Object Detection As A Safety Check For Human Factors In Operating Remotely Controlled Bridges

Graduation Thesis

Ernst de Groot | 1523570

Supervisor Zaanstad: Ing. C.L. Kraaijenbosch
Chair: Prof.dr.ir. P.H.A.J.M. van Gelder
1st Advisor: Dr. A.Y. Ding
2nd Advisor: Dr. D.F.J. Schraven

Construction, Management & Engineering
Delft Technical University
August 21st 2020

Preface

This master thesis is the result of 7 months of research concluding my time at Delft University of Technology. Although I've been interested in the field of artificial intelligence and data analytics for a couple of years now, and I did tinker around with it in my free time, this was my first deep dive into actually applying it in a real-world scenario. I'm glad I took on this challenge, to confirm to myself that this is indeed the type of work I would like to pursue professionally.

I would like to thank my committee for the opportunity to make the best of this thesis period by providing the essential knowledge and guidance, in a stimulating way. I would like to thank prof. dr. ir Pieter van Gelder for his ability to apply his extensive knowledge on safety theory to practical cases as the incidents in Zaandam. From the very first meeting, he showed genuine interest, philosophized about possible solutions, and by doing so, he gave me the comfort and motivation to take on this challenge. I would like to thank dr. Daan Schraven for his ability to decompose a problem and picking out the key points, in a very clear way. Also, his contributions to the readability of the report are very much appreciated. Lastly, but certainly not least, I would like to give special credits to Dr. Aaron Ding, as my first advisor. In our biweekly meetings, he contributed tremendously not only on the graduation thesis itself, but also on a personal level, talking about both present personal challenges, and future developments. This made a huge difference in both the process and the result.

I would also like to thank Ing. Kees Kraaijenbosch and the municipality of Zaanstad for providing the information, experiences and insights needed, to examine the possibilities of using an object detection support system. Before approaching Kees, I was hesitant whether the municipality was willing to cooperate in such a research, after all the attention from the media and the investigations done by the Dutch Safety Board. Their openness and willingness to investigate every possible way to increase the safety of the assets, were invaluable in bringing this research to a satisfying end.

Finally, I would like to thank my family, friends, and girlfriend, Fija, for supporting me in this difficult year, where I had to deal with the loss of my beloved father, Hans, and had to guide my struggling coffee business through the COVID-19 pandemic.

Ernst de Groot, August 2020

Summary

According to a report published by the Dutch Safety Investigation Board in early September 2019, the safety of remotely controlled bridges is not sufficient (Onderzoeksraad voor de Veiligheid, 2019, pg.58). This report was published after the occurrence of two severe accidents in Zaandam, on the Den Uylbrug and the Prins Bernhardbrug. On both occasions, the victims were standing on the movable part of the bridge deck during the opening of the bridge, and despite being visible for over a minute on the camera screens, were not observed by the operators, making the accidents human factor-based. The Dutch Safety Investigation Board concluded that part of the problem was safety mainly being considered a technical problem, instead of an integral one. In this research, the goal was to analyse how object detection could provide decision support for mitigating human factors for operating remotely controlled bridges. This was done by identifying the problems through literature studies, interviews and observations, and by building a proof of concept to mitigate these problems. Finally, this model was evaluated to gain experimental insights into the possibilities of object detection as a decision support tool.

The main problems identified are related to the operator's lack of situational awareness(SA). SA being "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in near future" (Endsley, 1988). When analysing the problem using Endsley's three-staged Situational Awareness Model, it showed that the difficulties were in the first level; the perception of elements in the current situation.

The model that was developed to help perceive the bridge users, was a custom trained Mask R-CNN ResNet 101 model, with images captured from CCTV systems from bridges in Zaandam, and its hit-rate was compared to a pre-trained model. The experimental insights show improvements of the custom model over the pre-trained model, and promising results for using object detection as a support tool to increase the safety of operating remotely controlled bridges.

Contents

List of Figures

Abreviations

ADR Action Design Research.

BIE Building, Intervention, and Evaluation.

CCTV Closed-Circuit Television.

CNN Convolutional Neural Network.

CP Central Post.

fps Frames Per Second.

GPU Graphics Processing Unit.

HMI Human Machine Interaction.

HOG Histogram of Oriented Gradients.

HOSS Human Operator Support System.

OVV Onderzoeksraad voor de Veiligheid.

ReLU Rectified Linear Unit.

SA Situational Awareness.

SCADA Supervisory Control And Data Acquisition.

SGD Stochastic Gradient Descent.

SVM Support Vector Machine.

1. Introduction

The Dutch Safety Board published a report early September 2019, stating that the safety of remotely controlled bridges is not sufficient(Onderzoeksraad voor de Veiligheid, 2019, pg.58). The minister of infrastructure and water management, Carla van Nieuwenhuizen addressed this issue in a letter to parliament, putting it on the political agenda (van Nieuwenhuizen Wijbenga, 2019). The research was conducted in response to an accident that occurred on the Prins Bernhardbrug in Zaandam, in November 2018. During the opening of the bridge, the presence of an elderly couple was missed by the bridge operator, despite being visible on the camera system for over a minute. After trying to hang on to the railing of the bridge, the couple couldn't hold it any longer, causing the man to fall in the water from approximately 20 meters, being unconscious after hitting a metal beam during his fall. Luckily, because of the decisive acting of bystanders, the couple survived the accident. Unfortunately, in February 2015, a similar calamity happened in Zaandam. This time the victim, a 57 years woman did not survive. This happened on the Den Uylbrug, a similar type of asset as the Prins Bernhardbrug. Here the technical side wasn't the problem either, but the decision making and the guidelines involved were at fault. (Onderzoeksraad voor de Veiligheid, 2016, pg.43).These are just two examples, but the OVV found at least thirteen more over the last 15 years. Many of these accidents were not caused by poor camera quality or malfunctioning mechanics/communication. Often human factors played a role in the happening of the event, also due to insufficient human - machine interaction(Onderzoeksraad voor de Veiligheid, 2019, pg.27).

In this research, the goal is to analyse whether the application of an object detection model could contribute to the safety of a remotely controlled bridge. Object Detection is a promising subsection within artificial intelligence, that has been developing rapidly over the last couple of years. As of late, models are developed with a strong combination of speed and accuracy that makes it an interesting technology to apply in the real world.

To analyse if this is a viable solution for the human factor-based problems as mentioned before, first the current situation will be described, to get an understanding of the way the remotely controlled bridges are controlled at the moment. Secondly, the interaction between humans and artificial intelligence applications/embedded systems will be covered. Subsequently, artificial intelligence will be discussed, with a focus on object detection.

After this literature study, the goal is to build a model that is capable of processing camera footage in real-time, reaching an accuracy that makes it likely to increase the safety on the bridge. First, the design and criteria of the object detection system will be discussed, based on user interviews and observations. Next up is the actual development of the model. Then, the application will be tested on speed and effectiveness, which leads to the discussion and conclusions. Finally, further recommendations will be stated to bring this research further.

2. Problem Definition & Research Questions

2.1. Problem Definition

The bridges the Dutch Safety Board focused on, are the Den Uylbrug and the Prins Bernhardbrug. Both are remotely controlled bridges located in Zaandam, and so are twelve others in the city, out of the total of twenty (Onderzoeksraad voor de Veiligheid, 2019, pg.13). The Prins Bernhardbrug is one of the most used bridges due to its connection with the A7, the longest national highway of The Netherlands(Rijkswaterstaat, 2019). The bridge was initially built in 1941, and was replaced in 2006. A significant feature of the asset is the strict separation between the bicycle lane and the motorway. The latter can also be found at the Den Uylbrug, which also crosses the river 'De Zaan'. For both bridges, the separate lanes aren't operated independently, but the same command is used to open both elements at the same time(Onderzoeksraad voor de Veiligheid, 2016, pg.20).

Because of the accidents that happened on these assets, the two bridges have frequently been in the news in the recent years. As a result, these bridges have been considered case studies in extensive discussions about safety of remotely controlled bridges, as at least thirteen more accidents happened in the last 15 years(Onderzoeksraad voor de Veiligheid, 2019, pg. 60). According to the reports on the Den Uylbrug and the Prins Bernhardbrug, the technical and legal requirements of the camera systems at the time of the accidents weren't the attention points, but mostly the human factors and the interaction between the users, bridge operator, and machine were(Onderzoeksraad voor de Veiligheid, 2019, pg. 27). In figure 1 screenshots are shown of the camera footage just before the casualties occurred, and in both cases the victims were visible on at least one camera for over a minute. Also noteworthy are the lighting conditions. Where dimly lit situations are presumably more difficult to interpret correctly, the accident on the Den Uylbrug happened at broad daylight.

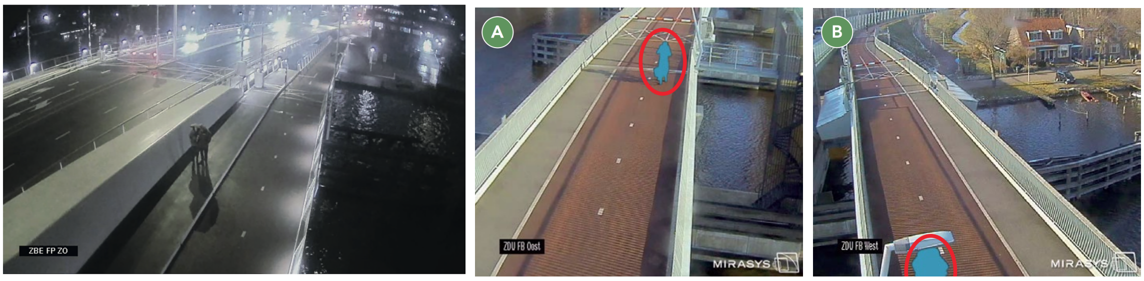


Figure 1.: *Screen captures Prins Bernhardbrug(1) and Van Uylbrug(2,3) (Onderzoeksraad voor de Veiligheid, 2019, 2016)*

In both accidents, the victims were convinced to be in a safe position, caused by confusing signing and lack of clear orientation points (Onderzoeksraad voor de Veiligheid, 2019). This can be seen as a human factor from the user side, but it shouldn't lead to an accident, because the bridge operator is required to scan the camera footage at the time of opening to be sure the situation is safe, before giving the definite command(fig 2).

Currently, measures are taken for the first mentioned human factors. This can be seen, for example, through the applications that are done in order to reduce confusion about the movable part and the non-moveable parts, by painting the bridge deck yellow. (NOS Nieuws, 2019). This marking will help the users of the bridge to orient themselves, but it will probably also add contrast to the images obtained by the cameras, making it easier to detect persons. Another measurement the municipality of Zaandam took, was placing additional cameras. This doesn't fully comply with the conclusions of the reports of the Dutch Safety Investigation Board. Additional cameras were recommended to get a better overview of the entire asset, but but it doesn't solve the risks regarding the human factors of the bridge operators. A deeper dive into the problems related to operating remotely controlled bridges can be found in section 5.1, where results of the interviews and observations are described.

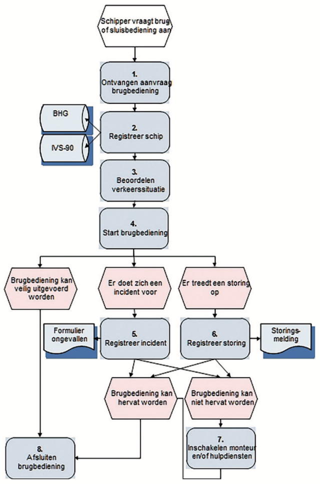

Schipper vraagt brug of sluisbediening aan
1. Ontvangen aanvraag brugbediening
BHG
IVS-90
2. Registreer schip
3. Beoordelen verkeerssituatie
4. Start brugbediening
Brugbediening kan veilig uitgevoerd worden
Er doet zich een incident voor
Er treedt een storing op
Formulier ongevallen
5. Registreer incident
6. Registreer storing
Storings-melding
Brugbediening kan hervat worden
Brugbediening kan niet hervat worden
7. Inschakelen monteur en/of hulpdiensten
8. Afsluiten brugbediening

Figure 2.: *Flowchart Bridge Operator (Onderzoeksraad voor de Veiligheid, 2019)*

A way to tackle these human factor based problems, could be achieved by implementing Object Detection. Object detection, an application of machine learning, is developing rapidly in recent years. Both the algorithms and the hardware side are improving in terms of speed, accuracy and accessibility. The main trade-off in object detection is between time en accuracy. The fewer calculations a model has to make, the faster it is, but it also makes it less accurate. On the other hand, an extensive model has to make a lot of computations, leading to a longer execution time(Huang et al., 2017). In the case of human detection on movable bridges, the same trade-off counts. This research should point out whether the speed and accuracy of such a model are already good enough to add value to the current situation.

However, within the field of Object Detections, multiple techniques coexist, and two methods in particular. The first method is background subtraction, histogram of oriented gradients feature extraction, and support vector machines combined. It is a relatively light-weight model that works well on static cameras, and has a good balance between speed and accuracy when it's properly configured. It's been the go-to technique for surveillance cameras for the last two decades, but in the last years the field moves towards the more computational-heavy technology of Convolutional Neural Networks(CNN). Although the model runs slower than the earlier mentioned technique, the accuracy of the new method has more potential. Because the Municipality of Zaanstad is already working with cameras manufactured by Bosch, the older technique is pretty accessible, as the cameras have Bosch Intelligent Video Analysis installed. This software is mainly used by security companies to detect unwanted trespassers, but hasn't been used on bridges yet.

