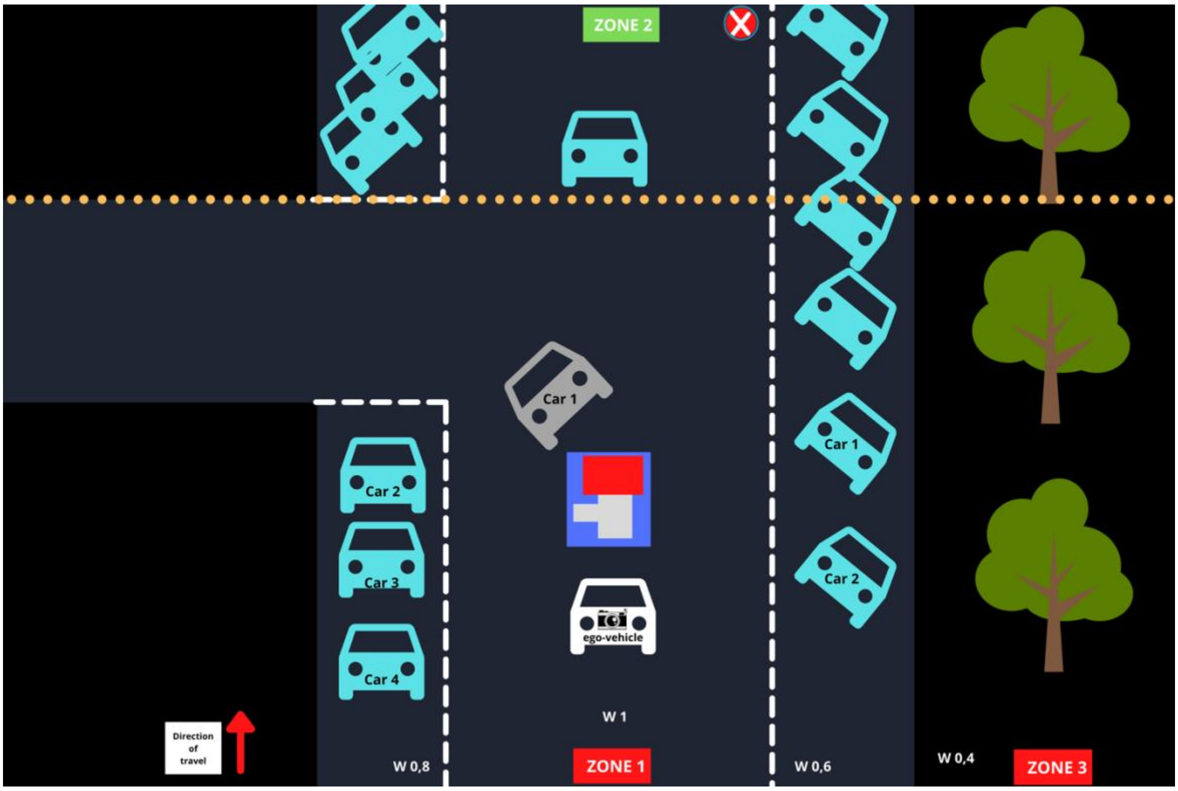


Figure 33. Top View - Nighttime [Location 2]

Since the Zone 2 has a weight of zero, it makes all of the targets in that specific zone equal to zero. The remaining targets are then split into two categories: Zone 1 and Zone 3. Table 14 shows the weight distribution of the targets in location 2 during nighttime.

	Zone 1				Zone 3			
	Car 1	Car 2	Car 3	Car 4	Car 1	Car 2	Tree	
Object Weight	0,9	0,9	0,9	0,9	0,9	0,9	0,8	
Lane Weight	1	0,8	0,8	0,8	0,6	0,6	0,4	
Zone Weight	1	1	1	1	0,1	0,1	0,1	
Direction Weight	0,5	0,7	0,4	0,3	0,4	0,4	0	
Distance Weight	0,9	1	0,4	0,2	0,2	0,1	0,1	
Time Weight	1	1	0,4	0,2	0,2	0,1	0,1	Total Sum
Overall Weight	0,450	0,504	0,0460	0,00864	0,000864	0,000216	0	1,00972

Table 14. Weight Distribution - Nighttime [Location 2]

$$Total\ Weight_{location\ 2-Nighttime} = \sum_{i=1}^{3} Total\ Weight_{Zone\ i}$$

$$Total\ Weight_{Location\ 2-Nighttime} = 1{,}00864 + 0 + 0{,}00108 = 1{,}00972$$

❖ Detected results

The observed findings of location 2 showed major changes in target count in each exposure level during the night. As a result, the analysis was carried out on all of the various exposure levels and sizes that revealed variations. When the overexposed image of 320 x 240 resolution detected more targets, all three sizes of normally exposed pictures detected fewer targets. In comparison, black and white and underexposed 320 x 240 resolution images failed to detect any targets at all.

i) Normally exposed

Starting with normally exposed 1280 x 960 resolution image, Figure 34 shows the detector failed to detect the cars in zone 3 from the scene. Even though these targets have lesser significance compared to other targets in the location, it affects the overall weight distribution. In addition, a new target has been identified in Zone 2, but since the zone has a weight of zero, it has no effect on the total weight. The new weight distribution for this image is calculated in Table 16.

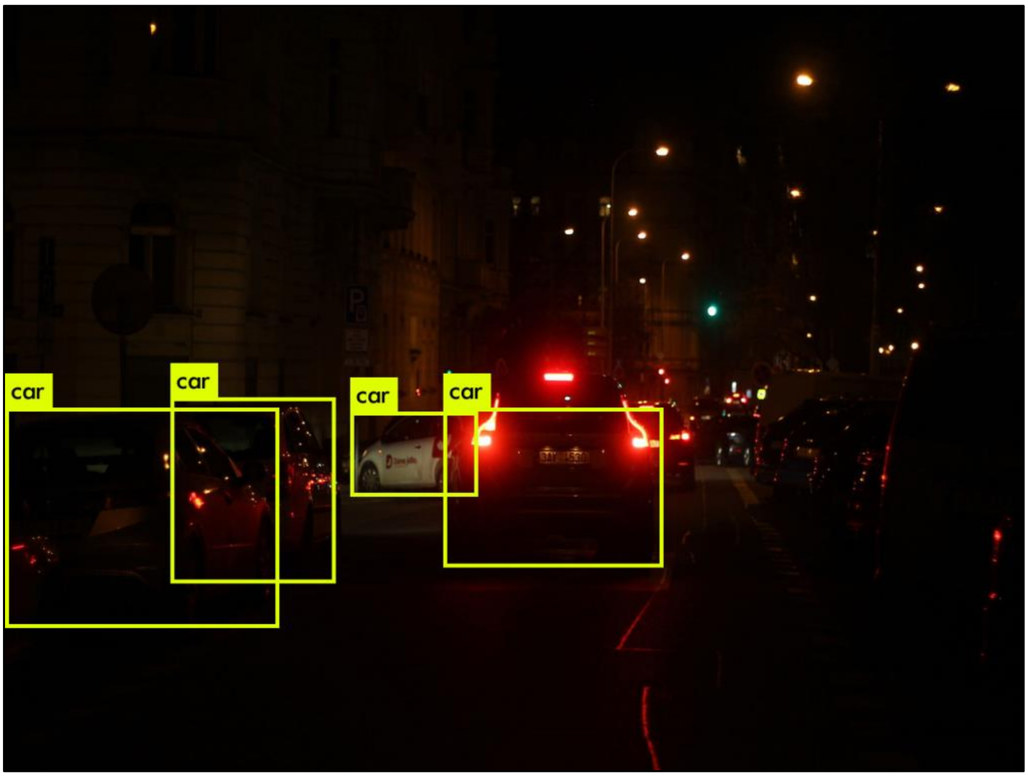


Figure 34. Normally exposed 1280 x 960 - Nighttime [Location 2]

	Zone 1			Zone 2	
	Car 1	Car 2	Car 3	Car 1	
Object Weight	0,9	0,9	0,9	0,9	
Lane Weight	1	0,8	0,8	0,8	
Zone Weight	1	1	1	0	
Direction Weight	0,5	0,7	0,4	0,2	
Distance Weight	0,9	1	0,4	0,1	
Time Weight	1	1	0,4	0,1	Total Sum
Overall Weight	0,450	0,504	0,0460	0	1,00

Table 15. Weight Distribution 1280 x 960 - Nighttime [Location 2]

Zone 2 and 3 has overall weight of zero each, leaving only zone 1. The weight difference is then calculated as:

$$Weight\ Difference_{overall} = Overall\ Weight_{defined} - Overall\ Weight_{detected}$$

$$Weight\ Difference_{overall} = 1{,}00972 - 1{,}00 = 0{,}00972$$

0,962 % decrease ↓

The naturally exposed 1280 x 960 resolution night image showed only a 0,962 percent decrease compared to the defined overall weight distribution.

For the second analysis of normally exposed pictures, both 640 x 480 and 320 x 240 resolution pictures showed similar detection results. Figure 35 shows the 320 x 240 resolution picture in normal light condition.

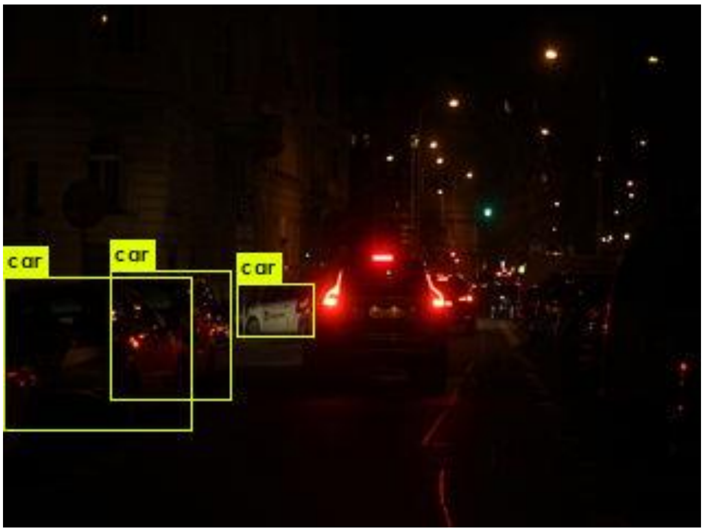


Figure 35. Normally exposed 320 x 240 - Nighttime [Location 2]

In Figure 35 it is seen that the detector failed to detect the Car 1 (driving car) taking the left turn from the scene. This target is critical to ego-vehicle’s path and has relatively higher significance compared to other targets. However, a target in zone 2 was detected similar to the previous analysis, which is again not counted toward the final weight distribution. The final weight distribution of this image is then calculated in Table 16.

	Zone 1		Zone 2	
	Car 2	Car 3	Car 1	
Object Weight	0,9	0,9	0,9	
Lane Weight	0,8	0,8	0,8	
Zone Weight	1	1	0	
Direction Weight	0,7	0,4	0,2	
Distance Weight	1	0,4	0,1	
Time Weight	1	0,4	0,1	Total Sum
Overall Weight	0,504	0,0460	0	0,55

Table 16. Weight Distribution - Normally exposed 320 x 240 – Nighttime [Location 2]

$$Weight\ Difference_{overall} = Overall\ Weight_{defined} - Overall\ Weight_{detected}$$

$$Weight\ Difference_{overall} = 1{,}00972 - 0{,}55 = 0{,}45972$$

45,52 % decrease ↓

There is a 45,52 percent decrease in total weight distribution, demonstrating how important the Car 1 is to the ego-vehicle's course and also showing that the defined weight distribution for this car in Table 14 was close to the ground truth.

ii) Overexposed

This analysis was performed on the overexposed 320 x 240 resolution image, and since there is more light in the image, the detector was able to detect more number of targets than the normally exposed images.

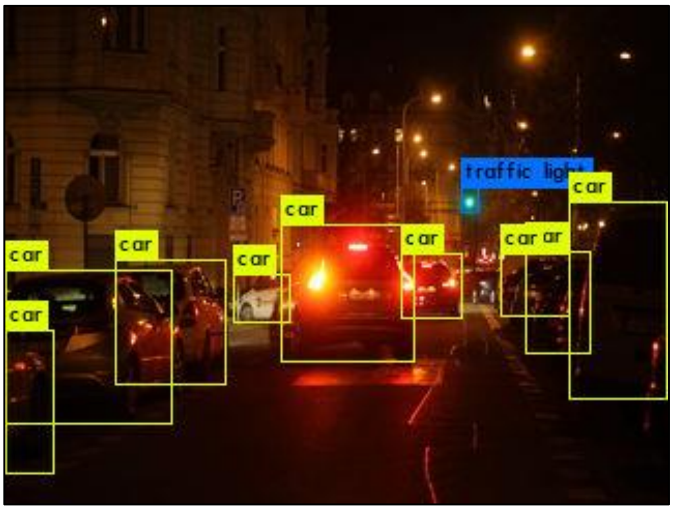


Figure 36. Overexposed 320 x 240 - Nighttime [Location 2]

There are three new targets detected in Zone 2 including a car on the driving lane and two parked on either side. But these targets won’t affect the overall weight distribution since they’re in Zone 2. Thus, the new weight distribution of the overexposed image would be then similar to that in Table 14.

$$Weight\ Difference_{overall} = Overall\ Weight_{defined} - Overall\ Weight_{detected}$$

$$Weight\ Difference_{overall} = 1{,}00972 - 1{,}00972 = 0$$

0 % decrease ↓

This 0% drop, also with new targets detected, demonstrates that the weights are appropriately assigned to the zones and its targets.

iii) Underexposed and Black and White

The 320 x 240 resolution underexposed, and black and white photographs failed to detect any of the targets in both images. A lack of sufficient light and quality may have resulted in the detector failing to pick up any details in these images.