In this research, both techniques will be discussed in the theoretical background chapter. However, because the Bosch system isn't available for testing purposes, it is difficult to compare both systems, apart from the theoretical differences. Therefore, this report will primarily focus on CNN based models. In the analysis, a comparison will be made between a pre-trained model, which is trained on the widely used COCO dataset, consisting of 1.5 million object instances in 80 different categories, and a custom trained model, using only images taken on remotely controlled bridges.

2.2. Research Questions

In this section, the questions are listed that should provide a holistic solution to the described problem statement. The main research question that should be answered in the thesis is the following:

How can Object Detection provide Decision Support for Mitigating Human Factors for Operating Remotely Controlled Bridges?

To answer this main question correctly, the following sub-questions can be stated:

1. How are remotely controlled bridges operated, and what are the human factors involved?
2. What are the most vital problems the proof-of-concept needs to solve?
3. How should the operator - object detection model interaction work?
4. To what extend is it possible to build a real-time proof-of-concept to perform object detection?

The first two questions are aimed to get a clear understanding of the current status, and to highlight the elements that keep the current situation from being the desired setting. The last two questions are about investigating the new situation, and the way the system should be implemented. To what extent is it possible to build a technically feasible system in the first place? Will the asset actually become safer, or do the new (secondary) risks outweigh the gains of the system? Together the sub-questions will form the basis to answering the main research question.

The methodology envisioned to be able to answer these questions in good order, is described in the next chapter.

3. Theoretical Framework

Introduction

In this chapter, the theoretical background of the related research domains will be covered, in order to get a feeling for the context of this report. The topics are ordered in the same way as the problem statement in the previous chapter, by first elaborating on the current situation and procedures(Bridge Operations), then on the identified problem area(Human Factors), and finally on the intended solution(Object Detection).

3.1. Bridge Operations

In this section, the current bridge operations will be discussed. First, the strategy and principles on which the operation procedure is built will be discussed, then the environment in which the operators need to do their work, and finally the operating steps the operator needs to take when controlling the remotely controlled bridge.

3.1.1. Operation Principles

The Ports and Waterways Department of the Municipality of Zaanstad has the primary mission to facilitate a fast and safe flow of the maritime traffic and an effective use of ports and waterways. Two of the main tasks to achieve this, is monitoring the activity on the water, and controlling the bridges and sluice of the Municipality of Zaanstad. Fourteen of these bridges are controlled from the Central Post(CP) in Zaandam (Draisma & van Heugten, 2014).

The operational model used to safeguard the risk management strategy is a logical and structured method for the identification, analysis, evaluation, handling and control of risks. It is developed in accordance with ISO-31000, and is applied to every level of the organization, from management functions, to the operators, and their primary work processes. The model is depicted in figure 3.

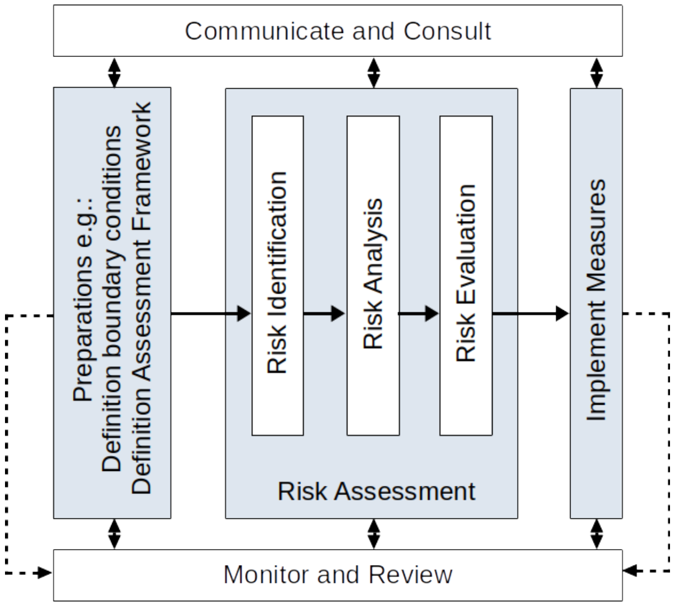


Figure 3.: *Operational Model Risk Management (Gemeente Zaanstad, 2017)*

The primary work processes are based upon the operation principles as described in the Noord-Holland's provincial decree on operating times and rules. These principles define the priorities of the different bridge users, and how they interact. Also, this decree outlines the situations where operation should be interrupted, or completely ceased. The most relevant principles for this research, are listed below.

Recreational Shipping
Within the standard operating hours, the operator is qualified to:

1. Open the bridge immediately when for both the recreational shipping and the road traffic the demand is low.
2. If the demand of the road traffic is high, regardless of the demand of recreational shipping, limit the bridge opening to once every half an hour.
3. When the demand of the recreational traffic is high, and the demand of the road traffic is low, wait for a few recreational ships to gather. In practice, this means a bridge opening every quarter of an hour.

Commercial Shipping
In case of a request by a commercial shipper, open the bridge as soon as possible, when the traffic situation allows it.

Connection
The connection of bridge openings isn't guaranteed. However, operators should try to achieve this as much as possible through mutual consultation.

Road Traffic
After a bridge opening, the operator should wait for the road traffic to normalize, before a new bridge opening can take place.

Process Cessation
The operation process is ceased in the following conditions:

1. Wind-speed of 8 Bft or more.
2. Visibility less than 50 metres caused by very thick fog.
3. When requested by the emergency services.
4. During big events.
5. By order of the supervisor.

Process Interruption
The operation process is interrupted in the following conditions:

1. When there is a great safety risk
2. When there is a malfunction or damage, making it unable to operate the bridge or to follow the guidelines.
3. In case of an incident on or around the bridge, making it impossible to operate the bridge, or to do so without leading to unacceptable risks.
4. When requested by the emergency services.

The decision to continue operation is taken by the operator, if the circumstances permit. When in doubt, the coordinator should be consulted.

3.1.2. Central Post

In order to design a good working human-machine interface, it should be clear in what kind of setting and organization the application is implemented, and how the organics of the workplace can be related to the problems perceived. Hereby, it should become apparent what the different implementation possibilities are, and how the current environment will be affected by the introduction of the new system. This research is applied to the bridge operation management of the municipality of Zaanstad. These operations are organized at the central post in Zaandam, from where the operators do their work. First, the house rules of the organization will be described, as these influence the behaviour of the operators. Thereafter, the current setup will be specified, to get a grasp of what the operators are working with on a daily basis.

House Rules

The central post consists of office spaces, meeting rooms, and the operation room. In this section, the house rules of the latter will be outlined, as these are most relevant for this research.

1. Concentration is required when operating the bridge. Side activities are forbidden(e.g. Answering the phone, conversations with supervisor or visitors).
2. During the entire process, the opening should be monitored on the screens.
3. No non-work related phone calls are allowed when working behind an operating desk.
4. No non-work related computer activities are allowed when working behind an operating desk.

5. Soft music is allowed
6. When working behind an operating desk, it is forbidden to read newspapers or magazines, use an e-reader or i-pad, or to have other materials on the desk.

Reading these house rules it is apparent that the primary focus is to minimize distractions, and maximize concentration, in order to operate the bridges as safe as possible. When an employee disobeys these rules, an official warning will be given. In case of a second occurrence, he is no longer allowed to work as an operator.

Resources

When the operator is working at his desk, he is responsible for a standard set of bridges. The following bridges are combined in the following order:

- Operator 1: Bernhardbrug, railway bridge, Alexanderbrug and Coenbrug.
- Operator 2: Julianabrug, Zaanbrug, Clausbrug and Prinses Beatrixbrug.
- Operator 3: Den Uylbrug, Schiethavenbrug, Nauernaschebrug and Reint Laan jr. brug
- Sluice Operator: Wilhelminasluice, Beatrixbrug and Wilhelminabrug.

The sluice operator works on a different kind of desk than the bridge operators. As the scope of this research are remotely controlled bridges, this setting will not be discussed. For the bridge operators, the desks are all similar. However, only one desk setup has some added features to be able to operate the railway bridge, but all other common elements are depicted in figure 4.

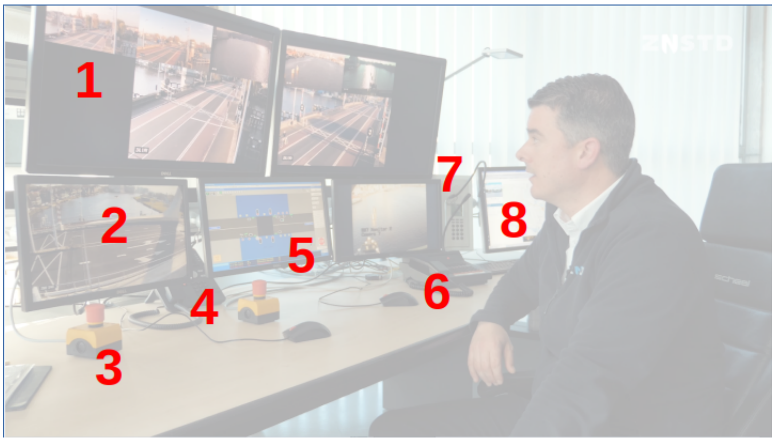


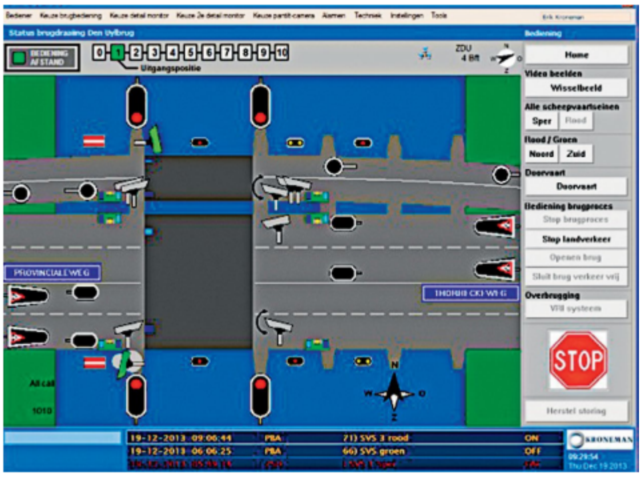
Figure 4.: *Operation Desk (Gemeente Zaanstad, 2019)*

#	Description	#	Description
1	Main monitor	5	SCADA Operating System
2	Side monitor	6	VHF maritime radio-telephone
3	Emergency Stop	7	Intercom
4	Telephone	8	Computer

As displayed in figure 4, the operator has four screens to his disposal for monitoring the situation on the bridge via the camera system. Every screen contains multiple camera streams. The operator can choose which streams to display in the big frame on the main monitors, and which in the small frames.

For communication, the operator has an intercom, a telephone, and a VHF maritime radio-telephone. The intercom is used to communicate directly with the users on the bridge, for when people neglect the red lights or hinder the opening procedure in a different way. The VHF maritime radio-telephone is used to communicate with shippers. This way, the shippers can make a request to open the bridge, and the operator can inform them about the process. Lastly, the telephone is used to contact external parties, like emergency services.

For operating the bridge, the SCADA(Supervisory Control And Data Acquisition) system is used. A screenshot of the SCADA system is depicted in figure 5. Here, all commands needed to operate the bridge can be controlled. Also, with this system, the operator can monitor the current configuration of the asset, read the log-files generated by the system, and information about the weather conditions is displayed. Besides the SCADA system, an external emergency stop is present, to cease ongoing processes in case of a dangerous situation.



```
19-12-2013 09:06:44   PBA   71) SVS 3 rood    ON
19-12-2013 06:06:25   PBA   66) SVS groen     OFF
```

Figure 5.: *SCADA system (Onderzoeksraad voor de Veiligheid, 2016)*

3.1.3. Operational steps

After having gained insights into the workplace and the overarching principles, the next step is to look into the actual opening procedure itself. The flowchart for operating the bridge is depicted in figure 6, where the blue elements are the existing steps, and the green elements are the envisioned additional steps of the object detection model. In appendix D, the workflow is described in a more detailed way. For each step in the process, it is reported what the operator should observe, what he should control, and what he has to monitor.