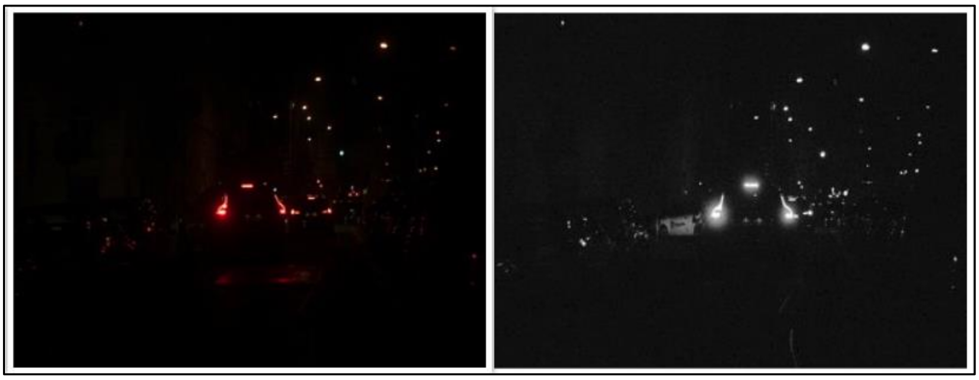


Figure 37. Underexposed and Black & White 320 x 240 - Detected - [Location 2]

If the new weight distribution for these images is then calculated, it would result in a 100 percent drop.

$$Weight\ Difference_{overall} = \ Overall\ Weight_{defined} - \text{O}verall\ Weight_{detected}$$

$$Weight\ Difference_{overall} \ = 1{,}00972 - 0 = \ 1{,}00972$$

100 % decrease ↓

- Detection faults

The detector failed to identify the signs for dead zone or any traffic signs at that location, similar to the detection flaws of location 2 in Section 5.4.1. Detection, in particular, becomes substantially more difficult under low light settings as compared to daytime. To increase the low-light performance of detecting systems used in autonomous vehicles, it is not only necessary to train these datasets, but use enhanced sensors to improve the quality of recognition. Only sensors with superior low-light performance can identify the existence of these targets, further assisting the detector in categorizing them accordingly.

5.5 Summary of Results

This section summarizes the results obtained from the experiments including the locations examined in Appendix A. All of the photographs from these eight sites were reduced from their original size of 5184 x 3456 pixels to three distinct sizes: 1280 x 960, 640 x 480, and 320 x 240 pixels, respectively. The experiment’s second approach was to convert the normally exposed 320 x 240 resolution image to black and white in order to minimize the size even further. These photographs were then run using the algorithm to compare the detection results to their benchmark photographs captured during both day and night.

Table 17 demonstrates the benchmark and detected weight distribution of each location during the day and night. The table also indicates the exposure level and image size at which the detected images varied, as well as the percentage of overall weight drop.

A weight drop of more than 20% is regarded too high since these images failed to identify the most significant targets on the ego-vehicle's route. This allows us to examine the aspects that had the greatest effect on weight drop in all locations in terms of image quality and lighting conditions. Only the identified pictures that differed in detection from the original picture are included in the table.

Location	Time	Weight	Image Exposure	Image Size	Weight ↓ drop %
Location 1	**Day**	**1,0352**	**Benchmark**	**5184 x 3456**	**N/A**
	Night	**0,5320**	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,753	Overexposed	320 x 240	27,62 %
	Night	0,516	Underexposed	320 x 240	3,00 %
	Night	0,450	Black & white	320 x 240	15,41 %
Location 2	**Day**	**0,9551**	**Benchmark**	**5184 x 3456**	**N/A**
	Night	**1,0097**	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,0001	Underexposed	320 x 240	0,01%
	Night	0,0097	Normal	1280 x 960	0,962%
	Night	0,4597	Normal	320 x 240	45,52%
	Night	0	Underexposed	320 x 240	100 %
	Night	0	Black & white	320 x 240	100 %
Location 3	**Day**	**0,5051**	**Benchmark**	**5184 x 3456**	**N/A**
	Night	**0,4906**	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,0010	Black & white	320 x 240	0,197%
	Night	0,4846	Underexposed	320 x 240	1,22%
	Night	0,4846	Black & white	320 x 240	1,22%
Location 4	**Day**	**0,8756**	**Benchmark**	**5184 x 3456**	**N/A**
	Night	**0,6570**	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,8756	-	-	0%
	Night	0,6570	-	-	0%
Location 5	**Day**	**0,6680**	**Benchmark**	**5184 x 3456**	**N/A**
	Night	**0,2394**	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,3870	Overexposed	320 x 240	42,06%
	Night	0,0864	Black & white	320 x 240	63,90%
Location 6	**Day**	0,2632	**Benchmark**	**5184 x 3456**	**N/A**
	Night	0,371	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,2160	Underexposed	320 x 240	17,93%
	Night	0,0360	Black & white	320 x 240	90,29%

Location 7	**Day**	0,7090	**Benchmark**	**5184 x 3456**	**N/A**
	Night	0,0648	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,4930	Overexposed	320 x 240	30,46%
	Night	0,0648	-	-	0%
Location 8	**Day**	**0,723**	**Benchmark**	**5184 x 3456**	**N/A**
	Night	**0,273**	**Benchmark**	**5184 x 3456**	**N/A**
	Day	0,273	Black & white	320 x 240	62,24%
	Night	0	Underexposed	320 x 240	100%

Table 17. Final Results

Table 17 shows that all of the images with any change in overall weight have a resolution of 320 x 240 pixels, with the exception of Location 2, where a normally exposed photograph at night with a size of 1280 x 960 pixels showed a weight drop of less than 1%. When closely examined, the majority of the photos that showed a substantial weight shift in detection were at night and in extreme lighting conditions, either overexposed or underexposed. Aside from locations 3, 4, and 7, black and white nighttime photographs at all sites revealed a minimum weight reduction of 15%. In the preliminary design of the experiment, the targets in each location were given different weights by category, with an aim to find the minimum conditions in which the algorithm could find the worthy of recognition objects. The numbers in Table 17 reveal that the weights assigned to the targets at these locations for the experiment were appropriate.

The experimental findings conclude that an image with a resolution of at least 320 x 240 pixels is required for the detector to effectively recognize all of the relevant targets in the ego-vehicle's course. Furthermore, based on the results, it is reasonable to conclude that, if the detection model in any autonomous system is designed to process black and white pictures to reduce the data load, the photos must be a minimum of 640 x 480 pixels in order for the detector to perform efficiently.

6. Conclusion

The aim of this work was to identify the minimum criteria of the picture quality required for object detection while maintaining relevant object detection level. The paper begins with investigating the concepts of existing sensor technologies for object identification, with a particular emphasis on the visual sensor camera. Each sensor was considerably researched, as were their object recognition methods, types, attenuation, and limitations, while the camera received a fairly in-depth study. An appropriate experimental method was chosen based on the environmental factors at work in the artificial intelligence industry. The experiment was performed using a deep learning based algorithm, on photographs taken from eight different locations around Prague city both during the day and night.

Two different approaches of data processing were employed in which the photographs were downsized into three different resolutions, with the smallest size colorized to black and white, before running the algorithm for detection. The results of this detection were then compared to the weights of targets in each location, assigned based on its worth of recognition to the course of the ego-vehicle. When the experimental recording was evaluated in this manner, both the lighting and minimum image quality requirements needed for the algorithm to perform effectively was found. It is observed that the detector is especially sensitive under severe lighting settings such as overexposure and underexposure. Furthermore, the detector performs poorly in detecting objects in lower resolution pictures of 320 x 240 pixels, resulting in considerable variations in detection. At night, the underexposed images also demonstrated poor detecting ability.

Based on the experimental results, the minimum image quality required for successful recognition is 320 x 240 pixels for a color image and 640 x 480 pixels for a black and white image. It can also be stated that the detector requires ambient illumination to work optimally. However, in a real-world setting, this is not the case. In real world, lighting may range from extremely bright to completely dark. Different weather conditions like fog, snow, rain etc. also have an impact on the lighting conditions outside. These considerations are taken into account and possible enhancements can be addressed further as a topic for future study.

7. Future Improvements

There are various ways the experiment can be improved to test the algorithm more thoroughly. The experiment in this thesis was only carried out in varied lighting and at different times of the day. However, performing these experiments in various weather situations, such as rain, fog, and snow, can provide more precise findings on how the algorithm functions. This will provide a better knowledge of how to determine the detector's limit. In addition, the experiment in this research was limited to static images. Testing it on motion pictures can provide an additional layer of analysis. The experiment was also limited to a single field of view (FOV); however, adding several FOVs to each site can further enhance the findings. Furthermore, the photographs could be captured from a car using a mounted camera for better results as it can better simulate a real ADAS system. These are some of the few aspects that can be addressed in the future to minimize the number of untested states and thus, improving the results in terms of ultimately determining the detector's limit and evaluating whether the experimental algorithm is still valid.

8. References

[1] Skolnik, M. I. (2020, October 29). Radar. Retrieved November 04, 2020, from https://www.britannica.com/technology/radar

[2] Rothman, P. (2018, October 16). Radar vs. LiDAR: A Technology Toe-to-Toe. Retrieved November 6, 2020.

[3] Sharma, Bhupendra (2020, October 16). What is LiDAR technology and how does it work? Retrieved November 05, 2020, from https://www.geospatialworld.net/blogs/what-is-lidar-technology-and-how-does-it-work/

[4] Thompson, J. (2019, October 31). LiDAR Explained: How Does It Work? Retrieved November 06, 2020, from https://levelfivesupplies.com/lidar-how-does-it-work/

[5] Hecht, J. (2019, September 25). Automotive Lidar: Safety questions raised about 1550 nm lidar. Retrieved November 06, 2020, from https://www.laserfocusworld.com/blogs/article/14040682/safety-questions-raised-about-1550-nm-lidar

[6] Goodnight, E. (2016, September 28). How Photography Works: Cameras, Lenses, and More Explained. Retrieved November 06, 2020, from https://www.howtogeek.com/63409/htg-explains-cameras-lenses-and-how-photography-works/

[7] Khader, M., & Cherian, S. (2020, May). An Introduction to Automotive LIDAR. Retrieved November 06, 2020, from https://www.ti.com/lit/wp/slyy150a/slyy150a.pdf?ts=1604652158830&ref_url=https%253A%252F%252Fwww.google.com%252F

[8] Science and Technology, M. O. (Ed.). (2016). Introduction to radar systems. Retrieved May 03, 2021, from http://mmust.elimu.net/BSC(ELEC_COMM)/Year_4/ECE%20451%20L_Radar_Eng_and_Facsimile/Introduction_to_Radar/Introduction_to_Radar.htm

[9] Perlstein, D. (2019, July 16). Description and clarification of sensors and various iot sensor types. Retrieved May 04, 2021, from https://axonize.com/blog/iot-automation/iot-sensors-bundles-platforms-oh-my-2/

[10] Whittaker, B. (2018, September 08). The Road to Autonomous Cars: LiDAR Sensors Drives Growth for LeddarTech. Retrieved November 07, 2020, from https://dronebelow.com/2018/05/24/the-road-to-autonomous-cars-lidar-sensors-drives-growth-for-leddartech/

[11] Curry, G. (2011). Radar Essentials: A Concise Handbook for Radar Design and Performance Analysis. Retrieved November 21, 2020, from https://m.eet.com/media/1121840/912radar_essentials_pt1.pdf

[12] ITU. (2015). Systems characteristics and compatibility of automotive radars operating in the frequency band 77.5 78 GHz for sharing studies. Retrieved November 21, 2020, from https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2322-2014-PDF-E.pdf

[13] Shwarz, R. (2018). Enabling Autonomous Driving. Retrieved November 21, 2020, from https://cdn.rohde-

schwarz.com/pws/dl_downloads/dl_common_library/dl_news_from_rs/219/NEWS_219__16_QAR__EN.pdf

[14] Majumdar, S., & Rajani, B. (1999). Numerical computation of Turbulent flow around Radome Structures. Retrieved November 21, 2020, from https://www.sciencedirect.com/science/article/pii/B9780080433288500291

[15] Cooney, L., & Cooney, M. (2011, June 20). Car radar system rules. Retrieved November 21, 2020, from https://www.networkworld.com/article/2229552/toyota-wants-us-to-relax-car-radar-system-rules.html

[16] Mokrane Hadj-Bachir, Philippe de Souza. LIDAR sensor simulation in adverse weather condition for driving assistance development. 2019. hal-01998668

[17] Sacyr. (2021, February 26). Lidar, the new laser eye for mobile phones. Retrieved May 08, 2021, from https://www.sacyr.com/en/-/lidar-el-nuevo-ojo-laser-de-los-telefonos-moviles

[18] McManamon, P. F. (2019). LiDAR Technologies and Systems. Retrieved November 22, 2020, from https://spie.org/samples/PM300.pdf

[19] Google MAPS

[20] Kretschmar, M. (2009). Please Enable Cookies. Retrieved November 22, 2020, from https://www.machinedesign.com/mechanical-motion-systems/article/21829396/sensor-noise-limits-resolution-when-monitoring-motion

[21] Stec, B., & Susek, W. (2018, May 06). Theory and Measurement of Signal-to-Noise Ratio in Continuous-Wave Noise Radar. Retrieved December 04, 2020, from https://www.mdpi.com/1424-8220/18/5/1445/pdf

[22] Vyacheslav Tuzlukov (2010), *Signal Processing Noise*, Electrical Engineering and Applied Signal Processing Series, CRC Press. 688 pages.

[23] Tatum Follow, M. (2017*). Signs, Signals and Markings*. SlideShare. https://www.slideshare.net/MaryTatum2/signs-signals-and-markings

[24] Enhancement of lidar backscatters signal-to-noise ratio using empirical mode decomposition method, Optics Communications, Volume 267, Issue 1, 2006, https://doi.org/10.1016/j.optcom.2006.05.069.