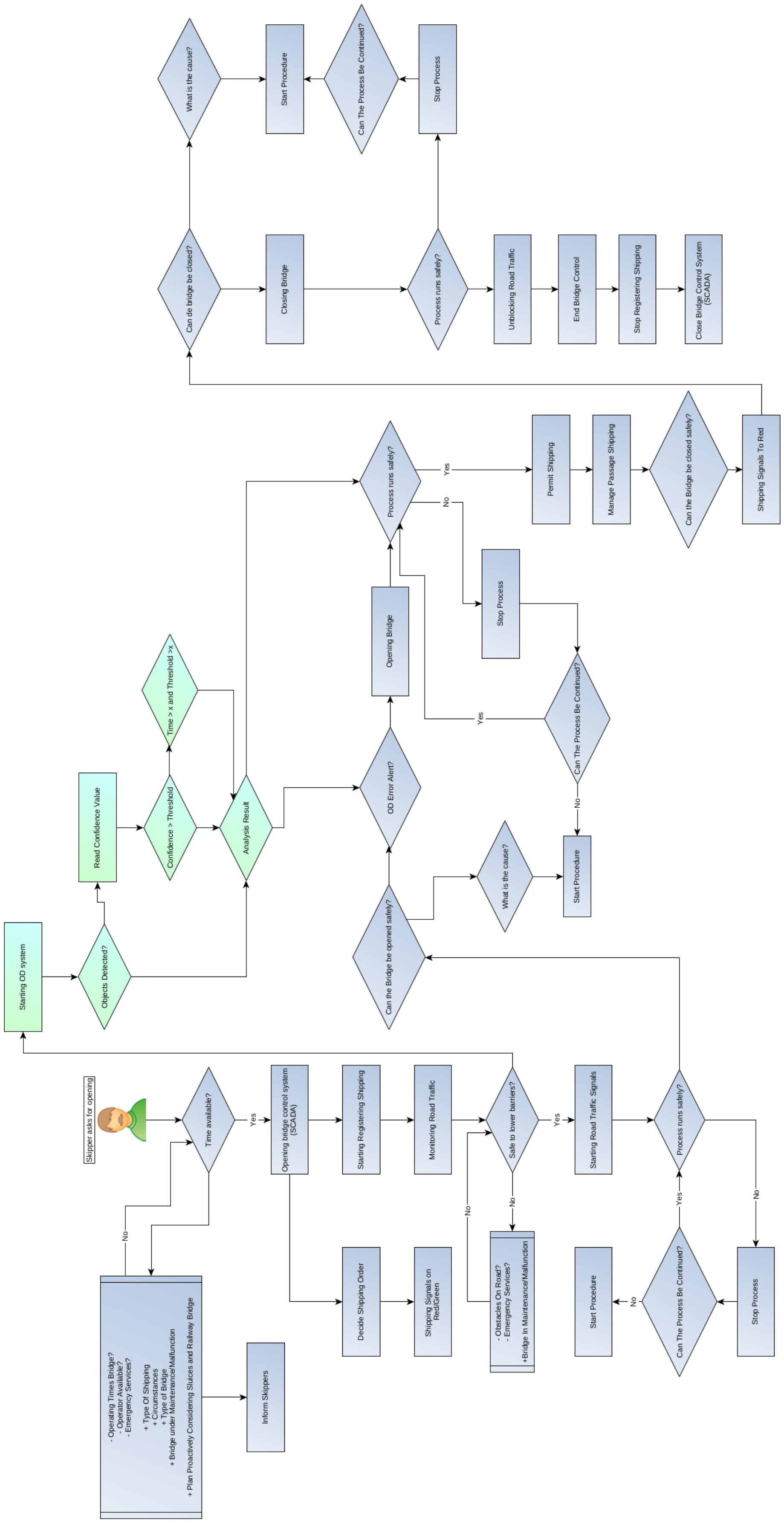


Figure 6.: *Workflow Bridge Operation (Draisma & van Heugten, 2014)*

3.2. Human Machine Interaction

In this section, the field of human machine interaction will be discussed. This is needed to make sure the desired object detection support system integrates well in the current situation as described in the previous section. However, before diving into this, human factors and related human studies will be discussed, forming the foundation for the human machine interface.

3.2.1. Human Factors

Before diving into the researches conducted in the field of human factors, a definition should be presented to provide a better upstanding of its scope. In this report, the definition as used by the International Ergonomics Association will be used:

"Ergonomics (or human factors) is the scientific discipline concerned with the understanding of the interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimise human well being and overall system performance." (International Ergonomics Association, 2000)

The main trigger that started the endeavours of studying human factors came from the technological developments during World War II. Weapon and transport systems grew in complexity, and great technological advances were made in factory automation and in equipment for common use. Through the difficulties encountered while working with the complex machinery, the need for human factors analyses became evident. These studies were procedeed by research on human physiology, industrial engineering, and human performance psychology (Proctor & Van Zandt, 2008).

When it comes to human factors studies specifically for bridge operators, let alone operators of remotely controlled bridges, there is not much to be found. However, it is possible to draw parallels between bridge operations, and other fields where operations are involved, such as the aviation sector. Here, the pilots and air traffic operators are also decision-makers in a complex traffic-related field. In 1996, the bureau of air safety investigation researched 75 fatal aeroplane accidents, and it turned out that in 70% of the occasions, the pilot's human factors were involved. Most of these factors were related to poor judgement and decision-making(fig. 7). The research also showed that the errors of judgement can be made by both experienced and inexperienced pilots. According to Endsley, the main reason for this poor judgement and decision makin in aviation is the lack of situational awareness (Endsley & Robertson, 2000). Researches conducted by Hartel, Smith, Prince and Merket, Bergondy, Cuevas-Mesa also identified inadequate situational awareness as one of the primary factors in accidents attributed to human error (Hartel, Smith, & Prince, 1991; Merket, Bergondy, & Cuevas-Mesa, 1997).

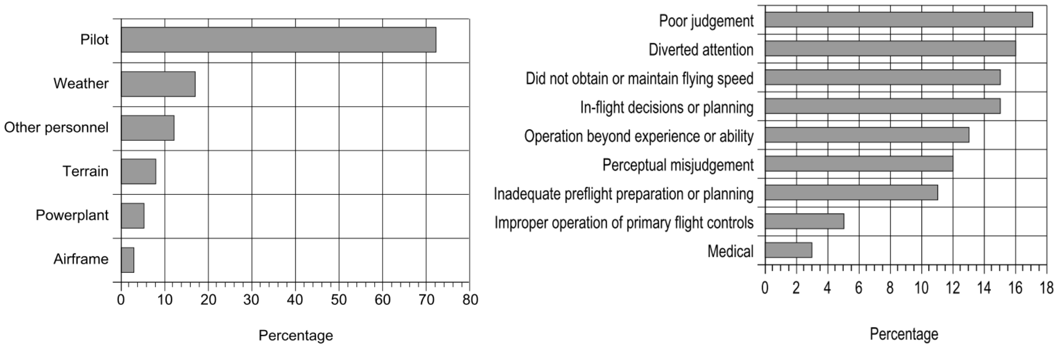


Figure 7.: *Fatal accidents to fixed wing aircraft. Broad Accident Factors(l) & Pilot Factors(r) (Bureau of Air Safety Investigation, 1996)*

Situational Awareness

The overarching theory for implementing an automated object detection system as a decision support tool, is related to enhance the operator's situational awareness(SA).To get the importance of this phenomenon, it is needed to understand the key concept of the type of accidents that are often regarded as casualties caused by human errors. The danger of pointing at human error as the cause of an accident, is that it can be a misleading perspective. One could easily think that it implies that people are merely careless or poorly trained, or that they are not reliable in the first place. However, in many of the accidents deemed as human errors, the human operator is striving against significant challenges, as he is working with a data overload and the challenge of dealing with a highly demanding complex system. They have to keep to a long list of procedures and checklists in order to get the system under control, which sometimes fails. A standard reaction of the industry, would be adding even more procedures and systems, which makes the system even more complex and more challenging to operate. To summarize this, one could state that the operator is held accountable for whatever failures and inefficiencies embedded in the system. In the end, the operators often are perfectly capable of performing their tasks physically, and have no difficulty knowing that is the correct thing to do, but they continue to be stressed by the tasks of understanding what is going on in the situation (Salvendy, 2012). It may feel contradictory to mention the flaws of the common industry's reaction of adding new procedures and systems in a report where an object detection sub-system is proposed. That is why the focus of this report is not solely on the technical side of the model, but equally on the HMI side of the potential solution, as the awareness exists of the possibility of doing more harm than good with an added element.

To conclude, this critical task of understanding what is going on in the situation is regarded as situational awareness, a research domain originated in the aviation domain. A widely used definition is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in near future" (Endsley, 1988). How situational awareness is part of the decision-making process for operators, is depicted by Endsley's situational awareness model(figure 8). After the figure, the individual steps will be explained.

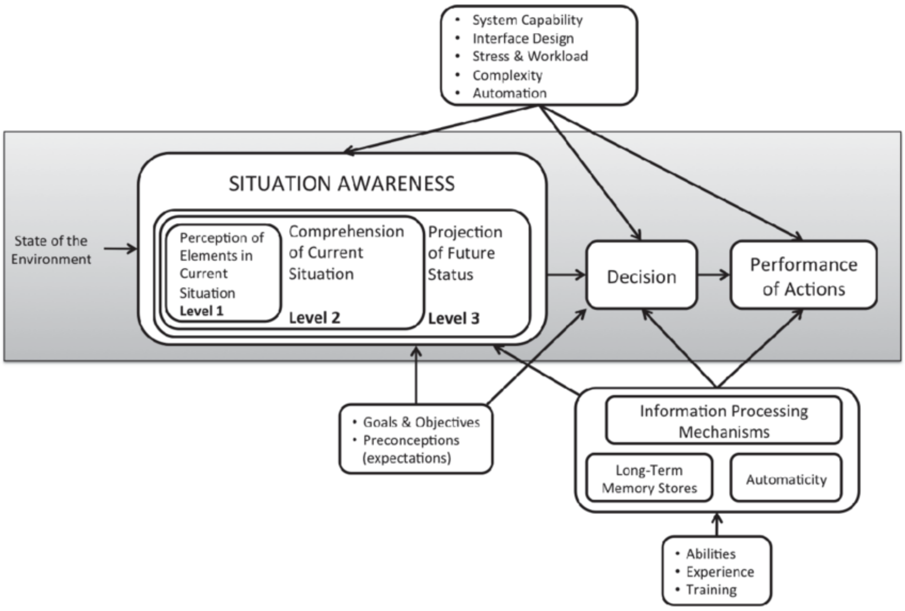


Figure 8.: *Endsley's Situational Awareness Model (Endsley, 1995)*

Level 1: Perception of Elements in Current Situation
The first step in the situational awareness model is to perceive the status, attributes, and dynamics of relevant elements in the environment. In case of the bridge operator, there are a lot of elements to consider. In the office alone there is the information represented by the SCADA system showing the current conditions of the bridge configuration, a display with emergency services, colleagues working on other bridges, indicating incoming ships, shipping requests from all controlled bridges over VHF maritime radio-telephone, and now and then the system picks up requests meant for bridges outside of Zaanstad's region. Besides these elements, there is of course the situation on the bridge that needs to be operated. Different kinds of traffic react in different ways, and have different characteristics to work with, like form, color, speed, and behaviour. This applies for both water and land traffic.

Level 2: Comprehension of Current Situation
Level two is the comprehension of the situation based on a synthesis of disjoined level 1 elements. In this step, the move is made from knowing about the presence of the elements to analysing the significance of the elements in light of the operator's goals. An operator could spot a person in the first level that is responding to the warning lights, and act accordingly by waiting for the barriers to lower, and the bridge to open. This person isn't threatening the goals of the operator, as he can still open the bridge safely, and optimize the traffic flow. However, a person that is standing on the movable bridge deck after the operator has given the command to open the bridge, is of much more significance, as this does obstruct the operator's goals.

Level 3: Projection of Future Status
The third step is one that benefits from the operator's experience, and this is by spotting an element, and the ability to predict how it will behave shortly after. If a vehicle is approaching recklessly, it is no immediate threat at its current position, but a projection of the future status could indicate that it will endanger the operation in a later stage.

This projection of future status will provide extra decision time for the operator.

Change Blindness and Inattentional Blindness

Without the first step in Endsley's model, perception of elements in the current situation, situational awareness can't be reached. The problem is that the perceptual system doesn't always respond appropriately to changes within the visual environment, even if there is a detailed picture of the scene. So, even if a person is visible on the camera footage, and the operator is monitoring this screen, it is not certain that this person will be perceived. The two phenomenons that are related to this are change blindness, and inattentional blindness.

Change blindness is 'the surprising failure to detect a substantial visual change', and inattentional blindness 'the failure to notice an unexpected, but fully-visible item when attention is diverted to other aspects of a display'((Jensen, Yao, Street, & Simons, 2011)). A famous example of the latter is the 'gorillas in our midst' experiment conducted by Simons and Chabris((Simons & Chabris, 1999)), where the observers were shown a videotape with the task to count the number of passes between two teams of three basketball players. Within this 75 seconds long video, a gorilla walked through the frame, being clearly visible for 5 seconds, as shown in figure 9. Around 50% of the observers did not spot the gorilla because of inattentional blindness.

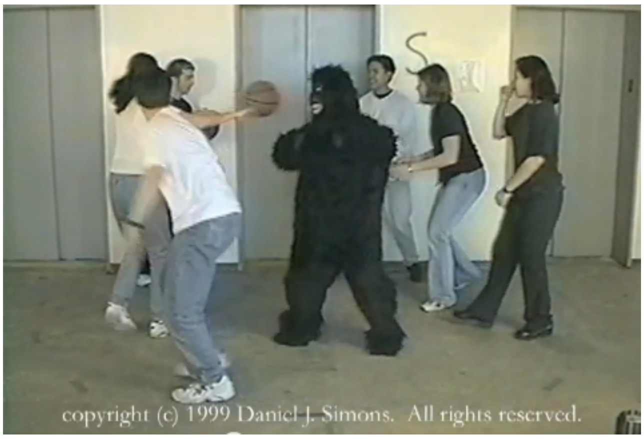


Figure 9.: *Inattentional Blindness (Simons & Chabris, 1999)*

So how does this translate to perceiving the situation on the bridge? When an operator is monitoring the displays with an expected shape of a person or vehicle in mind, it is possible to miss out on an instance of interest, because the appearance of this instance deviates from the expected shape. Take the incident at the Bernhardbrug for example, the victims were standing very close to each other, in dark clothes, covered by a dark umbrella. When an operator is monitoring with a more distinct appearance in mind, the chance of detecting these victims could have been decreased. It could also be that an operator is looking out for maritime traffic, because of a ship having technical difficulties for example. In this case, the operator's attention is diverted to this situation, making him inattentionaly blind for anomalous pedestrian behaviour.