[25] Huang Hao, Tian Lan, Yingchao Zhang, Guoqiang Ni, "The analysis of signalto-noise ratio of airborne LIDAR system under state of motion," Proc. SPIE 7843, High-Power Lasers and Applications V, 78431E (16 November 2010)

[26] Mullan, D. (2019, September). Calculating the Signal to Noise Ratio of a Camera. Retrieved December 05, 2020, from https://andor.oxinst.com/learning/view/article/ccd-signal-to-noise-ratio

[27] Gillmor, C. (2018, May 09). How does a digital camera sensor work? Retrieved December 17, 2020, from https://medium.com/tech-update/how-does-a-digital-camera-sensor-work-1342974250fd

[28] Lucid Vision Lab. (2020). Understanding The Digital Image Sensor. Retrieved December 17, 2020, from https://thinklucid.com/tech-briefs/understanding-digital-image-sensors/

[29] Augustyn, A. (2020, February 28). Diffraction. Retrieved December 17, 2020, from https://www.britannica.com/science/diffraction

[30] Fellers, T. J., & Davidson, M. W. (n.d.). Concepts in Digital Imaging Technology: CCD Noise Sources and Signal-to-Noise Ratio. Retrieved December 18, 2020, from https://hamamatsu.magnet.fsu.edu/articles/ccdsnr.html

[31] K.K, H. (n.d.). What is photon shot noise? Retrieved December 18, 2020, from https://camera.hamamatsu.com/jp/en/technical_guides/photon_shot_noise/index.html

[32] R. Paschotta, article on 'radiometry' in the *Encyclopedia of Laser Physics and Technology*, 1. edition October 2008, Wiley-VCH, ISBN 978-3-527-40828-3

[33] Fiete, R.D.. (2010). Modeling the imaging chain of digital cameras. 10.1117/3.868276.

[34] Weik M.H. (2000) Snell's law. In: Computer Science and Communications Dictionary. Springer, Boston, MA. https://doi.org/10.1007/1-4020-0613-6_17633

[35] Farrell, Joyce & Catrysse, Peter & Wandell, Brian. (2012). Digital camera simulation. Applied optics. 51. A80-90. 10.1364/AO.51.000A80.

[36] R. C. Gonzales and R. E. Woods, Digital Image Processing, 3rd ed., Prentice-Hall, Upper Saddle River, New Jersey (2008).

[37] T. Lule, S. Benthien, H. Keller, F. Mutze, P. Rieve, K. Seibel, M. Sommer, and M. Bohm, “Sensitivity of CMOS Based Imagers and Scaling Perspectives,” IEEE Transactions of Electronic Devices, Vol. 47, No. 11, November 2000, pp. 2110–2122.

[38] R. Ramanath, W.-E. Snyder, Y. Yoo, and M.-S. Drew, “Color Image Processing Pipeline,” IEEE Signal Processing Magazine, Vol. 22, No. 1, January 2005, pp. 34–43.

[39] Nakamura, Junichi (2005). Image Sensors and Signal Processing for Digital Still Cameras. CRC Press. ISBN 978-0-8493-3545-7

[40] Lukac R., Plataniotis K.N. (2006) Digital Camera Image Processing. In: Furht B. (eds) Encyclopedia of Multimedia. Springer, Boston, MA. https://doi.org/10.1007/0-387-30038-4_56

[41] G. Wyszecki and W. S. Stiles, “Color Science, Concepts and Methods, Quantitative Data and Formulas,” John Wiley, N.Y., 2nd Edition, 1982

[42] G. C. Holst and T. S. Lomheim, CMOS/CCD Sensors and Camera Systems, JCD Publishing, Winter Park, Florida and SPIE Press, Bellingham, Washington (2007).

[43] C.Y. Fang, C.S. Fuh, P.S. Yen, S. Cherng, S.W. Chen, An automatic road sign recognition system based on a computational model of human recognition processing, Computer Vision and Image Understanding, Volume 96, Issue 2, 2004, Pages 237-268, http://www.sciencedirect.com/science/article/pii/S1077314204000761

[44] G. Sharma, H.J. Trussell Digital color imaging IEEE Trans. Image Process., 6 (1997), pp. 901-932

[45] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system", *IEEE International Conference on Intelligent Robots and Systems* 2004, pp. 70-75, Oct. 2004.

[46] Y. Gu, T. Yendo, M. Panahpour Tehrani, T. Fujii and M. Tanimoto, "Traffic sign detection in dual-focal active camera system," 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, 2011, pp. 1054-1059, doi: 10.1109/IVS.2011.5940513.

[47] U. L. Jau, C. S. Teh and G. W. Ng, "A comparison of RGB and HSI color segmentation in real — time video images: A preliminary study on road sign detection*", International Symposium on Information Technology 2008*, pp. 1-6, Aug. 2008.

[48] Y. Gu, T. Yendo, M. Panahpour Tehrani, T. Fujii and M. Tanimoto, "Traffic sign detection based on shape and color", *Workshop on Picture Coding and Image Processing 2010, pp. 111-112, Dec. 2010.*

[49] Toth, Štefan. (2012). Difficulties of Traffic Sign Recognition.

[50] Wali, Safat & Abdullah, Majid & Hannan, M. A. & Hussain, Aini & Samad, Salina & Ker, Pin Jern & Mansor, Muhamad. (2019). Vision-Based Traffic Sign Detection and Recognition Systems: Current Trends and Challenges. Sensors. 19. 2093. 10.3390/s19092093.

[51] Li, Jia and Z. Wang. “Real-Time Traffic Sign Recognition Based on Efficient CNNs in the Wild.” *IEEE Transactions on Intelligent Transportation Systems* 20 (2019): 975-984.

[52] Y. SOLDI Et Al. , "CNN based traffic sign recognition for mini autonomous vehicles," *39th International Conference Information Systems Architecture and Technology, ISAT 2018* , vol.853, NYSA, Poland, pp.85-94, 2018

[53] Brownlee, J. (2019, October 07). How to perform object detection With Yolov3 in Keras. Retrieved April 12, 2021, from https://machinelearningmastery.com/how-to-perform-object-detection-with-yolov3-in-keras/

[54] Li, E. (2020, March 17). Dive really deep into Yolo v3: A beginner's guide. Retrieved April 12, 2021, from https://towardsdatascience.com/dive-really-deep-into-yolo-v3-a-beginners-guide-9e3d2666280e

[55] Adaloglou, N. (2020, March 23). Intuitive explanation of SKIP connections in deep learning. Retrieved April 13, 2021, from https://theaisummer.com/skip-connections/

[56] Redmon, J., &amp; Farhadi, A. (2018, April 08). YOLOv3: An Incremental Improvement. Retrieved April 13, 2021, from https://arxiv.org/pdf/1804.02767.pdf

[57] Sahoo, S. (2018, November 29). Residual blocks - building blocks of resnet. Retrieved April 13, 2021, from https://towardsdatascience.com/residual-blocks-building-blocks-of-resnet-fd90ca15d6ec

[58] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp.770–778.

[59] Ju, & Yun, Luo & Wang, & Hui, & Chang,. (2019). The Application of Improved YOLO V3 in Multi-Scale Target Detection. Applied Sciences. 9. 3775. 10.3390/app9183775.

[60] Seif, G. (2021, February 14). Visualising filters and feature maps for deep learning. Retrieved April 13, 2021, from https://towardsdatascience.com/visualising-filters-and-feature-maps-for-deep-learning-d814e13bd671

[61] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690.

[62] Garbade, D. (2018, September 12). Understanding k-means clustering in machine learning. Retrieved April 13, 2021, from https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

[63] K. (Ed.). (2019, August 27). Logistic regression for machine learning and classification. Retrieved April 13, 2021, from https://kambria.io/blog/logistic-regression-for-machine-learning/

[64] Strubell, Emma & Ganesh, Ananya & Mccallum, Andrew. (2019). Energy and Policy Considerations for Deep Learning in NLP. 3645-3650. 10.18653/v1/P19-1355.

[65] Hao, K. (2020, December 07). Training a single AI model can emit as much carbon as five cars in their lifetimes. Retrieved April 26, 2021, from https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

[66] Toews, R. (2020, July 21). Deep learning's carbon emissions problem. Retrieved April 26, 2021, from https://www.forbes.com/sites/robtoews/2020/06/17/deep-learnings-climate-change-problem/

[67] Brownlee, J. (2019, August 07). What is natural language processing? Retrieved April 26, 2021, from https://machinelearningmastery.com/natural-language-processing/

[68] University of Reading. (2008, July 23). Understanding digital images. Retrieved April 26, 2021, from https://www.reading.ac.uk/web/files/using-images/Understanding1.pdf

[69] Tabora, V. (2019, September 25). The relationship of image resolution to image size. Retrieved April 26, 2021, from https://medium.com/hd-pro/the-relationship-of-image-resolution-to-image-size-1f6a28ea30bb

[70] Sawalich, W. (2018, August 24). Understanding digital image resolution. Retrieved April 26, 2021, from https://www.dpmag.com/how-to/tip-of-the-week/understanding-digital-image-resolution/

[71] Wheeler, M. (2019, August 19). Understanding digital images. Retrieved April 26, 2021, from https://www.psm-marketing.com/understanding-digital-images/

[72] Redmon, J. (2018). YOLO: Real-Time Object Detection. Retrieved April 14, 2020, from https://pjreddie.com/darknet/yolo/

[73] Yohanandan, S. (2020, June 09). Map (mean average precision) might confuse you! Retrieved May 02, 2021, from https://towardsdatascience.com/map-mean-average-precision-might-confuse-you-5956f1bfa9e2

[74] Dempster, P. (2014, July 14). CCD and CMOS Sensors. Tech Briefs. Retrieved June 06 from https://www.techbriefs.com/component/content/article/tb/supplements/ptb/-features/articles/20063

9. Appendices

List of Tables

List of Figures

Appendix A: Remaining Locations

Similar analysis from Section 5.4 was carried out on the remaining locations. Both their daytime and nighttime photos, along with their weight distribution and top views, are included here. The calculations of these locations are included in the excel file named Appendix_Calculations.

I. Daytime

i) Location 4

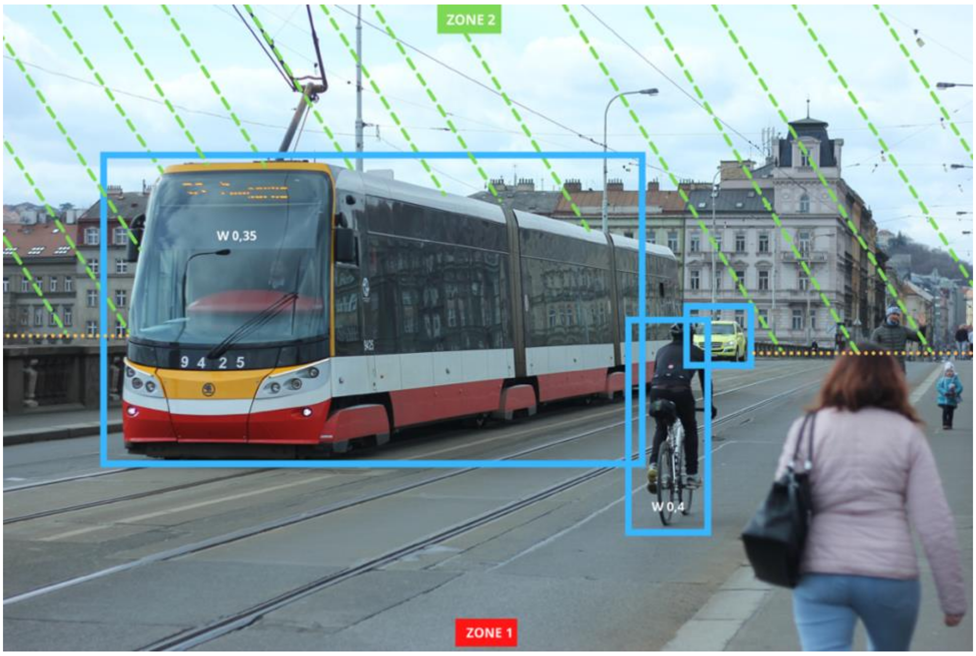


Figure I: Palackeho bridge [Location 4]

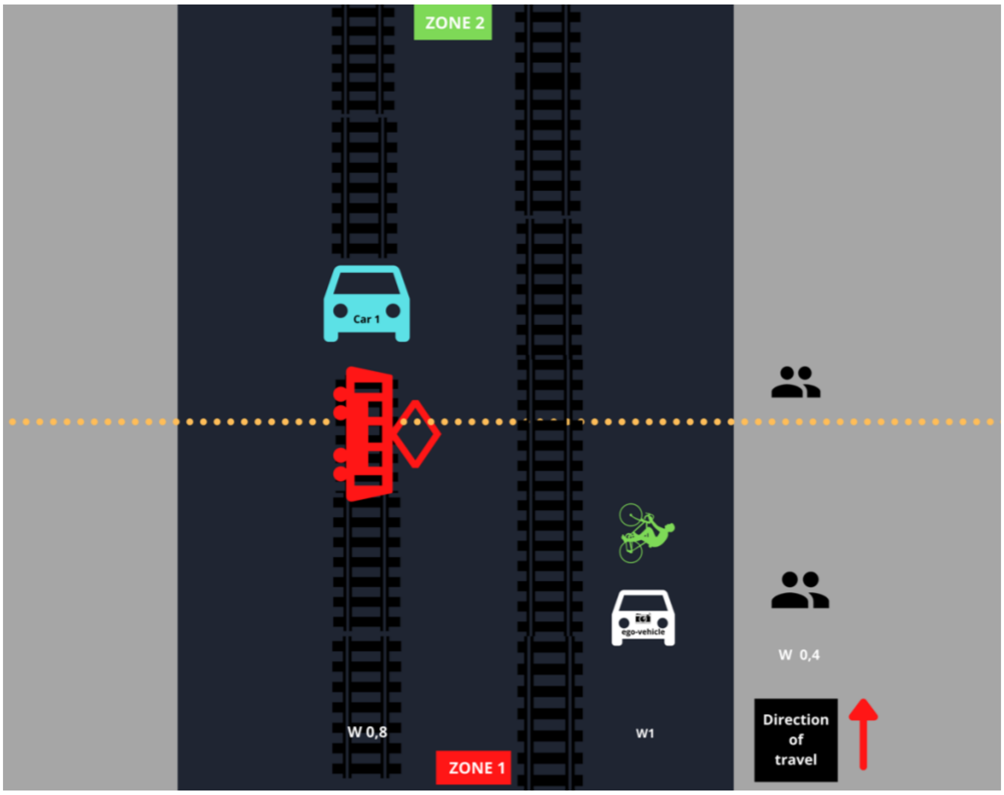


Figure II: Top view [Location 4]

The location in Figure I is on the Palckeho bridge is a two-way road with both trams and cars sharing the same lanes. Since the bridge is narrow and relatively short, the location has been split into two different zones. We picture ourselves as the ego-vehicle behind the cyclist.

ii) Location 5

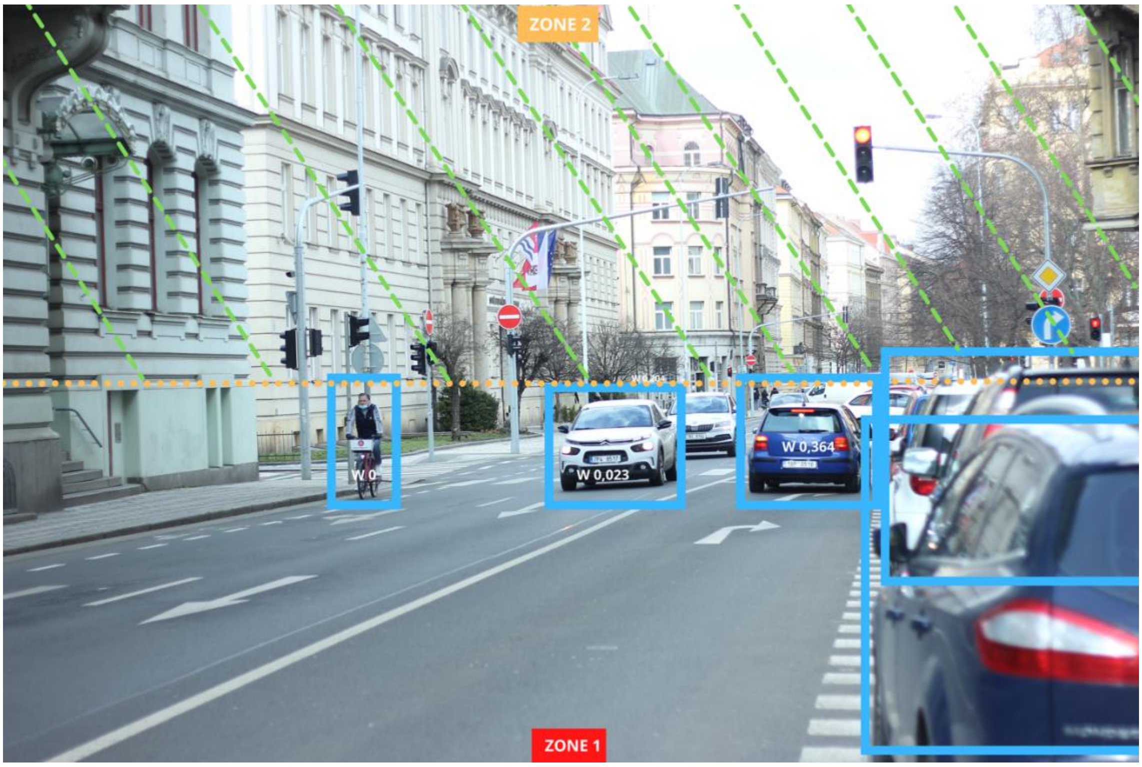


Figure III: Zborovska road [Location 5]

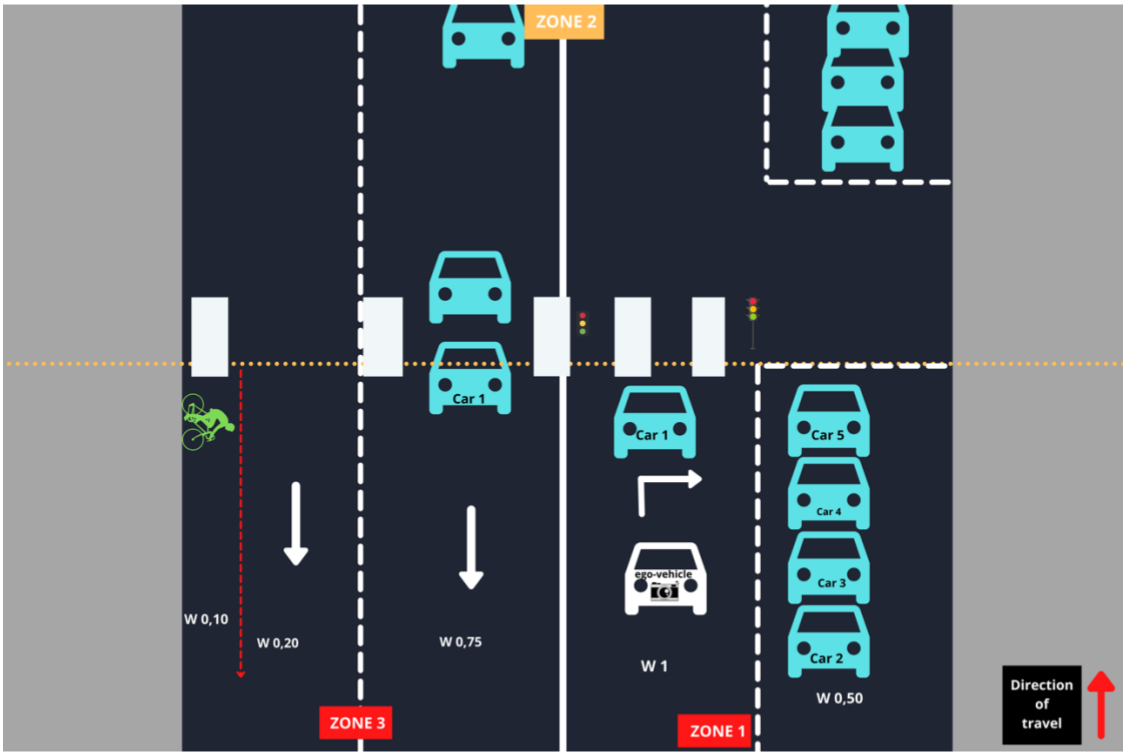


Figure IV: Top view [Location 5]

Figure III depicts a three-way road with two lanes in the opposite direction of the ego-vehicle's path and one lane in the same direction. The location was divided into three zones, with Zone 2 bearing a weight of zero. Here, the ego-vehicle is behind the dark blue car waiting for the traffic signal.

iii) Location 6

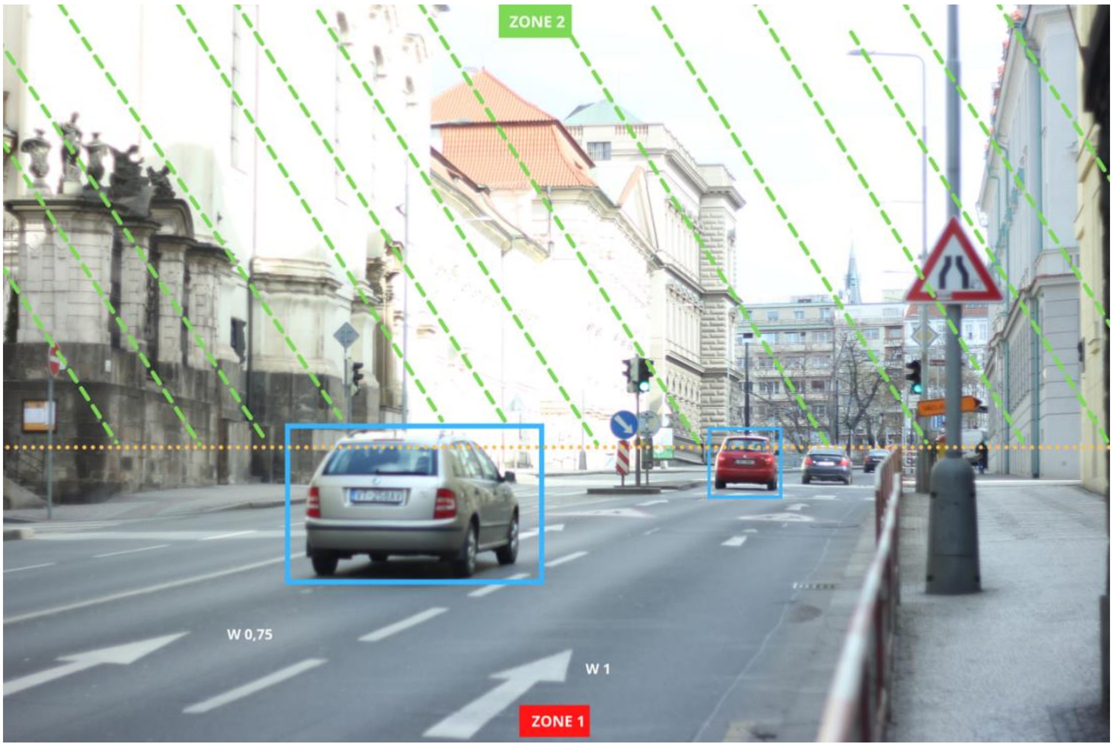


Figure V: Resslova road [Location 6]

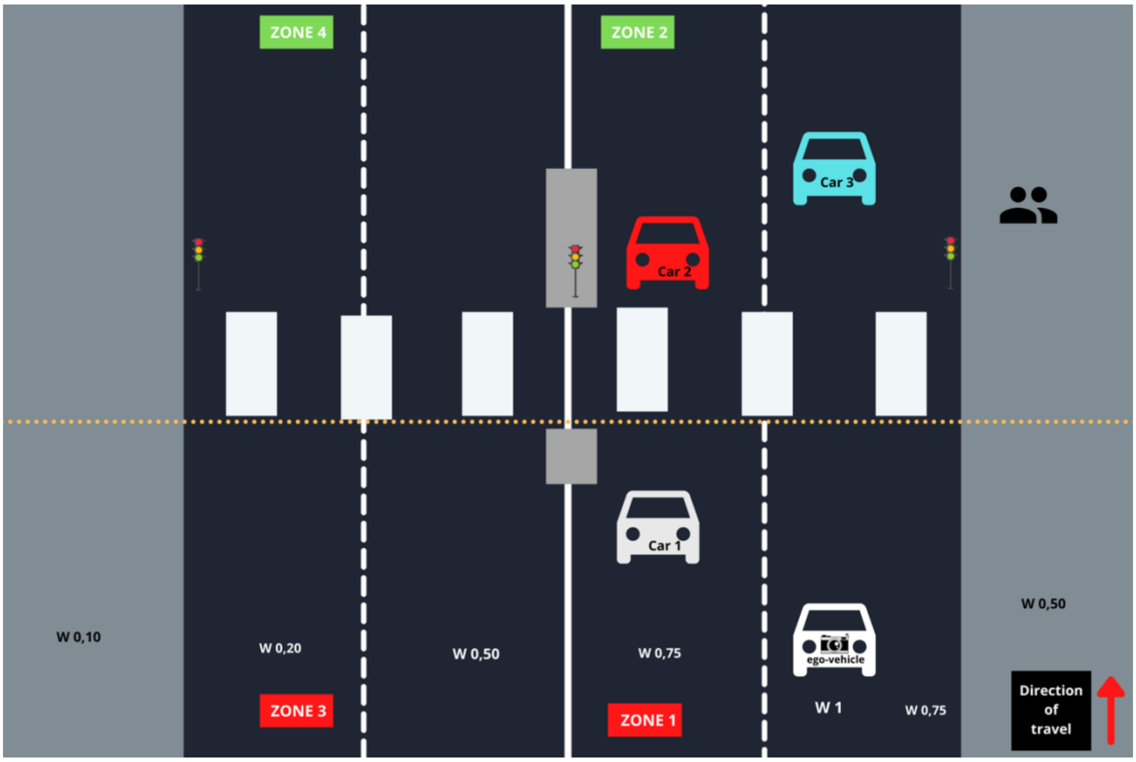


Figure VI: Top view [Location 6]

iv) Location 7

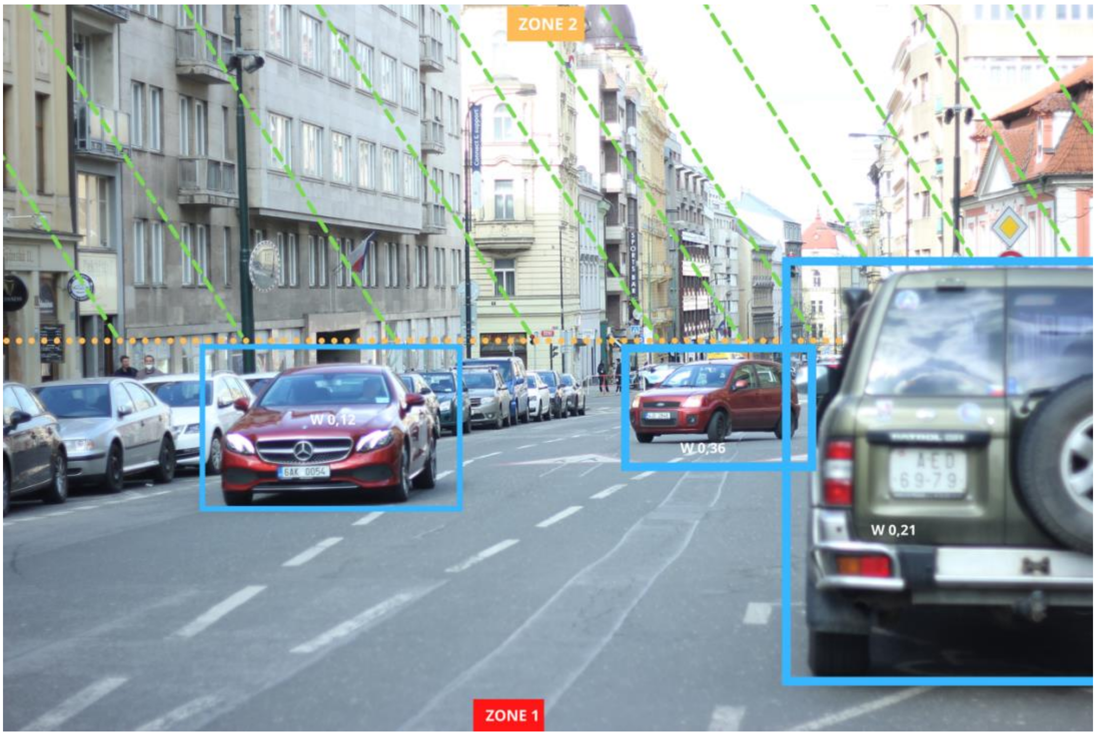


Figure VII: Štěpánská road [Location 7]