Where these scenarios are imaginable, the phenomenon of change blindness is probably a bigger risk when operating based on cameras, as more complex processing is typically required for successful performance in change blindness tasks than inattention blindness ones (Eysenck & Keane, 2015). According to Jensen et al., five separate processes must be engaged successfully by the operator, for change detection to occur (Jensen et al., 2011).

1. Attention must be paid to the change location.
2. The pre-change visual stimulus at the change location must be memorized.
3. The post-change visual stimulus at the change location must be memorized.
4. The pre- and post-change representations must be compared.
5. The discrepancy between the pre- and post-change representations must be recognised at the conscious level.

Often, in the real world, people are aware of changes in the visual environment because they detect motion signals accompanying the change (Jensen et al., 2011). This makes it more difficult to detect changes when the motion signals are missed.

When monitoring a bridge, the operator needs to keep his eyes on several camera displays. Because of this, the operator could miss the motion signals of a person entering the bridge on one screen, when monitoring another. If this person stops moving, no additional motion signal will be visible, possibly leading to change blindness. Subsequently, leading to a possibly unsafe situation. Research by Durlach, showed that when multiple events occur simultaneously, operators can fail to detect important changes even when they are not fatigued, stressed or multitasking (Durlach, 2004). When looking at both accidents discussed in this report, in both occasions the victims were waiting for the bridge to open with minimal movement, while the operator had to monitor and control the complex opening procedure. This combination could possibly have led to change blindness, making the operator unaware of their presence. A change detection tool could mitigate this risk (Durlach, 2004). Using an object detection system could possibly achieve this.

3.2.2. Human-Machine Interaction

Within the field of human-machine interaction, the need of identifying users' emotional and social drives and perspectives is recognised. This means getting familiar with their motivations, expectations, trust, and social norms, and relating these topics to work practices, communities and organisational social structures as well as organisational, political, and economic drivers. Therefore, HMI researchers turn to qualitative research methods to get insight into these elements, and to understand the qualities of a particular technology and how people use it in their lives. These subjects are difficult to put into numbers, and because of that quantitative research will not cover the issues properly. (Adams, Lunt, & Cairns, 2016)

Ethnographical field study techniques(observation and contextual interviews) are employed to provide qualitative data about potential and/or actual users of the product, with the main principle to come up with a set of behavioural patterns, that help categorize modes of the use of the potential product (Cooper, Reimann, & Cronin, 2007, pg.20).

Personas

To help categorizing the modes of use, and use them in the building process, personals can be used. Personas help thinking and communicating about how users behave, how they think, what they wish to accomplish, and why. These are based on the behaviors and motivations of real people as observed during the research phase, and they're used to represent them during the design phase (Cooper et al., 2007, pg.75). When a certain type of persona is needed to complete the entire set of future users, but a representative person can't be found during the research, a provisional persona could be made. Provisional personas are structured similarly to real personas but rely on available data and designer best guesses about behaviors, motivations, and goals. Although they are based on assumptions opposed to qualitative research, making use of provisional personas yields better results than no user models at all(Cooper et al., 2007, pg.87).

The process of creating personas, can be divided into the following steps(Cooper et al., 2007, pg.97/98):

1. Identify behavioral variables
2. Map interview subjects to behavioral variables.
3. Identify significant behavior patterns.
4. Synthesize characteristics and relevant goals.
5. Check for redundancy and completeness.
6. Expand description of attributes and behaviors.
7. Designate persona types.

Represented Models

The insights in the user's behavior and motivations that are captured in the personas, help understanding how to represent the working of the machine or application to the user. By representing it in a way that is close to the way the user perceives the working of a system, the user's understanding of the system will be increased. The three types of models that are relevant when building the represented model are the following:

Implementation Model
The implementation model, also known as the system model, is the representation of how a machine or a program actually works.(Cooper et al., 2007, pg.30). In the case of an object detection model, this describes the complete process of gathering image data, the way of processing these frames to get to the predictions, and the steps taken to present these detections to the operator.

Represented Model
Generally, users don't need to know all details of how a complex model actually works, so they come up with a cognitive shorthand for explaining it, which is extensive enough to cover their interactions with it, without having to know the entire inner mechanics (Cooper et al., 2007, pg.30). For this object detection application, for example, the operator doesn't have to know how a convolutional neural network manages to detect and classify objects. As long as the operator regards this system as a tool that checks camera footage for objects and reports it to them when necessary, he will be able to interact with

it.

Mental Model
The represented model is the way the designer chooses to represent a program's functioning to the user(Cooper et al., 2007, pg.30). The closer this represented model comes to the mental model of the user, the easier it will be for the user to understand and use the program, as depicted in figure 10. In the implementation model, the object detection system will work with classes and scores, depicting the certainty of a prediction, and when that score is higher than a stated threshold, at a certain moment during the bridge operation process, the detection will be presented to the bridge operator. This threshold may change for different light/weather conditions. The worst represented model would ask the user to change the values appropriately, a better model would work with different setting for the operator to choose from, based on the weather conditions, but the best model would just fetch the weather information itself, and adapt to it. This would come closest to the mental model of a system that accurately detects objects.

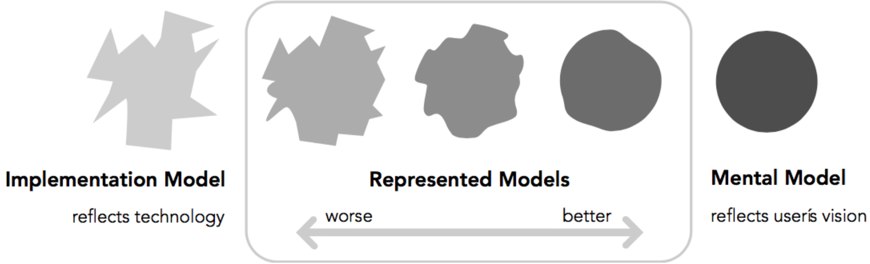


Figure 10.: *Represented Model Positioning (Cooper et al., 2007, pg.30)*

Human Centered Automation

To further enhance the operator's understanding of the system, the Human-Centered Automation principles as put together by (Wickens, Lee, Gordon-Becker, & Liu, 2014) should be considered. Safeguarding these in the design, should achieve maximum harmony between human, system, and automation. Below, the six principles are listed, stating both the importance of the principle, and the way of adapting to the object detection model.

1. Keeping the Human Informed
Humans should have the 'big picture' of the process. To be able to have this, the human should be informed of what the automation is doing, and why. Therefore, the underlying processes of an automated task should be well displayed.

For the automated detection system, the operator should be able to see on the basis of what the system decides to notify. This way, the operator could prevent false positives and false negatives of unnecessarily stopping processes, or putting people at risk, as the outcome of the automated model is not always in line with the actual situation.

2. Keeping the Human Trained
When working with an automated system, work becomes more abstract as it depends upon understanding and manipulating information generated by the system (Zuboff, 1988). In case of the operator, it could be a choice to communicate the prediction scores/confidence

rates of the detected instances. The operator should then judge the presented information by taking into account the limitations and possibilities of the system. For example, with a confidence rate of 0.4 in broad daylight, the probability of it being a false detection, is way bigger than the same confidence value on a rainy evening. With updated features and other changes to the system, the operator should be aware of the changed conditions and their implications. This is also where the Human Operator Support System(HOSS) dilemma makes its entrance(Wieringa & Wawoe, 1998). A HOSS reduces the task complexity with the goal to reduce the mental load experienced by the operator, in this case, the stress of being the only set of eyes that needs to spot a user on the bridge in times of opening. However, decreasing the task complexity may lead to an increase of the system complexity by adding the HOSS. This sub-system is also part of what the operator needs to understand. Especially during stressful moments, not fully understanding this HOSS may end up causing a higher experienced workload for the operator, making it counter-productive.

For a HOSS that is able to reduce the task complexity without adding to much on the system complexity side, there is not much too worry about. However, adding a deep learning computer model does increase the system complexity by a lot. For many people, such a machine learning model is a black box, and even when familiar with the working of such an algorithm, it is still difficult to diagnose certain behaviour.

However, training on the understanding of the automated system is not the only meaning of the 'keeping the human trained' principle. It also goes for taking over the automated tasks in case the system fails. Take for example an automated pilot. When the system fails, the actual pilot should still be able to fly and land the plane safely. In moments like these, it is important to have an operator that is experienced and trained in dealing with the manual procedure. This is somewhat of a catch 22. A failing autonomous system could end up in a critical situation where an experienced operator is needed to get out of the situation. However, outside of experiencing these kind of situations, it can be difficult to train on them.

3. Keeping the Operator in the Loop
This principle is focussed on keeping the operator's situational awareness high, so the operator is able to jump back into the control loop when the system fails. In this case, it could mean that if the detection system could perform all steps in the perceptual, cognitive and action stages, the operator's situational awareness would degrade, making it difficult to manually override the system in case of opening the bridge when dealing with a false negative.

4. Selecting appropriate stages and levels when automation is imperfect.
When designing a human-machine interface, the level of automation and the stages where it is active should be considered. In terms of the bridge opening procedure, it is possible to distinguish four basic steps. Detecting instances(step 1), judging the instances based on confidence, type, situation (step 2), recommend appropriate action(step 3), and perform the actual action(step 4). When the automation is imperfect in the early stages(step 1 and 2), the operator can work around it in the later stages, and not much harm is done. However, when the imperfection takes place in the last stages, for example opening a bridge because of a missed instance, the effects can be much more serious.

5. Making the automation flexible and adaptive.
Autonomation studies showed that the amount of automation needed to vary from person to person and to the same person over time (Wickens et al., 2014). This is particularly the case in situations that are not fully predictable. A parallel could be drawn with cruise control in a car. Some drivers prefer to use it, others don't, and for some people their preference could change for the type of traffic situation. In the case of the object detection system, for example, it could be a choice to make the confidence threshold of the system flexible, so the operator can decide to accept more false positives, because he experiences challenging conditions, where the need for an automated support system is high.

Another option is to make the object detection system use adaptive automation, a form of automation that can adjust its method of operation based on changing situational demands(Scerbo, 1996). The level of automation can be steered by certain user, task, and environment based conditions. There are numerous ways of making the object detection algorithm adaptive. The system could for example determine the operator's age, and subsequently adjust the estimated reaction times of the operator, as age affects reaction time and movement time (Light, Reilly, Behrman, & Spirduso, 1996). Taking it one step further, it is also possible to monitor heart rate and body temperature to monitor workload imposed on the operator and fatigue. Doing this, the system can adjust the level of automation for reducing the workload, or increase the intensity of notifications and lower the confidence threshold to make up for the reduced attention of the operator caused by fatigue.

These options may have their benefits, but there are also some reasons for not implementing elements with such adaptive character. Again, the system complexity increases, and dealing with rapidly changing system configurations may negatively affect the operator's understanding and comfort when working with the system (Wickens et al., 2014).

6. Maintaining a positive management philosophy
Besides the technical challenges of automation, there is also a social part that is important to keep in mind. Operators could see the new technology as a system that eventually replaces them, and the idea of cooperating on bringing a system further that possibly makes their jobs redundant in the future can count on resistance. According to a survey conducted by HR agency ADP in 2018, 23 percent of the 1300 Dutch respondents feared losing their job to automation in the near future (ADP, 2018), this phenomenon, also known as 'Automation Anxiety', has been around for ages. An early example was the strike of New York's lamplighters on the night of April 24, 1907, because their work lost value due to the uprise of electric streetlights (Frey, 2019). A positive management philosophy should make clear that the automated sub-system is designed for support instead of replacement. Highlighting that the human remains the master, and the automation the servant increases the chance of acceptance (Billings, 1996).

On a more conceptual level, there is an element that plays a big role in both accepting a new human colleague and accepting a machine as support, and that element is trust. Trust is a major determinant of reliance on and acceptance of automation, standing between people's beliefs towards automation and their intention to use it (Ghazizadeh, Peng, Lee, & Boyle, 2012). The elements where human-automation trust and interpersonal trust are based upon, however, differ. Interpersonal trust can be based on the ability, integrity, or benevolence of a trustee (Mayer, Davis, & Schoorman, 1995), whereas human-automation trust is based upon performance, process, or purpose (J. D. Lee & See, 2004). On the performance side, if the operator understands that in challenging conditions, an extra set of eyes could make the asset safer, and lower the experienced workload, the trust increases. On the process side, the operator will remain the decision-making permissions, and also the monitoring of the bridge isn't replaced by the system. So, the expertise and experience of the operator will remain its importance. Lastly, on the purpose side of the system, its sole purpose is to support the operator to make the work less stressful, and making the bridge safer. Keeping these elements in mind when introducing and managing the system, the chances of acceptance will increase.

Cry Wolf Syndrome / Alarm Fatigue

For the implementation of an object detection system, an effect that is interesting to research is the effect of false detections on the operator's behaviour. This is closely related to the so-called Cry-Wolf syndrome, named after Aesop's fable "The Boy Who Cried Wolf", also known as alarm fatigue.

Alarm fatigue is referring to the sensory overload when clinicians are exposed to an excessive number of alarms, which can result in desensitization to alarms and missed alarms(Sendelbach & Funk, 2013). A sector where this problem has been dominant in the last decades is Health Care (ECRI Institute, 2020). The research demonstrated that 72% to 99% of clinical alarms are false, which led to alarm fatigue.

From which point false alarms will seriously affect the reliance of an operator on the system is not certain. Researchers vary from stating there is no effect at all (Wickens et al., 2009) to stating it can jeopardize safety (Ruskin & Hueske-Kraus, 2015). According to Mileti and Sorensen the effectiveness of people's responses to warnings is not diminished by the so-called 'cry wolf' syndrome, as long as they've been informed of the reasons for the previous "misses". False alarms, if explained, may actually enhance the awareness of a hazard and its ability to process risk information. (Mileti & Sorensen, 1990).

Response Time

When designing the HMI for the Object Detection Model, it is important to be able to go from the perceptual stage, via the cognitive stage to the action stage as quick as possible. For this system, this means going from perceiving the detected instance by the model, to performing the bridge commands as soon as possible. Below, the implications for each phase are described.

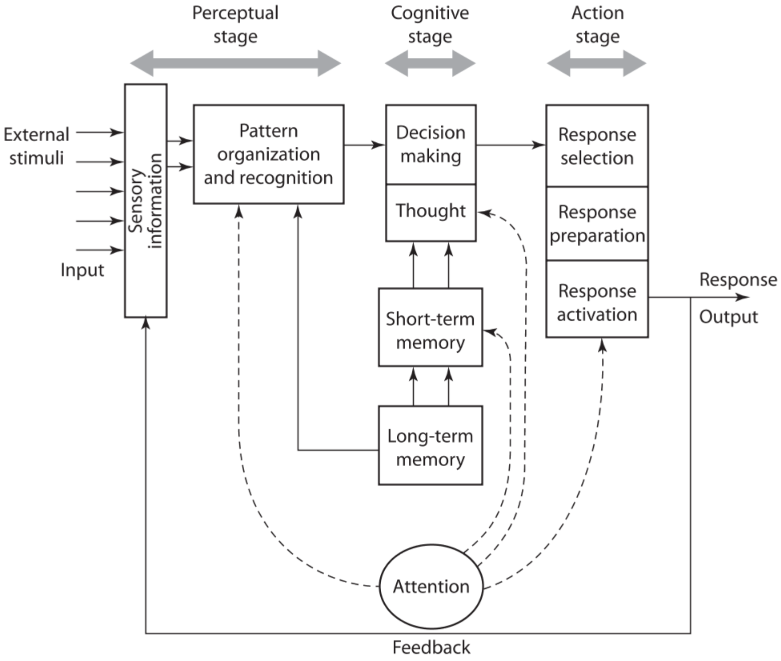


Figure 11.: *Flowchart Bridge Operator (Proctor & Van Zandt, 2008)*

Perceptual stage
This is the stage where the information provided by the object detection system, should be displayed to the operator as clear as possible. Different options of displaying a detected instance are depicted in image 12. The top two variants are able to notify the operator that on that particular screen an instance was detected. The area of focus doesn't go into more detail than that, so the remaining effort for the operator to see what the system actually detected, can be substantial. In case of a false positive, it will not become clear what the detected object was, which decreases the operator's understanding of the system. This happened during the pilot in March with the Bosch system. The screen only turned red when an instance was detected so the operator had to search for the instance of interest. The system had approximately one false alarms every time the bridge opened according to one interviewee, and combined with the unclear way of presenting information, the operator lost the sense of reliance and compliance with the system, and switched it off.

For the bottom two alternatives, the focus area is a lot smaller, and point directly at the detected instance. If there is a false positive, the operator sees it straight away, and potentially sees the familiarities with the category that should have been detected, understanding why the system reacted the way it did.

In case of an actual correctly detected instance, the main difference between the display forms is the time it takes to spot the instance of interest. Reducing this time to a minimum could be crucial in a dangerous situation. The respondents of the survey preferred the option in the bottom left corner, but to ensure the right option is used, empirical research should be done by measuring the time it takes for an operator to point out the detected instance.

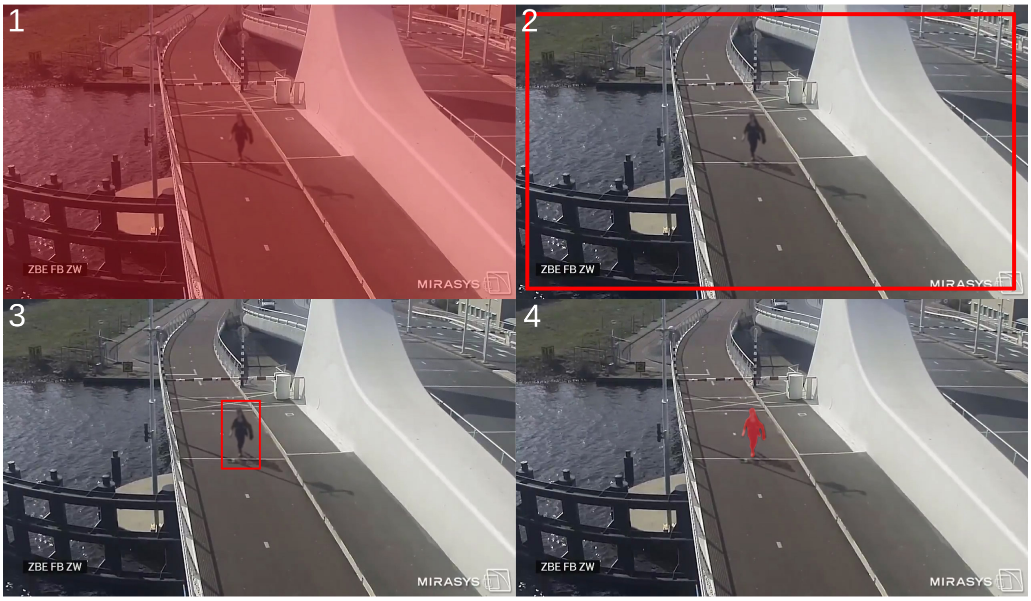


Figure 12.: *Detection Options*

Cognitive stage
In this stage, the operator needs to make a decision based on the perceived information in the previous stage. As response time is still critical, the time needed to make a decision should be minimized as well. This is where Hick's Law, also known as the Hick-Hyman law, is implemented. Hick's Law states that the time required to make a decision is a function of the number of available options (Lidwell, Holden, & Butler, 2010). It demonstrates that a person's reaction time increases as the number of choices increases (Erlandson, 2007). This applies to simple decision-making tasks in which there is a unique response to each stimulus. For example, there are bridges where the bridge decks for motorized and non-motorized transport are separated, and controlled separately. If there are also two emergency buttons, one for each section, the reaction time will be longer than when one emergency button is used, as the number of options is bigger.

Action Stage
Where Hick's Law dealt with the reaction time of the operators, Fitts' law does the same with movement time. According to Fitt's Law, the smaller and more distant a target is, the longer it will take to move to a resting position over the target. Also, it states that the faster the required movement and the smaller the target, the greater the error, because of the speed-accuracy trade-off (Lidwell et al., 2010). For emergency buttons, to ensure rapid movements, it is therefore wise to place them near the operator and/or making them large. A downside of this could be that the chances of accidentally activating them are higher, so a balance needs to be found.

Sensory Psychophysics

Part of working with human machine interface is the understanding of how humans are able to perceive and act on information arriving through the senses. Ernst Weber and Gustav Fechner founded the study of psychophysics and are regarded as the fathers of modern experimental psychology(Proctor & Van Zandt, 2008). They showed how controlled experimentation could reveal the characteristics of human performance. Weber researched the ability of humans to spot the difference in magnitude between two stimuli. These stimuli could be weight, sound, light, and other elements that can be perceived by human senses. The discovered relation became Weber's Law, which can be formulated as follows:

$$\frac{\Delta I}{I} = K \tag{3.1}$$

Where I is the intensity of one stimulus, ΔI is the extra intensity needed for another stimulus of the same kind to be minimal noticeably different from the first, and K is a constant(Weber, 1978). Constant K is called the Weber Fraction. The Weber Fraction remains relatively constant for each particular sense, but different types of sensory judgements have their own Weber Fraction, as depicted in fig 3.1 (Goldstein, 2009).

Electric Shock	0.01
Lifted Weight	0.02
Sound Intesity	0.04
Light Intensity	0.08
Taste (Salty)	0.08

Table 3.1.: *Weber Fractions for a Number of Different Sensory Dimensions (Teghtsoonian, 1971)*

In this research, Weber's Law could help to set the intensity of different notifications, where a clear distinction between alerts could distinguish different levels of severity.

3.3. Object Detection

Object Detection through Deep Learning has been applied in various fields, as deep learning helps to improve the detection and classification performance of computer vision related challenges, solving issues in areas where approaches based on hand-crafted features provided only limited solutions(Karg & Scharfenberger, 2020). In the medical domain, for instance, convolutional neural networks have been applied with promising results. A research performed by Bejnordi et al. showed that deep learning algorithms managed to outperform the diagnostic performance of a panel of 11 pathologists, mimicking their routine pathology workflow in a simulation exercise(Bejnordi et al., 2017). In the steel industry, the same technology was used to diagnose steel surface defects, with scores surpassing the traditional machine learning approaches. The Convolutional Neural Network (CNN)-based algorithm managed to achieve a detection performance of 99.44%(S. Y. Lee, Tama, Moon, & Lee, 2019).

As there is no literature available on the use of object detection on remotely controlled bridges, it is needed to look into other domains where parallels with the intended use-case can be made. Maybe the closest related topic is autonomous driving, where the detection of vulnerable road users is also a crucial component. With the rising interest in autonomous driving, several kinds of research have been done on the application of convolutional neural networks in the domain of infrastructure and transport. Successful studies have been committed to detecting traffic signs(Peemen, Mesman, & Corporaal, 2011), lanes(Kim & Lee, 2014), and reading license plates(H. Li, Wang, You, & Shen, 2018). These objects are limited to a number of forms and shapes, especially if the model is built for a specific country, so it only has to cope with the given standards of that region. However, traffic users vary to a large extend. There are cars in many different forms, and the appearance of pedestrians is different among ages, ethnic backgrounds, cultures, etc. According to researches done in the last years, convolutional neural networks also outperform the current hand-tailored feature detectors and machine learning models when it comes to these complex elements (Bautista, Dy, Mañalac, Orbe, & Cordel, 2016), (Tomè et al., 2016), (Ribeiro, Nascimento, Bernardino, & Carneiro, 2017).

Apart from the less case-specific attributes that helped the rapid development of deep learning models, being the changes in network architecture, the improved ability to handle a big amount of data, and the availability of faster processing with GPU's e.g, there are also more distinct components that led to the enhancements of deep learning in detecting vulnerable road users. The most important one is the availability of many public datasets focused on this category, as one of the prerequisites for composing an accurate convolutional neural network is having a lot of data to perform training on(Karg & Scharfenberger, 2020).

One of these big public datasets is the Caltech dataset, which was at the time of publication in 2009, twice as big as the existing pedestrian datasets of that time. The Caltech dataset consists of 250,000 frames, subtracted from approximately 10 hours of 640x480 30Hz video taken from a vehicle driving through regular traffic in an urban environment. In these frames, 2300 unique pedestrians were annotated, consisting of 350,000 bounding boxes(Dollár, Wojek, Schiele, & Perona, 2009). This dataset helps both training pedestrian oriented computer vision models, and evaluating the effectiveness of the generated model. However, these datasets are all build for eye-height, so it is to be seen if training on these datasets is also sufficient for the cameras at bridges, which are normally about 6 meters high. This will be addressed in the analysis phase.

Karg et al. also state the functional requirements for a pedestrian detection model, that can be translated into the application of a similar model on remotely controlled bridges. The following needs are distinguished(Karg & Scharfenberger, 2020):

1. Pedestrian detection should work at any lighting condition, for pedestrians both far and near the camera.
2. Pedestrian detection should be able to handle different sizes, poses, appearances, and views.
3. Pedestrian detection should work in challenging weather conditions.
4. Detection should work in complex environment and traffic situations.
5. Pedestrians can be detected even when they are covered by carried objects, vehicles or other persons.
6. Pedestrian Detection is able to detect individual pedestrians in a crowd and extract the most relevant and critical pedestrian that a vehicle may need to brake for.

For the case in this report, the last requirement is less relevant. Any detected person or vehicle is considered a reason to activate the notification system. There should not be a difference whether it is a group that crosses a bridge whilst it is opening, or just one person.

In this report, the goal is to work with the existing infrastructure, so the camera systems as they are being used at the moment. This is because they are technically in order, and this would minimize the costs of scaling the solution. However, the use of CNN's isn't just limited to color images, but can also be used on multispectral data. This way, thermal image information can also be used to boost the performance of the object detection model. Combining these two spectrums has shown possibilities in obtaining higher confidence rates (Ding, Wang, Laganière, Huang, & Fu, 2020; C. Li, Song, Tong, & Tang, 2019; Chen, Xie, & Shin, 2018). Analysing the gains of using multispectral images compared to using only the color spectrum in the specific use case of remotely controlled bridges, can be handled in further research, as stated in the discussion.