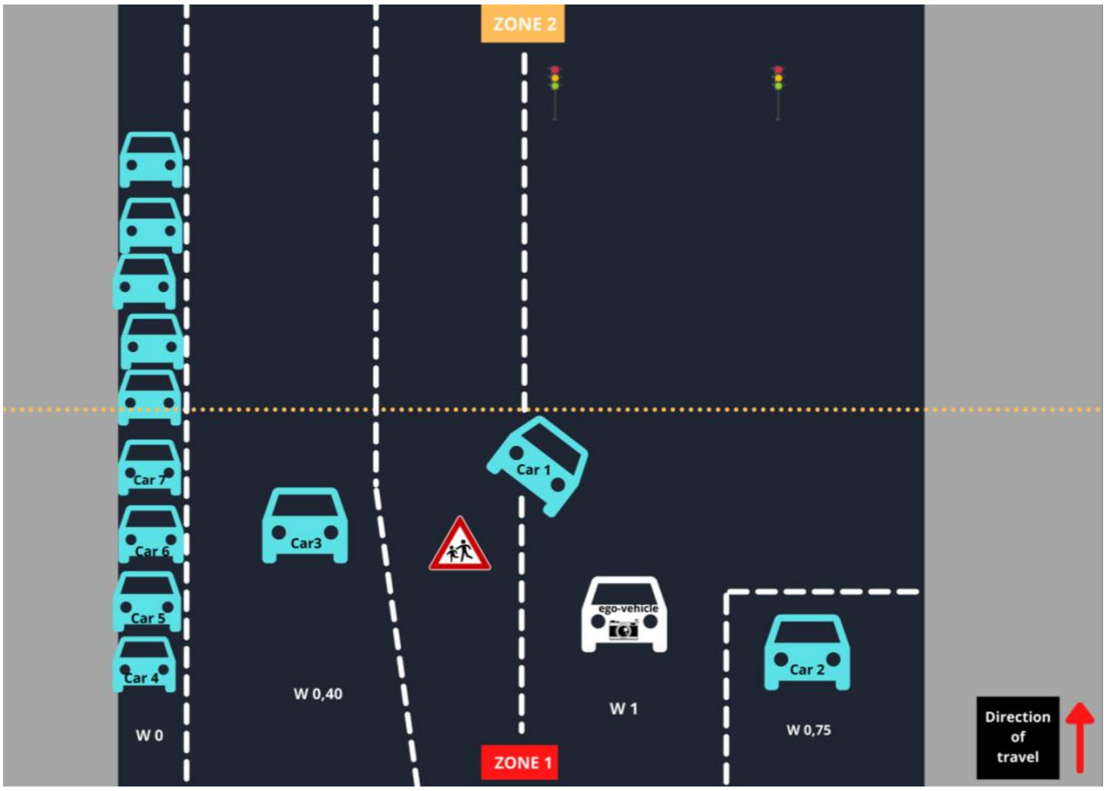


Figure VIII: Top view [Location 7]

Location 7 is on the Štěpánská road which is a two-way street one lane in each direction. As seen in Figure VII, the ego-vehicle is coming from the direction next to the parked green car.

v) Location 8

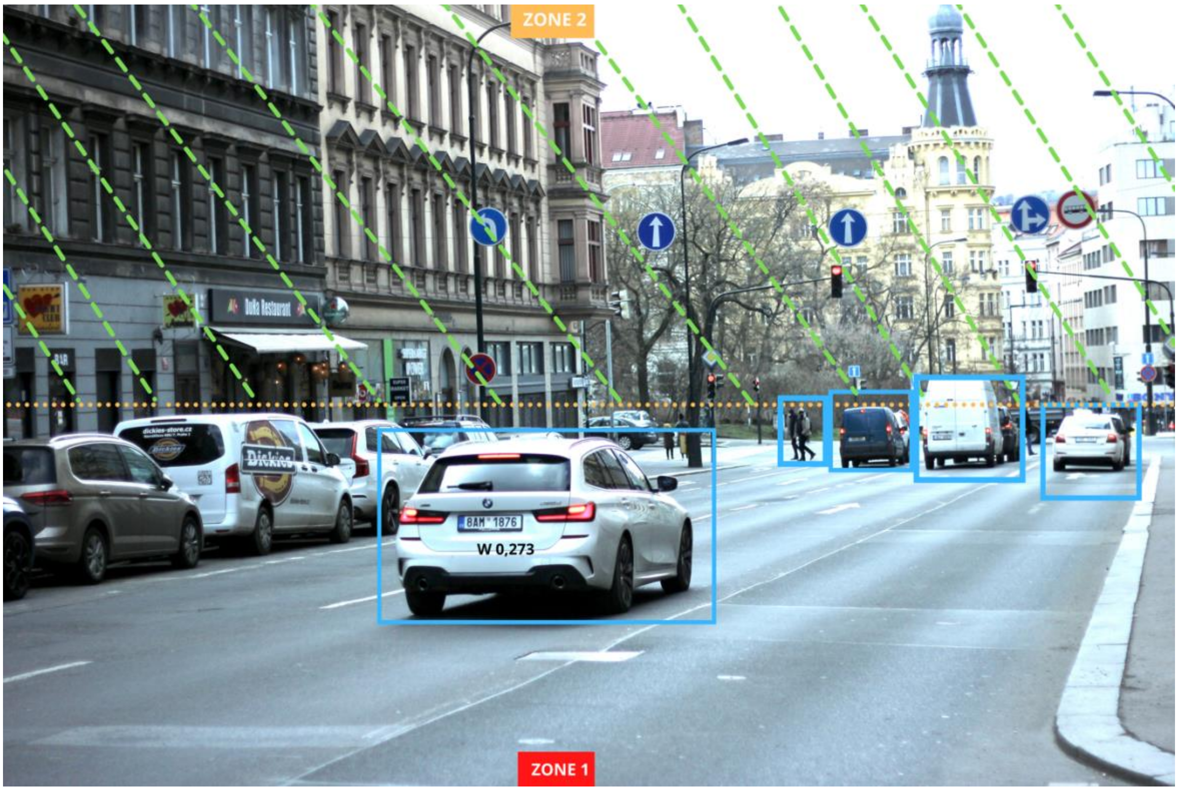


Figure IX: Žitná road [Location 8]

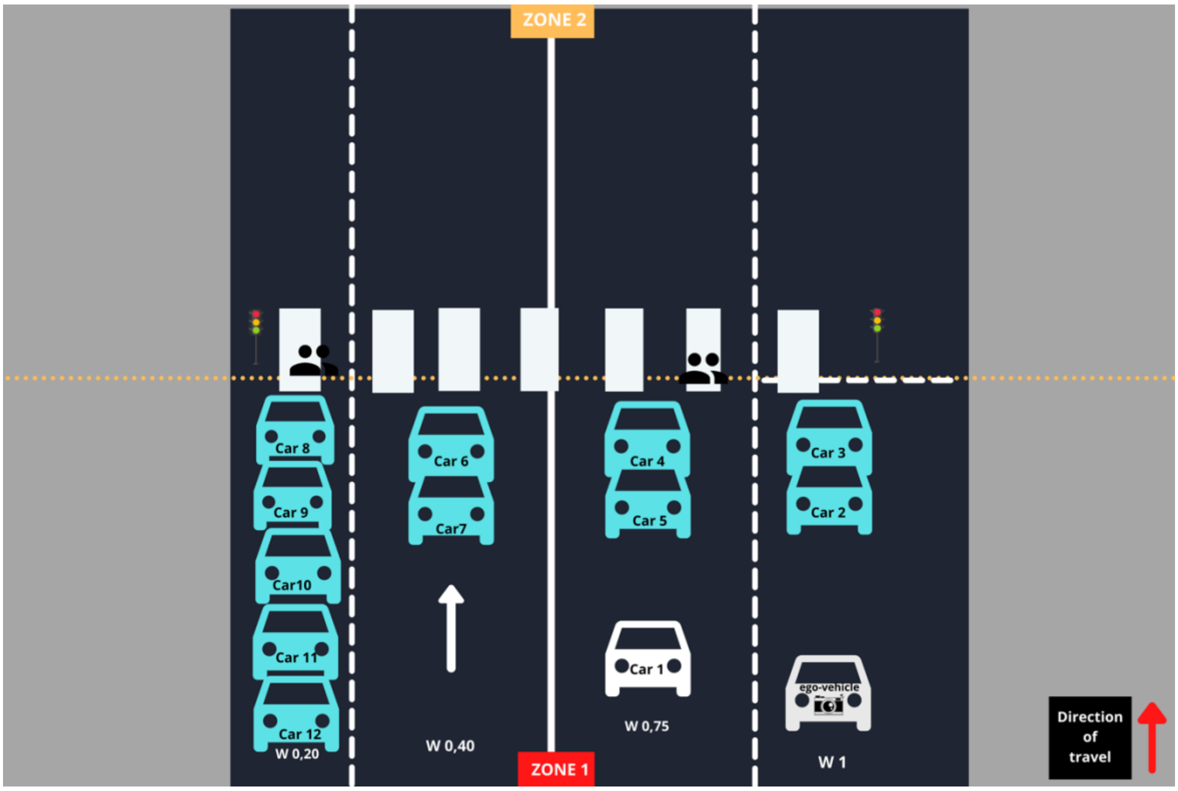


Figure X: Top view [Location 8]

II. Nighttime

i) Location 1

Since location 2 has already been covered in Section 5.4.2 for the nighttime analysis, the remaining nighttime photographs of other sites, as well as their top views, are investigated here.

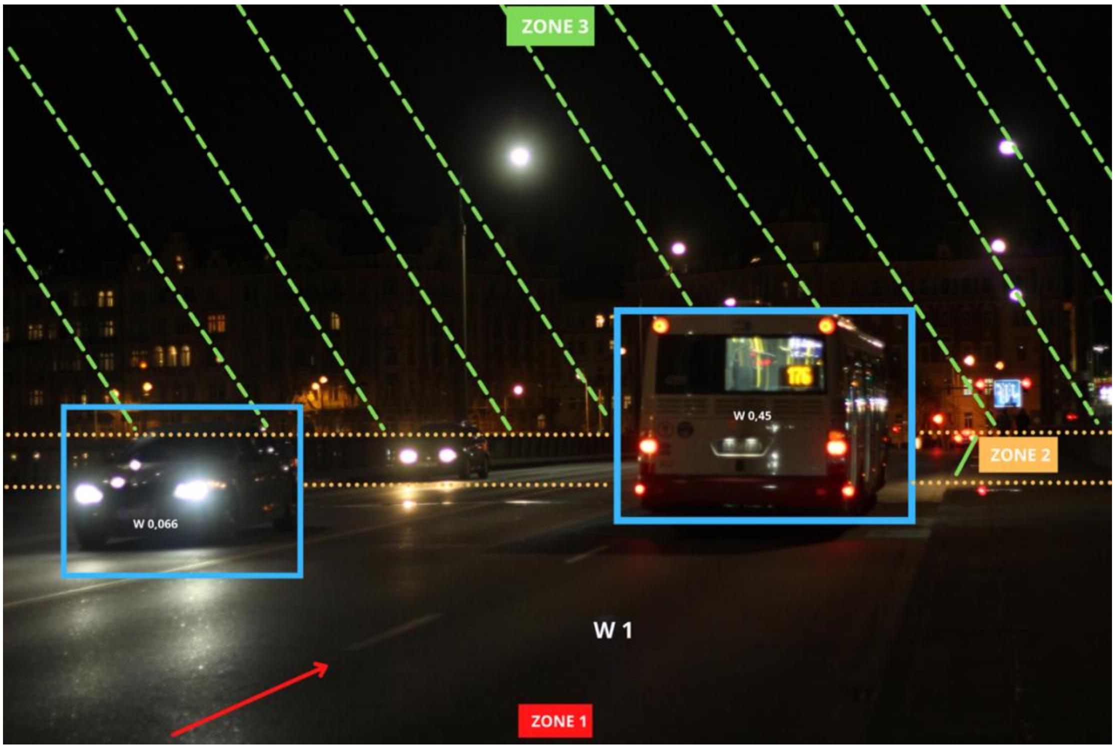


Figure XI: Jirasek Bridge - Nighttime [Location 1]

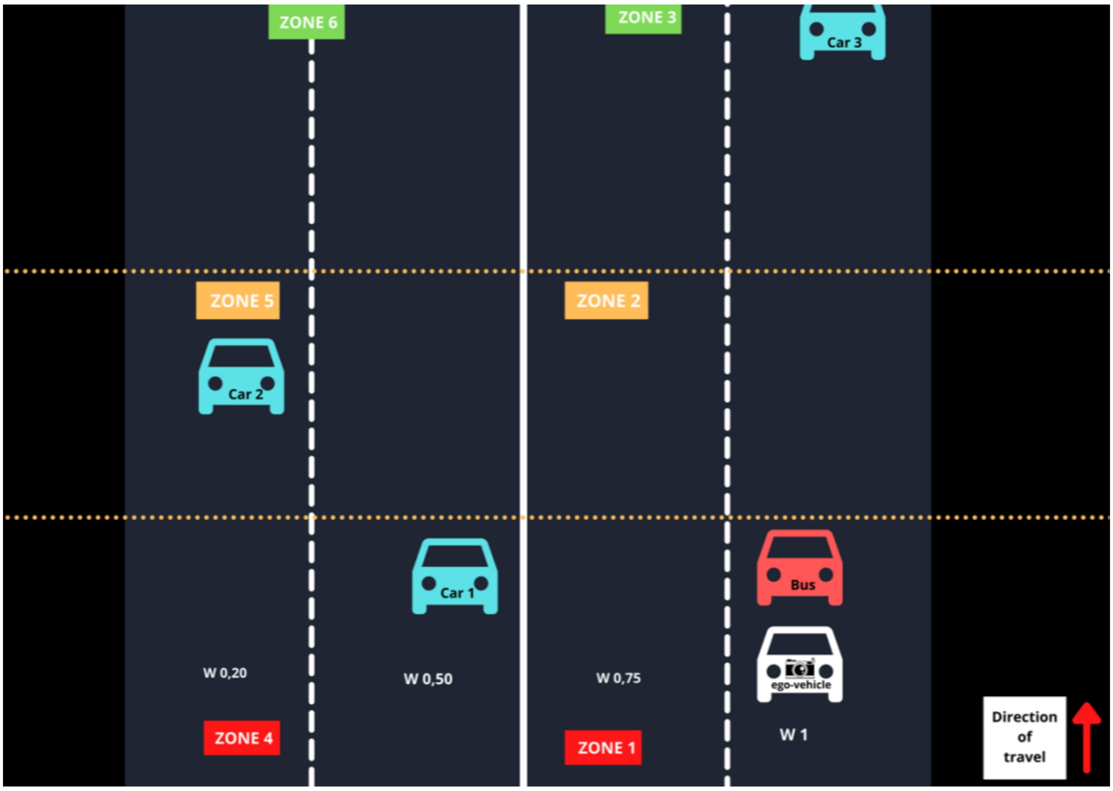


Figure XII: Top view - Nighttime [Location 1]

ii) Location 3

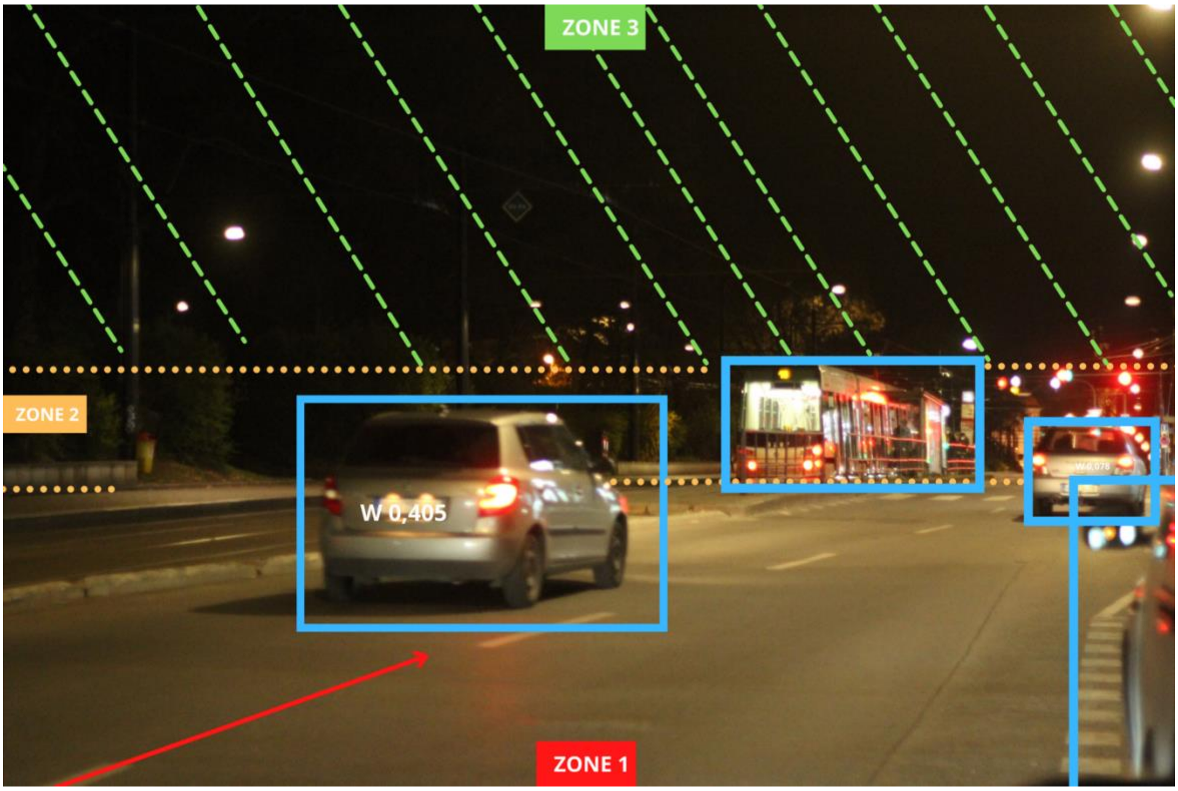


Figure XIII: Karlovo námesti - Nighttime [Location 3]

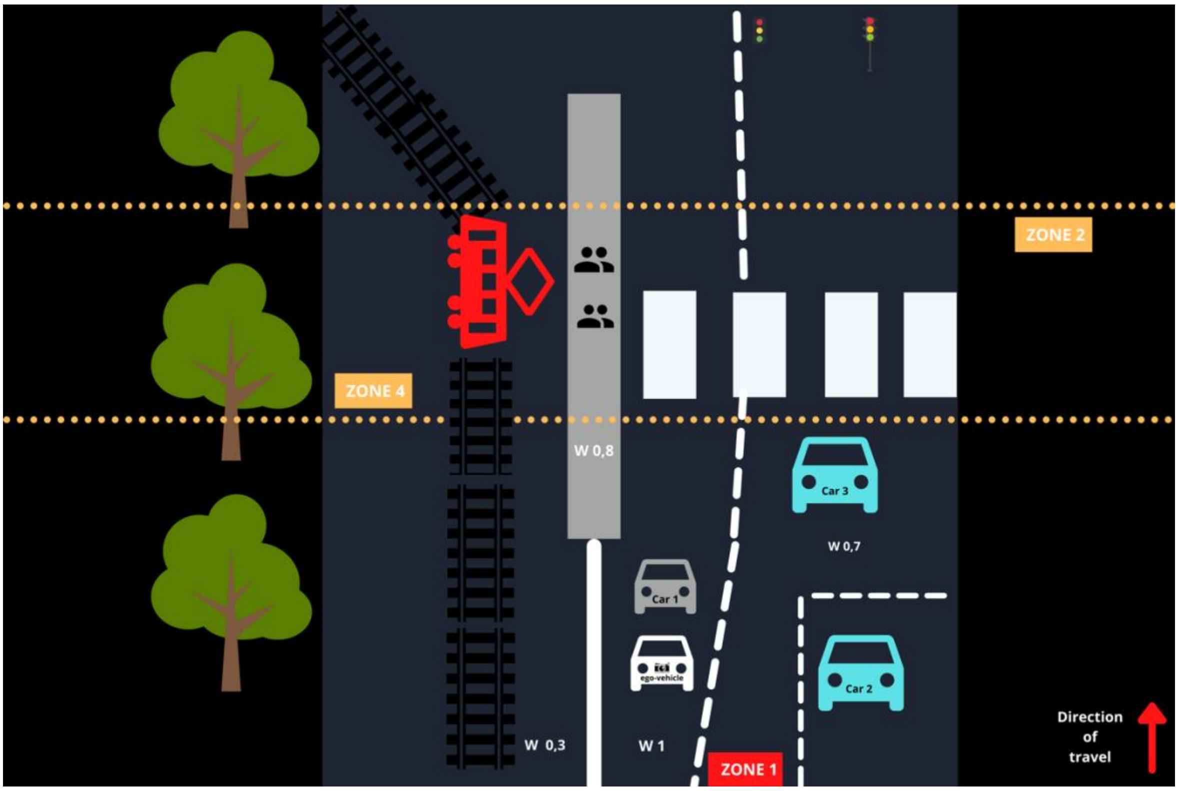


Figure XIV: Top view - Nighttime [Location 3]

iii) Location 4

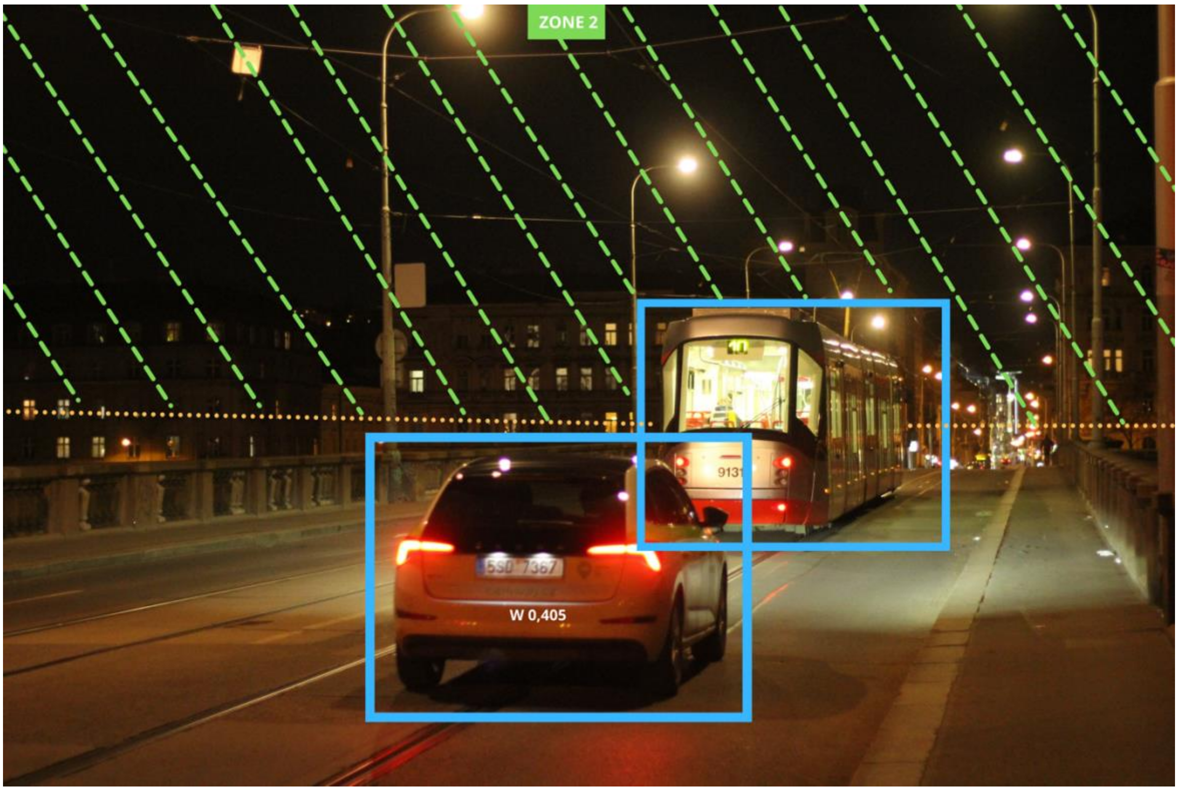


Figure XV: Palackeho bridge - Nighttime [Location 4]

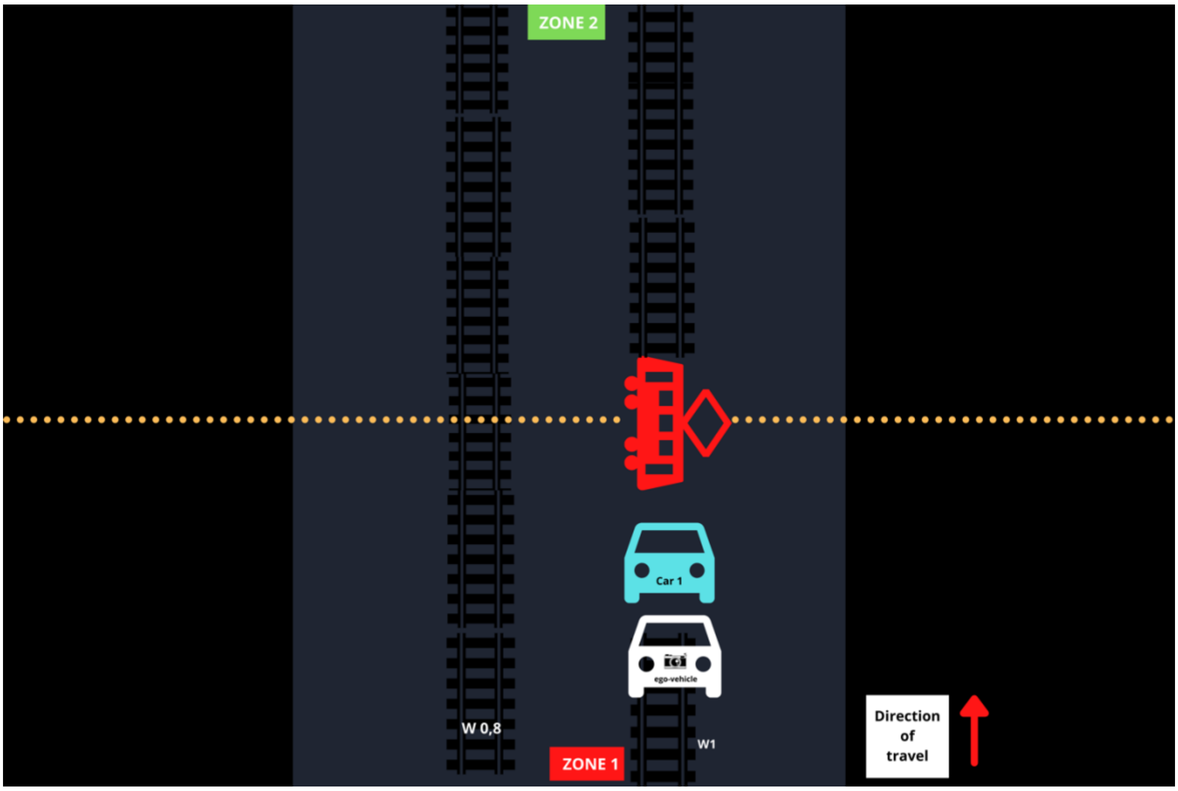


Figure XVI: Top view - Nighttime [Location 4]