This report deals mainly with CNN's, which is a Deep Learning approach, but there is also a machine learning approach that has been popular over the last decades. This is by using a Histogram of Oriented Gradients(HOG) descriptor and a Support Vector Machine. According to research, CNN shows promising improvement over these more traditional methods (Liu et al., 2020; Lemley, Abdul-Wahid, Banik, & Andonie, 2016; Antipov, Berrani, Ruchaud, & Dugelay, 2015). This is also visible in figure 13, where the benchmarks on the popular VOC2012 dataset show a significant increase since the arrival of Deep Learning. Accuracy aside, CNN tends to need a fair amount of processing power, as stated earlier in this section, making it a fairly slow process, depending on the complexity of the model. This research should show whether it is possible to find a good balance between accuracy and speed for the use of Object Detection on the use case of remotely controlled bridges, as this has not been done before.

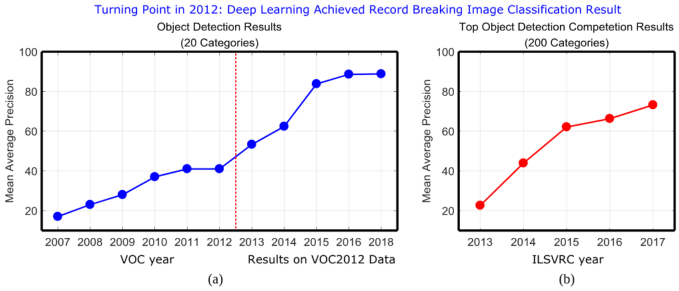


Figure 13.: *Significant improvement since arrival of Deep Learning (Liu et al., 2020)*

3.3.1. Background Subtraction, Histogram of Oriented Gradients and Support Vector Machine

Before researching the deep learning oriented approach, the more traditional, machine learning approach of Background Subtraction, Histogram of Oriented Gradients and Support Vector Machines will be explained.

Background Subtraction

Background subtraction is a concept used to detect moving objects in videos taken from a static camera. Several algorithms have been developed to achieve this, and it is applied to various fields, such as surveillance, analyses of sport videos, etc. Because this technique is applied to static cameras, it is possible to detect moving objects by comparing each new frame with a representation of the scene background. The main advantage of using background subtraction is that the outcome is an accurate segmentation of the foreground regions of the scene background (Elgammal, 2014).

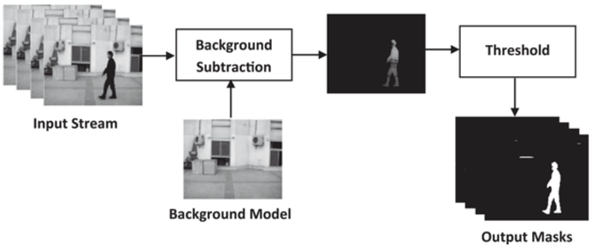


Figure 14.: *Background Subtraction (Shaikh et al., 2014)*

The heart of the technique is background modeling and background adoption to both sudden and gradual changes in the background. With a completely static background, this process isn't too challenging, but when working with natural scenes, backgrounds are generally dynamic. Changes in illumination, swaying vegetation, rippling water, fluttering flags, and such behaviours can be expected. This is the main hurdle for this technique (Jiang & Wang, 2012).

Several researchers have tried to mitigate these problems, and came up with significant improvements by using Gaussian mixture model amongst others, to make the background adaptive to these dynamic elements (Stauffer & Grimson, 2000; L. Wang, Tan, Ning, & Hu, 2003).

Typical outdoor challenges for video analytics are:

- Low-light situations
- Fast illumination changes
- Snow / hail reducing visibility
- Shaking / vibrating camera
- Moving background like grass / trees in the wind or water / waves
- Low contrast of object to background
- Object moving toward the camera instead of crossing the field of view
- Object rolling / crawling towards the fence
- Fast objects near the camera
- Deep object shadows
- Groups of objects

Histogram of Oriented Gradients

Histogram of Oriented Gradients(HOG) descriptors have been shown to be distinctive and robust under small affine transformations and illumination changes. The descriptors are constructed by dividing the image into a dense grid of uniformly spaced cells and then computing the orientation histograms of the image gradient values on each cell.

Local normalization of the gradient strengths takes care of the illumination and contrast differences. The vector of the components of the normalized cell histograms for all the block regions forms the HOG descriptor(Collumeau, Leconge, Emile, & Laurent, 2011).

Support Vector Machines

Support Vector Machines(SVM) are learning systems that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimisation theory that implements a learning bias derived from statistical learning theory(Andrew, 2001). For the classification of the detected objects, support vector machines are a popular choice. Three of the main strengths of SVMs compared to other analysis methods are the fact that SVMs tend to work well in datasets that have a very large number of variables and a relatively small sample size. Secondly, SVMs can learn both simple and highly complex models, and finally the strong built-in protection against an element that is deleterious to modern high-dimensional modeling; overfitting(Statnikov, Aliferis, Hardin, & Guyon, 2013, pg.9). Overfitting means that a model learns the training set too well, but underperforms on unseen data. This can be observed when the loss of the training set is way lower than the loss of a new dataset/the testing set. The other side of the spectrum is underfitting. Here, the losses in the training set are high, and so are the losses when running the model on an unseen dataset. A good fit model is a model that suitably learns the training dataset and generalizes well to the new dataset(Brownlee, 2018a, pg.246). A visual example is shown in figure 57, in appendix B.

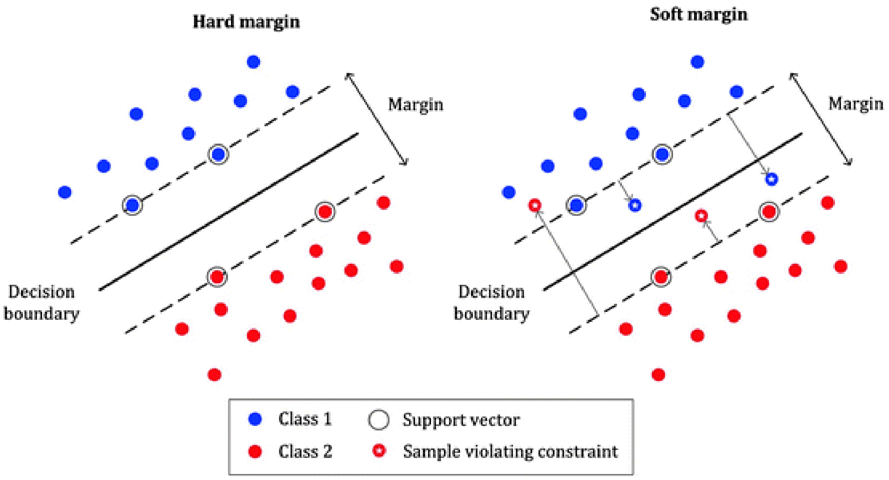


Figure 15.: *Soft Margin & Hard Margin SVM (MLMath.io, 2019)*

In figure 15, the working of an SVM is depicted. In this case, a dataset is classified into two classes, which is closest to the basic form of a SVM, as it originally is a binary classification methodology. The goal of the SVM is to pick a boundary line(or hyperplane) to make the margin between the classes as big as possible. The data points closest to the boundary line, circled in the figure, are the support vectors. The difference between the hard margin(left) and the soft margin(right) is the inclusion of outliers/anomalies. The hard margin forces a hyperplane to separate the classes(with a high order polynomial), often leading to over-fitted models. The use of soft margins allows the SVM to wrongly classify samples that are outliers, keeping a good fit for classifying unseen data samples.

As mentioned earlier, SVMs are capable of dealing with highly complex datasets. This is done through so-called multiclass SVMs, which can be regarded as multiple binary classifications combined. The two methodologies to do so are 'One vs. One' and 'One vs. Rest', as displayed in figure 16. 'One vs. One' SVM is generally the fastest method of training, and tends to have higher accuracy, according to Hsu & Lin(Hsu & Lin, 2002) and Allwein(Allwein, Schapire, & Singer, 2001), although Rifkin & Klautau(Rifkin & Klautau, 2004) disagree with the statement on the accuracy, after proper configuration.

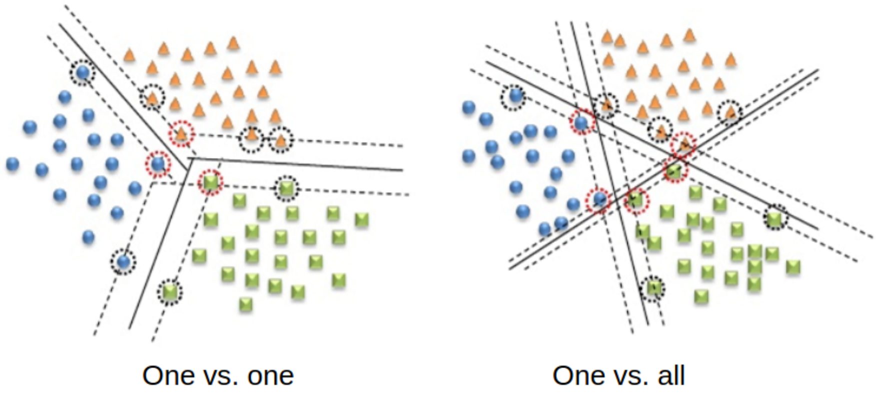


Figure 16.: *Multiclass SVM (Dürr, 2014)*

Elements Combined

Combining the three elements above, a solid object detection model can be constructed. The Background Subtraction is used to find the area of interest, subsequently the HOG-descriptor is used to subtract the features, and finally the Support Vector Machine is used to classify the output of the HOG-descriptor into useful categories. This typical architecture, although without the background-subtraction, is described by (Suleiman & Sze, 2016) and depicted in figure 17. The descriptor is able to process high definition video footage in real-time with a frame-rate of 60fps.

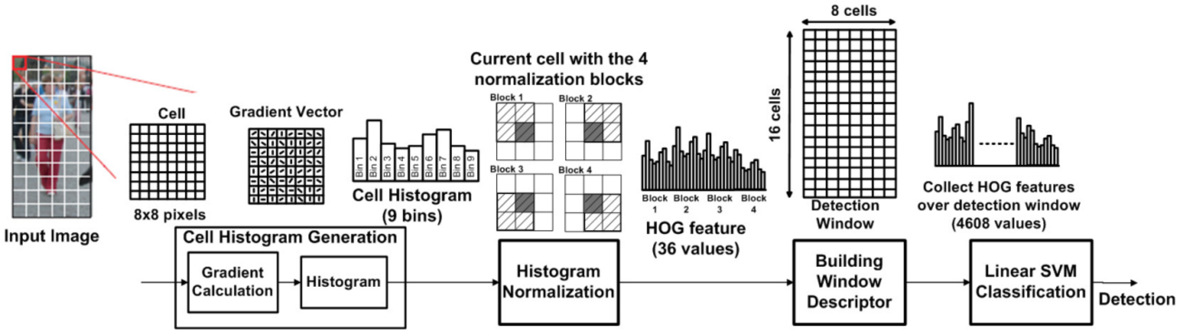


Figure 17.: *Object Detection Algorithm using HOG features and SVM (Suleiman & Sze, 2016)*

3.3.2. Convolutional Neural Networkss

In this section, CNNs will be explained, based on the basic architecture as depicted below. The functionalities of the separate parts will be discussed from left to right, from the input layer, to the output layer.

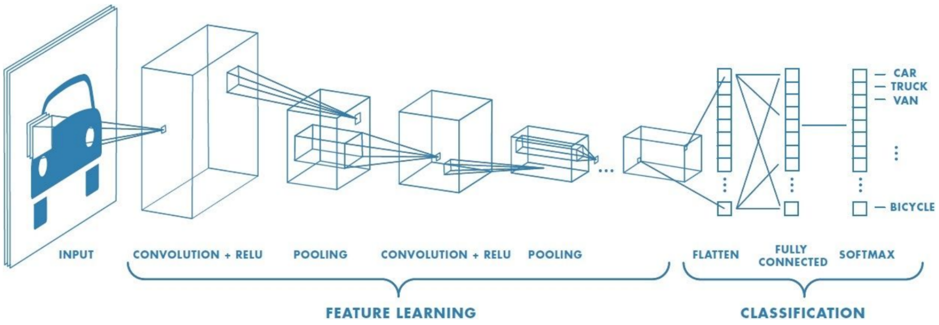


Figure 18.: *Sample Convolutional Neural Network Architecture (Stanford University, 2018)*

Input Layer

The input of a Convolutional Neural Network(CNN), are images. When working with video footage, the goal is to extract frames from the video, and process them individually. Each frame is analyzed from scratch, so the CNN doesn't work with the relations between the individual frames. It is possible to build a model that does keep this relationship in mind. For example, tracking models can be built with a CNN architecture, linking the different frames within a camera stream, or combining them with other cameras observing the same region (Chu et al., 2017; Xu, Li, & Deng, 2016).

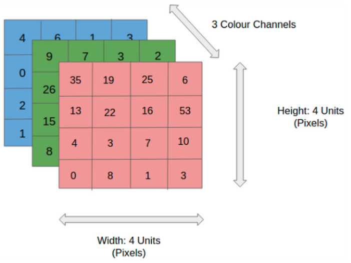


Figure 19.: *Decomposed RGB (The Learning Machine, 2020)*

When dealing with a monochrome input image, the input of the CNN is a matrix with the M(Height of Image) * N(Width of Image), as the matrix contains all pixel values of the image, which are values in the range of 0 to 256. When dealing with a color image, the image can be decomposed into three channels, the red, green, and blue channels(RGB). In this case, the input is a three-dimensional matrix with the number of channels as an extra dimension. In image 20 it is shown how a three-dimensional RGB image can be deconstructed into three two-dimensional images. As yellow is a combination of red and green, it is clear to see how the values in this region are high in the red and green channel, and low in the blue layer, as a high pixel value is depicted light, and a low pixel value dark.