iv) Location 5

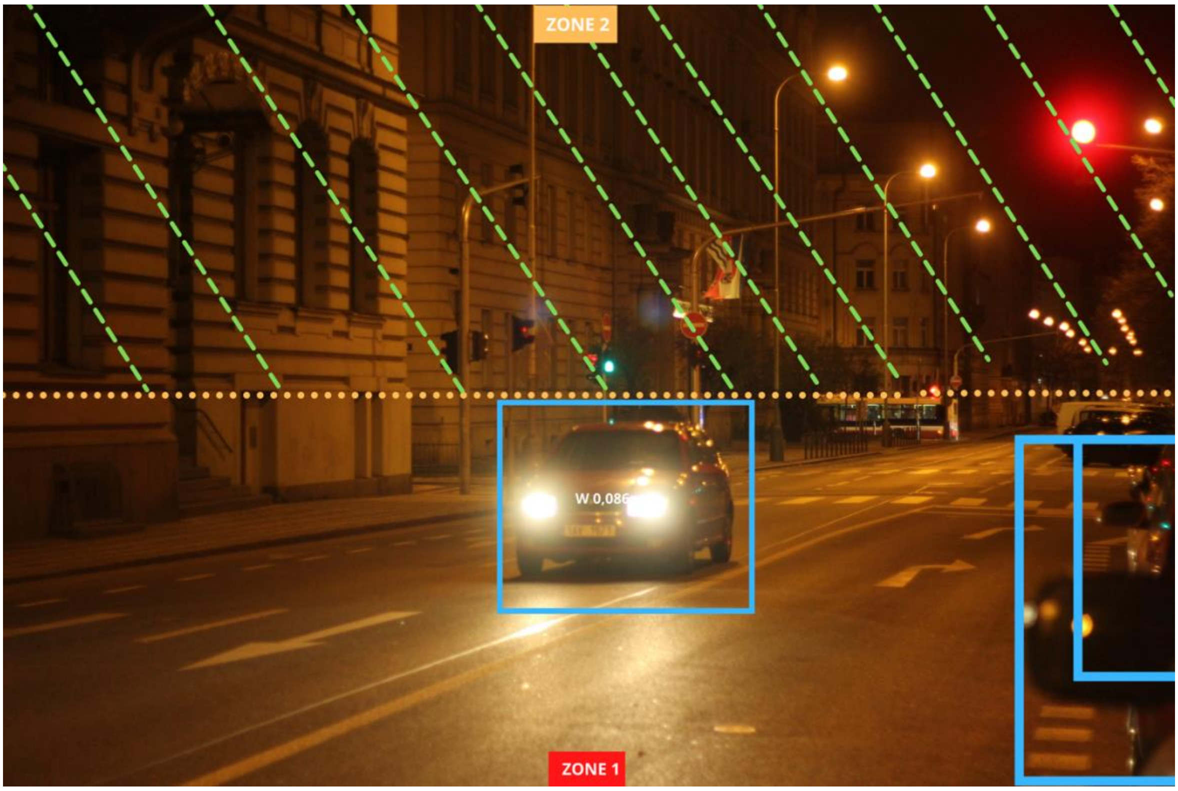


Figure XVII: Zborovská road - Nighttime [Location 5]

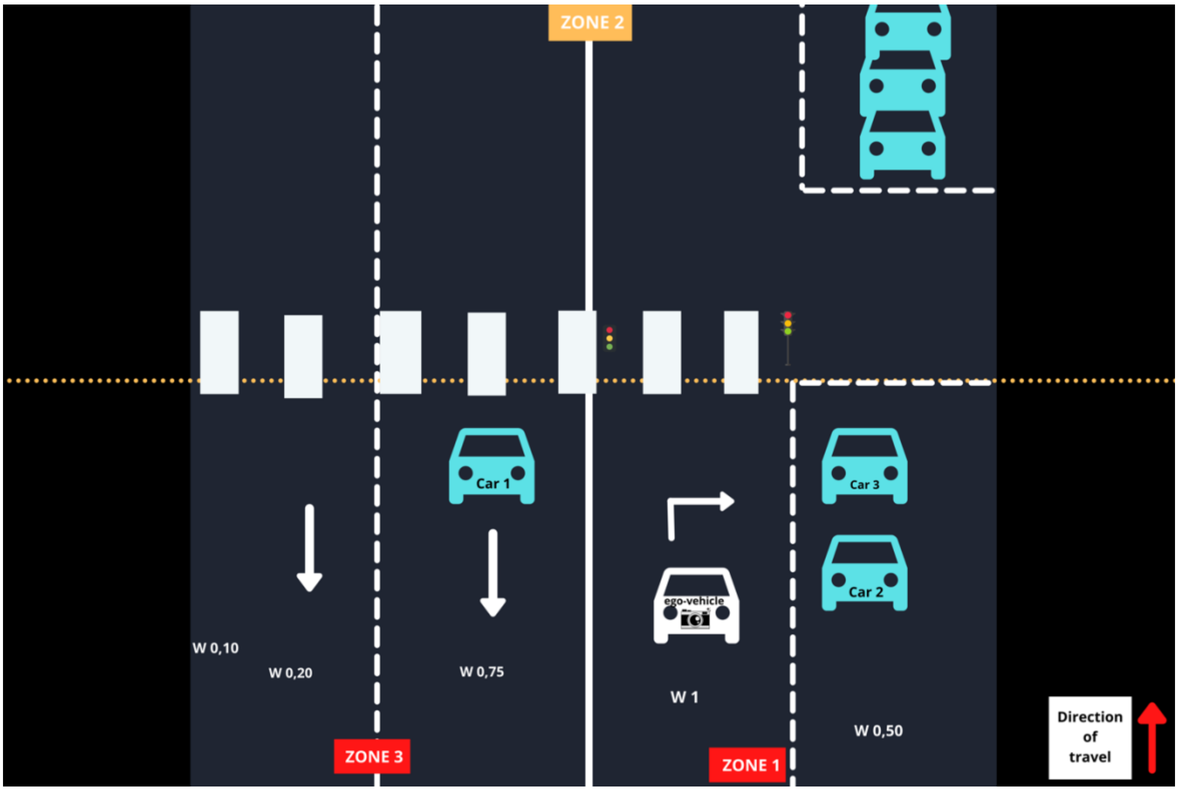


Figure XVIII: Top view - Nighttime [Location 5]

v) Location 6

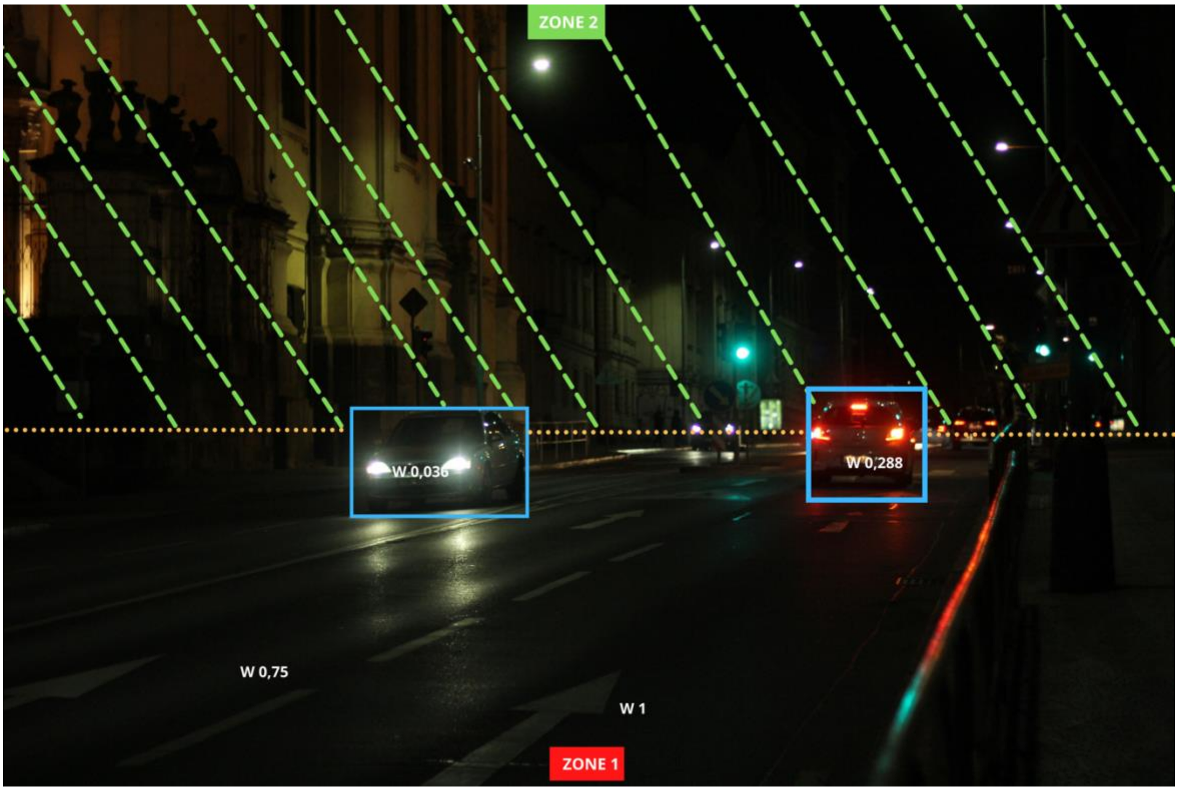


Figure XIX: Resslova road - Nighttime [Location 6]

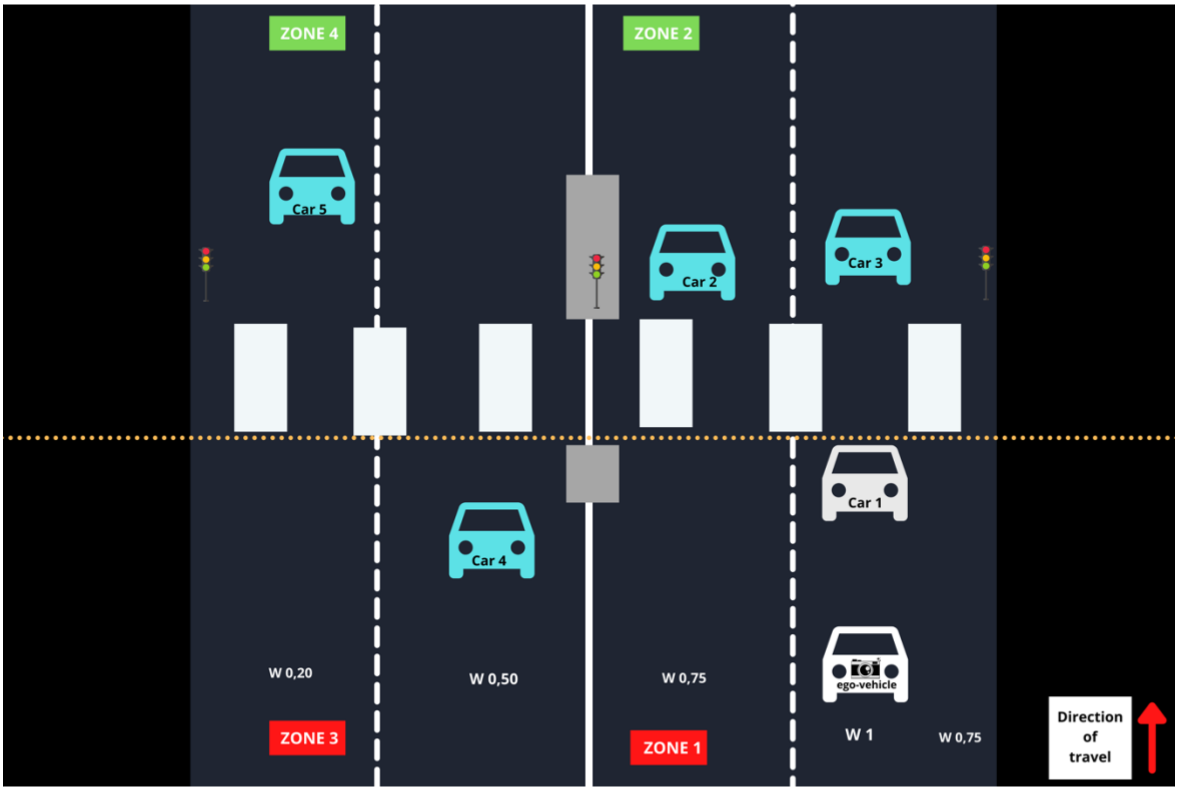


Figure XX: Top view - Nighttime [Location 6]

At night the location 6 had more targets on the road compared to the daytime which resulted in a higher overall weight.

vi) Location 7

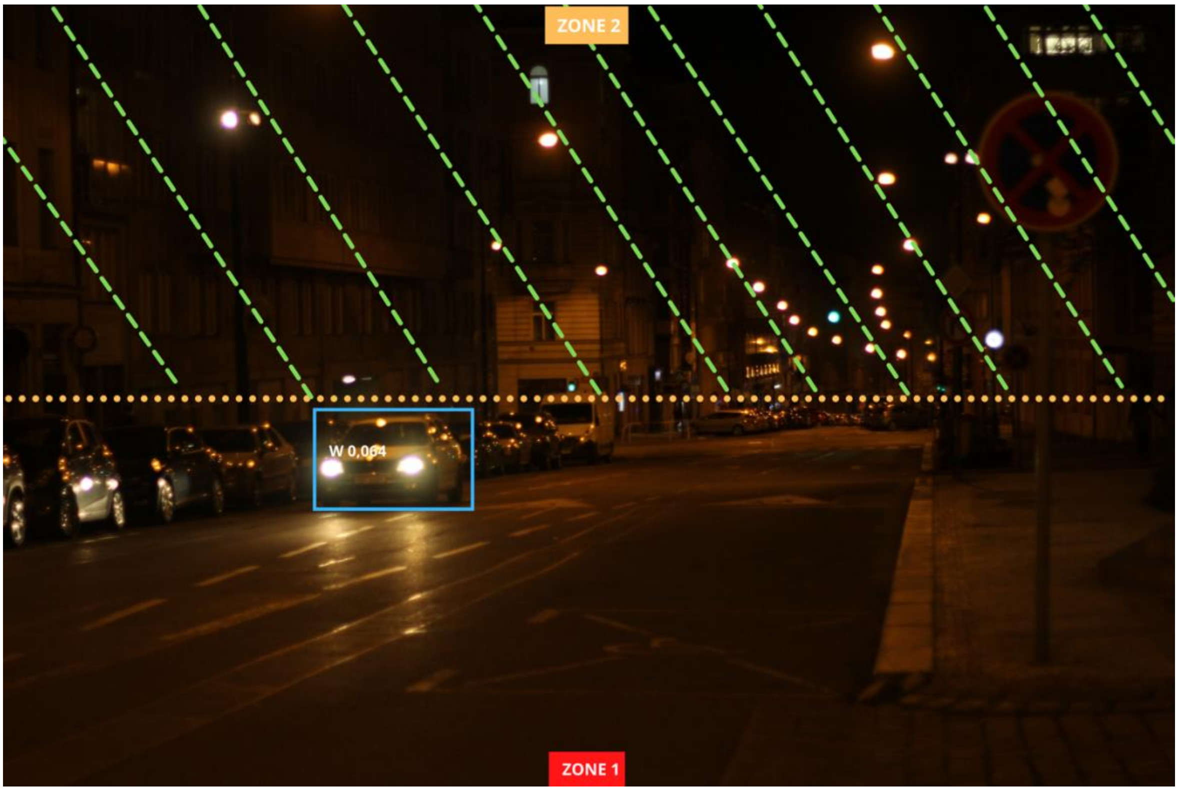


Figure XXI: Štěpánská road - Nighttime [Location 7]

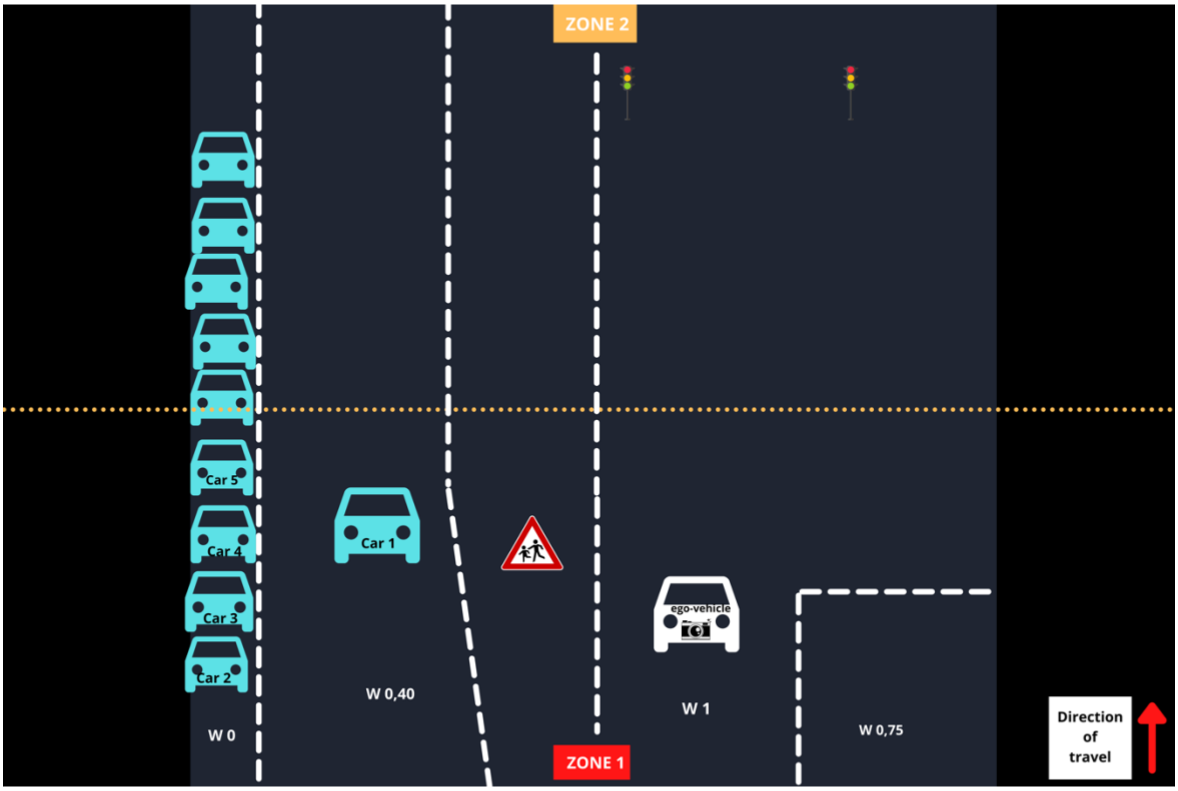


Figure XXII: Top view - Nighttime [Location 7]

vii) Location 8

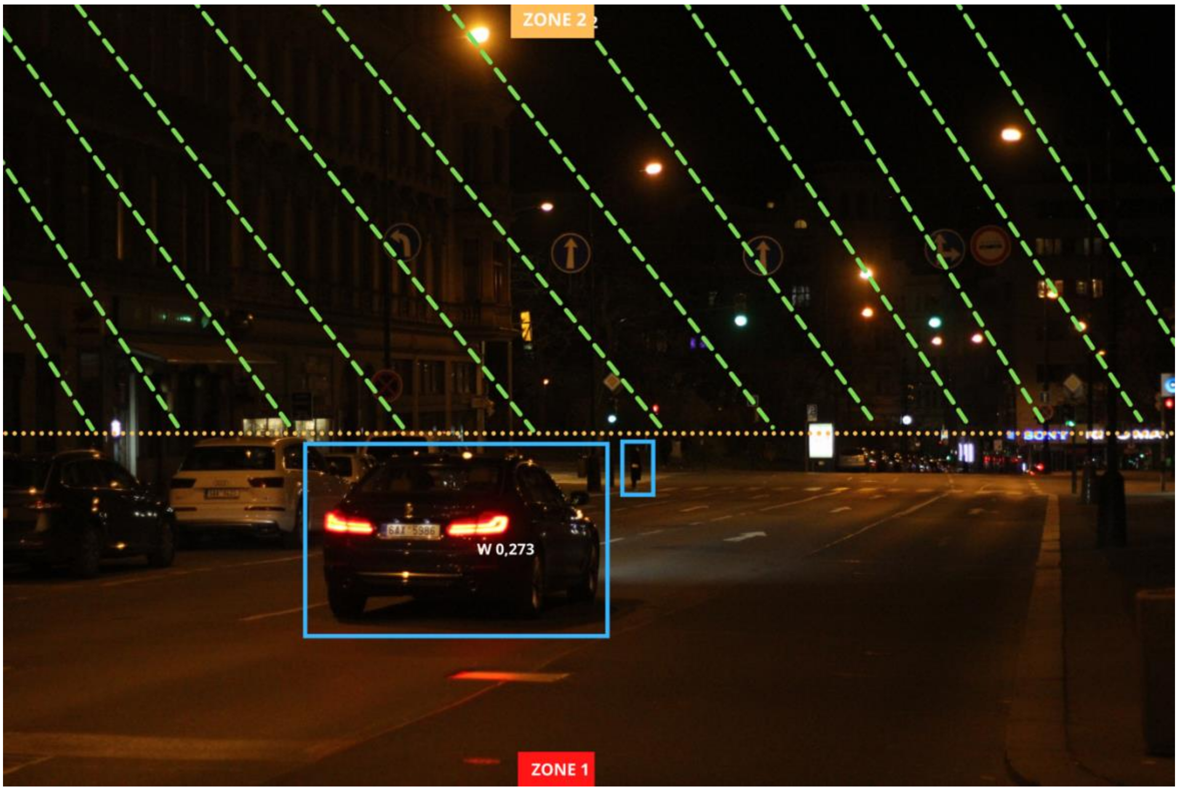


Figure XXIII: Žitná road - Nighttime [Location 8]

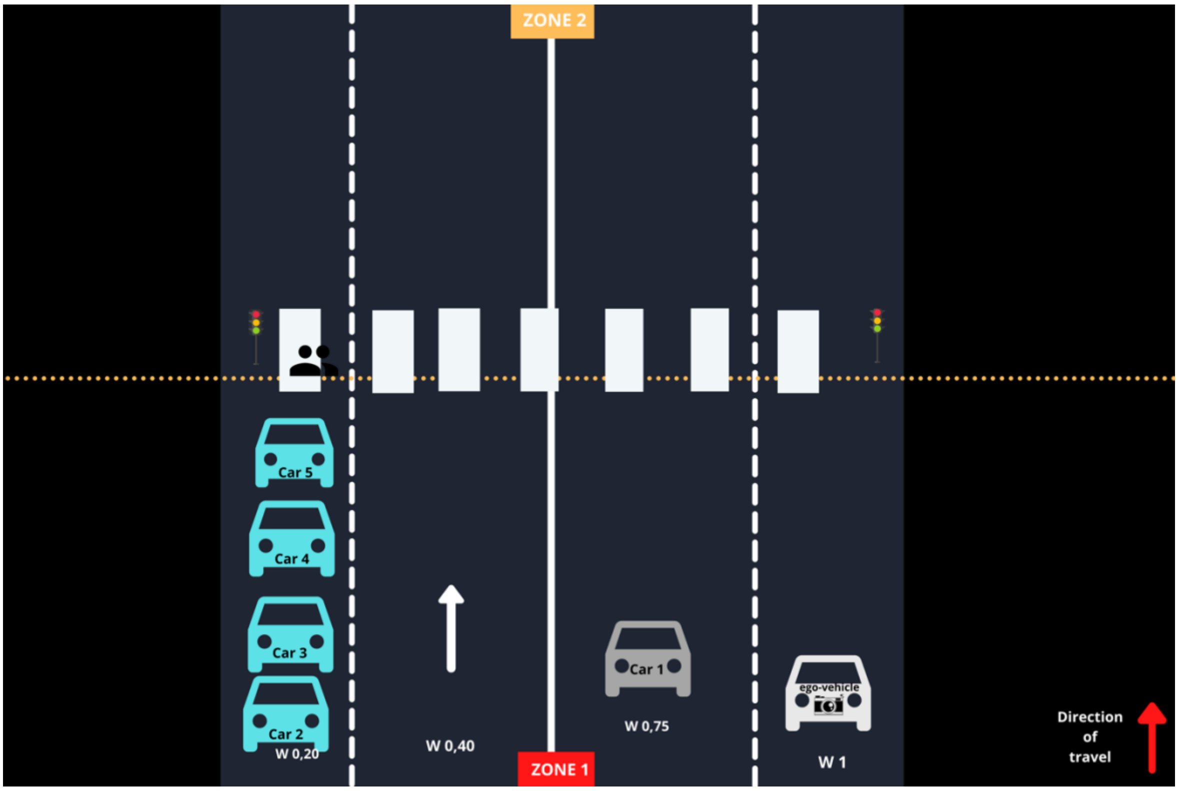


Figure XXIV: Top view - Nighttime [Location 8]

Appendix B: Image Size Calculations

RGB	Black & White
W = 1280 pixels H = 960 pixels BitDepth = 24 bits/pixel = (1280 x 960) x 24 bits/pixel = 1 228 800 pixels x 24 bits/pixel = 29 491 200 bits / 8 bits/byte = 29 491 200 bits x 1 byte / 8 bits = 3 686 400 bytes ≈ 3,68 MB	W = 1280 pixels H = 960 pixels BitDepth = 8 bits/pixel = (1280 x 960) x 8 bits/pixel = 1 228 800 pixels x 8 bits/pixel = 9 830 400 bits / 8 bits/byte = 9 830 400 bits x 1 byte / 8 bits = 1 228 800 bytes ≈ 1,22 MB
W = 640 pixels H = 480 pixels BitDepth = 24 bits/pixel = (640 x 480) x 24 bits/pixel = 307 200 pixels x 24 bits/pixel = 7 372 800 bits / 8 bits/byte = 7 372 800 x 1 byte / 8 bits = 921 600 bytes ≈ 0,9216 MB	W = 640 pixels H = 480 pixels BitDepth = 8 bits/pixel = (640 x 480) x 8 bits/byte = 307 200 pixels x 8 bits/byte = 2 457 600 bits / 8 bits/byte = 2 457 600 x 1 byte / 8 bits = 307 200 bytes ≈ 0,3072 MB
W = 320 pixels H = 240 pixels BitDepth = 24 bits/pixel = (320 x 240) x 24 bits/pixel = 76 800 pixels x 24 bits/pixel = 1 843 200 bits / 8 bits/byte = 1 843 200 x 1 byte / 8 bits = 230 400 bytes ≈ 0,2304 MB	W = 320 pixels H = 240 pixels BitDepth = 8 bits/pixel = (320 x 240) x 8 bits/pixel = 76 800 pixels x 8 bits/pixel = 614 400 bits / 8 bits/byte = 614 400 x 1 byte / 8 bits = 76 800 bytes ≈ 0,0768 MB

Table I: Image Size Calculations