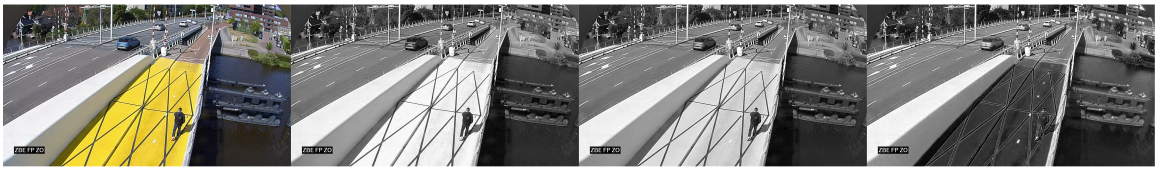


Figure 20.: *RGB(1),R(2),G(3),B(4)*

Troughout the feature learning phase, the channels are separated, and are combined just before the classification process into a 1d array.

Feature Learning

Convolutional Layer/Feature Engineering

The convolutional layer is the most important component of the CNN, and is also known as the feature detector of a CNN. Here, filters are convolved with a given input matrix to generate an output feature map. This input matrix can be the raw data of an imported image, as described in the previous section, but as many CNN architectures contain multiple convolutional layers, it can also be the result of a previous convolution. The goal of these convolutional layers is to subtract patterns or features from the image. In the early convolutional layers, this is often used to subtract global features, as edges, and in later convolutional layers, this is going more into detail, like textures and facial features. An example of such an early convolutional layer is depicted in figure 21, where the upper part of the image shows the filters used to detect the edges.

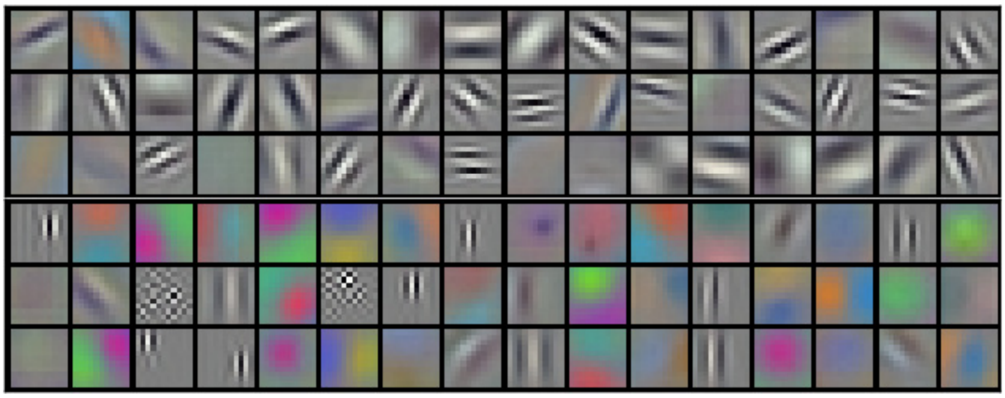


Figure 21.: *Filters of AlexNet's First Layer (Krizhevsky et al., 2012)*

A filter, also known as a kernel, is a rectangular grid of discrete numbers (Khan, Rahmani, Shah, & Bennamoun, 2018). Within the convolutional layer, these are the parameters that are learned. When dealing with 32 filters for example, each of dimension 5x5, the total of learnable parameters is 32*5*5=832 (Michelucci, 2019). There also exist CNNs where an additional bias is added in the convolutional layers, which would add one learnable parameter per filter. These filters, are applied across the width and height of the input matrix in a sliding windows manner with a predefined stride(s), and are applied for every depth of the input image. While doing this, the dot product for each position is stored in a new matrix, as shown in figure 22.

The result of this convolution process is called the feature map, also referred to as the activation map. Both names are logically named after the process. The feature map stores the features as detected by the related filter, and the activation map stores the pixels where the filter was 'activated', which means that the filter lets information pass through it from the input volume into the output volume. As seen in figure 22, the computed feature maps are smaller sized than the input matrix. However, there are applications where this is not desirable, as they require more dense predictions at the pixel level. Moreover, keeping the spatial size constant allows to design deeper networks by avoiding a quick collapse of the output feature dimensions (Patterson & Gibson, 2017).

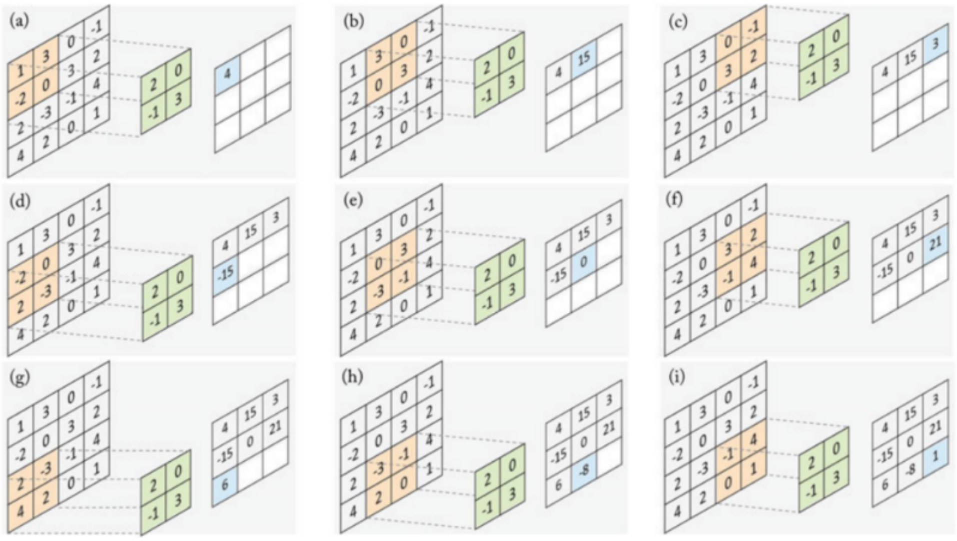


Figure 22.: *Convolution Process Visualized (Khan et al., 2018)*

The solution for keeping the spatial size the same, is adding zero padding to the input matrix. This is done by adding rows of pixels on the top and bottom, and columns of pixels on the right and left of the input image before doing the convolution (Venkatesan, Li, Venkatesan, & Li, 2018). The number of added rows and columns can be calculated by using the feature map dimension equation, as stated in fig 3.2.

$$h' = \lfloor \frac{h - f + s + p}{s} \rfloor, w' = \lfloor \frac{w - f + s + p}{s} \rfloor \tag{3.2}$$

where:

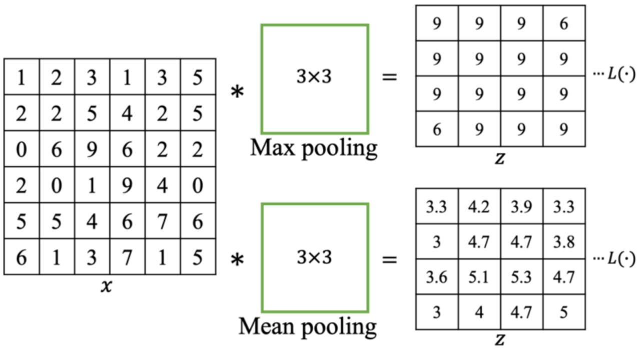
h' = height feature map
w' = width feature map
h = height input matrix
w = width input matrix
f = filter size
s = stride
p = padding

Max/Mean Pooling

A popular manner to reduce the dimensions of the outputs of the convolutional layers, is making use of max pooling. It may feel counter-intuitive to reduce the dimensions after explaining that zero-padding is added to maintain the spatial dimensions, but this way of reducing dimensions will condense the relevant information (Michelucci, 2019). By gradually aggregating information, yielding coarser and coarser maps, the global representation of an instance can be learned, while keeping all of the advantages of convolutional layers at the intermediate layers of processing (Zhang, Lipton, Li, & Smola, 2020).

Another benefit of using pooling layers is that it makes the network more tolerant to slight translations. For example, when working with edges, if an edge is moved up one pixel, and the resulting feature map would not be subjected to the pooling layer, the output of this slightly translated image would be vastly different. As these pixel shifts are unavoidable in real life, because of camera vibrations, for example, mitigating the sensitivity of the convolutional layers by the pooling layer is benifitial (Zhang et al., 2020).

The two most common ways of pooling are Max and Mean Pooling, where the first mentioned is the most popular choice. However, it is dependent on the data and features at hand whether max pooling is the most accurate option (Boureau, Ponce, & Lecun, 2010).



1	2	3	1	3	5
2	2	5	4	2	5
0	6	9	6	2	2
2	0	1	9	4	0
5	5	4	6	7	6
6	1	3	7	1	5

9	9	9	6
9	9	9	9
9	9	9	9
6	9	9	9

3.3	4.2	3.9	3.3
3	4.7	4.7	3.8
3.6	5.1	5.3	4.7
3	4	4.7	5

Figure 23.: *Max Pooling & Mean Pooling Visualized (A. Wang, 2019)*

Classification

After the pooling layer, a fully connected neural network(fig 24) is used to subsequently classify the extracted features. The output of the last pooling layer before the classification is a 3D feature map, and the input required for a fully connected neural network is a 1D feature vector. So, before the output of the feature learning phase can be processed in the classification phase, the 3D volume is flattened, into an 1D vector.

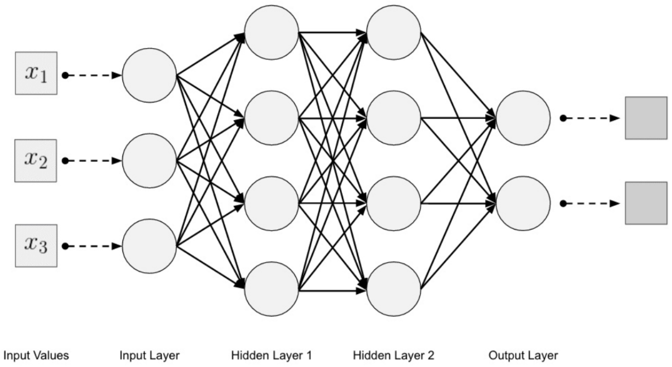


Figure 24.: *Fully connected multilayer feed-forward neural network topology (Patterson & Gibson, 2017)*

The fully connected neural network has an input layer, one or more hidden layers, and an output layer. Each layer has one or more artificial neurons. The architecture of a neuron is displayed in figure 25. The net input of the activation function is the dot product of the weights and the input features. The depends on the type of activation function what the output of this function will be. The different activation functions are described in more detail in the next section.

The input features in this dot product can be the input value of the flattened layer, when the neuron is part of the first hidden layer. It can also be the output of an activation function in the layer before, if the neuron is not in the first hidden layer. The other part of the dot product is the weights on the connections. These are coefficients that scale the input signal to a given neuron. This can either amplify or minimize the input features.

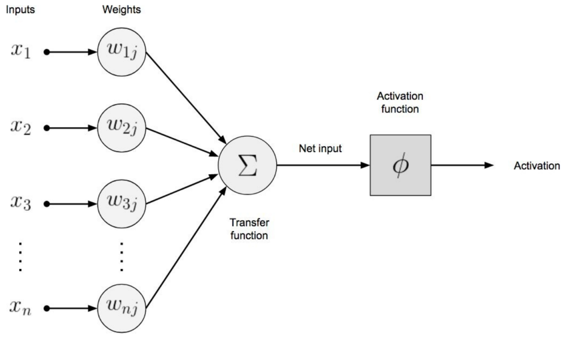


Figure 25.: *Artificial neuron for a multilayer perceptron (Patterson & Gibson, 2017)*

Often, biases are added to the input on an activation function. These are scalar values that ensure that at least a few nodes per layer are activated regardless of the strength of the input signal. Biases enable the network to learn by giving action in the event of low signal. This way it allows the network to try new interpretations or behaviors (Patterson & Gibson, 2017).

Combining these elements, the activation value of a node can be noted as follows:

$$a_i = g(W_i \cdot A_i + b) \tag{3.3}$$

where:

a_i = Activation value
W_i = Vector of all weights leading into neuron i
A_i = Vector of activation values for the inputs to neuron i
b = Bias value
g = Activation function

Before going into detail about the activation functions, the purpose of the output layer should be discussed. The neural network maps an input space to an output space, so the output layer gives an output based on the input from the input layer. Therefore, the output layer provides the predictions or answers of our model, depending on the goal of the model, based on the provided input data. In case of a classification problem, the output of the model is a vector with predictions, with every node in the output layer representing a separate category. In case of a regression, the outcome would have been a real-valued output. To get to these predictions, the output layer uses either a softmax or sigmoid activation function for returning the final values. The difference between both will be discussed in the next section.

Activation Functions

The activation functions make it possible to deal with more complex problems, as it enables to model not to work as a linear model. If these nonlinear blocks wouldn't be incorporated in the architecture, it wouldn't matter how many layers the model would have, as any linear combination of linear functions collapses down to be a linear function. As the name suggests, the activation function controls the way a perceptron is activated. The way of doing this depends on the type of activation function that is used. It can set a threshold deciding from which value the neuron fires or not (0 or 1), or it uses a function deciding to what extend the neuron is firing.

Sigmoid $\sigma(z) = \frac{1}{1+e^{-z}}$

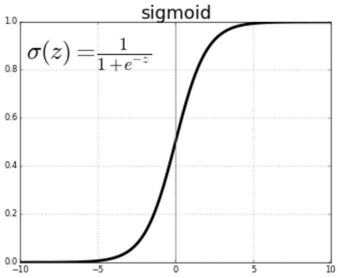


Figure 26.: *Sigmoid Function (Sharma, 2019)*

Historically, the Sigmoid and Tanh functions were the activation functions of choice for most neural networks. However, there has been a shift to another activation function. This was done because of difficulties during the training of the neural networks. The problem with both the Sigmoid and Tanh function is called the vanishing gradients problem.

Because training is done by backpropagation, making use of local gradients, there is a problem when a neuron saturates close to either zero or one, as the gradient in this region is very close to zero. This becomes increasingly problematic when dealing with multiple layers. To understand this phenomenon, it helps to look at SGD(Stochastic Gradient Descent). In short, SGD looks at the partial derivative of a weight parameter with respect to the total error, and multiplies this local gradient with a learning rate, and subtracts this from the initial weight. Doing this in an iterative manner, with all trainable parameters in the network, the total loss will be minimized to a minimum.

$$\omega^i \leftarrow \omega^i - \eta \frac{\partial Etotal}{\partial \omega^i} \tag{3.4}$$

where:

ω^i = weight
η = learning rate
$Etotal$ = Error

These local gradients are calculated by making use of the chain rule. The more layers a neural network has, the more attributes the equation will get. Using multiple Sigmoid activation functions will add multiple values between 0 and 0.25 to the equation, as those are the lower and upper limits of the sigmoid's derivative, with high risks of values very close to zero. Multiplying all these small values may lead to a near negligible local gradient, and looking at the equation in figure 3.4, will hardly update the weight. Especially when taking into account that the learning rate is often a value somewhere between 1 and 0.00001 (Patterson & Gibson, 2017). Not being able to adjust the weight with large enough steps to adjust the total error, equals not being able to fully train the network.

To resolve this problem, ReLU(Rectified Linier Unit) was introduced in the object detection field. It was first introduced in the field of biology by Hahnioser, in 2000 (Hahnioser, Sarpeshkar, Mahowald, Douglas, & Seung, 2000), but was popularized in the field of object detection by Nair in 2010(Nair & Hinton, 2010), and is now the most popular type of activation functions, because of the much steeper profile. Faster and better learn-ability was achieved, demanding less computational power.

$R(z) = max(0, z)$

However, the basic variant of ReLU function isn't flawless either. When the activation value of the ReLU neuron becomes 0, then the gradients of the neuron will also be 0 during backpropagation, making the neuron unusable. This is referred to as the Dying ReLu Problem. This can be mitigated by assigning the right values for the initial weights and learning rates (Campesato, 2020), or by using variants that have been developed based on the ReLU activation function, like Leaky ReLU or ELU, which solve the problem of the activation value becoming 0, by dealing with the negative values in a different way 27.

When dealing with classification, as it is done in this report, there is one last activation function that needs to be addressed, and this is the Softmax activation function. As discussed in the previous section, when dealing with classification, either a softmax or a sigmoid function is used in the output layer. When dealing with binary classification,

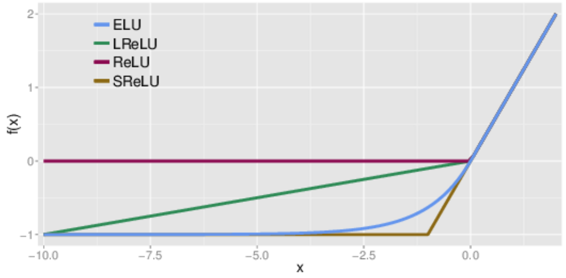


Figure 27.: *ReLU Variations (Clevert et al., 2016)*

often sigmoid is used to get the final probability, but softmax is also usable. However, in case of a multiclass classification problem, the softmax activation function is used to interpret the outputs as probabilities, by making them nonnegative and sum up to 1 (Zhang et al., 2020). This fact that with softmax the outputs are interrelated, makes it the activation function of choice.

$softmax(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j))}$

Loss Functions

As explained before, a well-trained CNN has weights that amplify the signal, and dampens the noise. The bigger the weight, the stronger the correlation of the signal to the outcome. During the training of an CNN, the weights and biases are re-adjusted. This process allocates the significance of the elements within the network, helping the model learn which features are tied to which outcomes, and changing the trainable parameters to reduce the loss in the output layer (Patterson & Gibson, 2017).

To be able to minimize the loss in the output layer, we need to understand how to calculate the loss of the network. Several loss functions can be used to reach this goal, but the shared goal of these functions is to quantify the difference between the desired output, and the predicted output by the model. For multiclass classification, this is often the multiclass cross-entropy function, also referred to as logarithmic loss. Cross-entropy will calculate a score that summarizes the average difference between the actual and predicted probability distributions for all classes in the problem (Brownlee, 2018b).

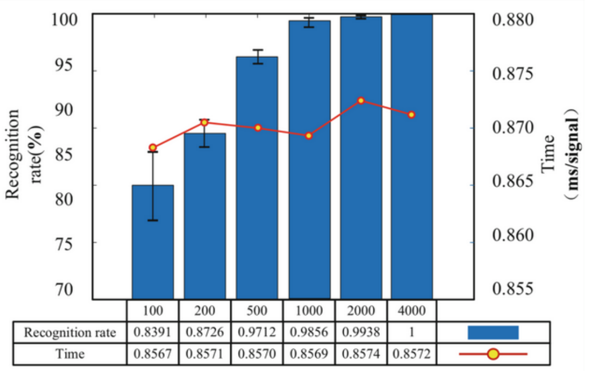
$$H(Q,P) = -\sum_{i=1}^{n} P(x_i)\log Q(x_1) \tag{3.5}$$

where:

P = Model Prediction
Q = Ground Truth
n = Number of classes
H = Cross-entropy

Training/Test set

For training and testing the network, an annotated dataset is needed. Hereby, the model can train itself by comparing its predictions, with the ground truth as annotated in the training set. In order to achieve a model with good accuracy and generalization, it is important to have a big dataset to train on (Hestness et al., 2017; Sun et al., 2020). Although the training time will increase significantly, running the model will only take marginally longer.



	100	200	500	1000	2000	4000	
Recognition rate	0.8391	0.8726	0.9712	0.9856	0.9938	1	🟦
Time	0.8567	0.8571	0.8570	0.8569	0.8574	0.8572	—○—

Figure 28.: *Number of training samples, identification rate and diagnoses time. (Sun et al., 2020)*

Constructing a big dataset is time consuming, and not always feasible when the available data is scarce. A way to deal with this is to make use of data augmentation. Data augmention is one of the most popular approaches to reduce the risk of overfitting by artificially creating training samples to increase the size of the dataset (Yoo et al., 2016). Traditional data augmentation techniques for image classification tasks are for example, flipping, distorting,adding a small amount of noise to, or cropping a patch from an original image (Inoue, 2018).

4. Methodology

4.1. Theoretical Framework

The goal of the theoretical framework is to get a good understanding of the different fields of interest in this project, to help formulating the requirements of the proof-of-concept. To do so, literature study should be conducted on the related topics. The following three subsections can be distinguished.

4.1.1. Bridge Operations

The bridge operations part of the theoretical framework should cover the context the support systems needs to work in. By going over the principles, the physical workplace, and the individual steps of the bridge opening procedure, the challenges and opportunities of the current situation are described. These will be complemented by user interviews and observations in a later stage of the research.

4.1.2. Human Machine Interaction

This is the more social side of the research, the research on Human Machine Interaction(HMI). Being a decision support tool instead of a fully automated application brings up challenges on how the interaction between the computer model and bridge operator should work. For example, if the proof-of-concept gives too many false positives or false negatives, the bridge operator could stop paying attention to the system's outcome. The functionality, position in the operational flow chart, and way of communicating the object detection signals will all have a certain effect on the operator. It also works the other way around. The capabilities of the operator of working with a proof-of-concept will contribute to the application's effectiveness. This subject matter has been around for decades. Although human-machine interaction studies specifically on bridge operations are limited, comparisons with other work fields will be drawn.

4.1.3. Object Detection

Object Detection will also mostly consist of material obtained from the literature. The first phase will be about artificial intelligence in general, and then the research will dive deeper into this area, going from machine learning to deep learning, to end up focusing on object detection. These steps are taken to set the context of this new technology right, instead of diving into it without an understanding of where it's coming from. In the Object Detection section the rise of the technology will be explained, the benefits and downsides, and most importantly; the way it works. Often Machine Learning is depicted as a black box, but analysing the processes in a neural network, and working with simple examples should provide enough clarity and insight to demystify the concept. Currently, Convolutional Neural Networks are considered to be the ideal solution for tackling object detection problems, so this will be the method that will be analysed. A typical sample of a CNN is depicted in figure 31. In the research the different layers of the convolutional neural networks will be explained, and so will the way a network trains itself through back-propagation.

4.2. Underlying Research Methodologies

4.2.1. Action Design Research

For this research the methodology of choice is the Action Design Research methodology(ADR). This methodology is invented for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting, where traditional design science does not fully recognise the role of organizational context in shaping the design as well as shaping the deployed artifact(Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011). In the methodology, four stages and seven principles can be distinguished, as seen in figure 29. The individual elements of the ADR process are explained in more details in appendice A, section A.1.

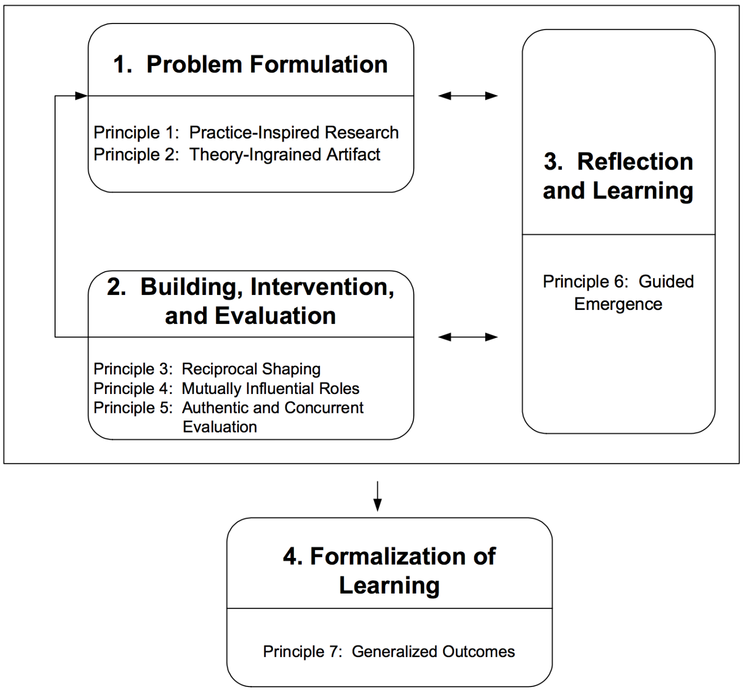

1. Problem Formulation
Principle 1: Practice-Inspired Research
Principle 2: Theory-Ingrained Artifact
2. Building, Intervention, and Evaluation
Principle 3: Reciprocal Shaping
Principle 4: Mutually Influential Roles
Principle 5: Authentic and Concurrent Evaluation
3. Reflection and Learning
Principle 6: Guided Emergence
4. Formalization of Learning
Principle 7: Generalized Outcomes

Figure 29.: *ADR Method: Stages and Principles (Sein et al., 2011)*

4.2.2. Goal-Directed Design Research

The second underlying research methodology is goal-Directed design research. Goal-Directed Design combines techniques of stakeholder interviews, market research, detailed user models, scenario-based design, and a core set of interaction principles and patterns. It provides solutions that meet the needs and goals of users, while also addressing business/organizational and technical imperatives (Cooper et al., 2007). The global steps are depicted in figure 30. By combining the goal of addressing the problems of users, while keeping organizational and technical facets in mind, shows similarities with ADR. In the next section will be explained why the combination of the two seemed a sane option to make. In appendix A, section A.2, the individual steps of Goal-Directed Design Research are explained in detail.

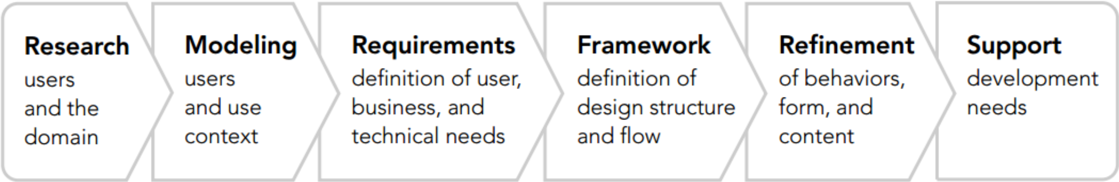

Research
users and the domain
Modeling
users and use context
Requirements
definition of user, business, and technical needs
Framework
definition of design structure and flow
Refinement
of behaviors, form, and content
Support
development needs

Figure 30.: *Steps Goal-Directed Design Research (Cooper et al., 2007)*

4.3. Methodologies Combined

As described in the previous section, the research methodology for this research is a combination of ADR and Goal-Directed Research. The ADR lacked a stepwise design approach, but the principles and generalization facets make it a solid foundation for developing a solution that works well in organizational context. To complement the methodology with a stepwise design-approach, goal-Directed design research is used. Together, the general research workflow in fig 32 was developed. In figure 53, the phases are divided into more detail. In essence, the ADR forms the methodology on macro level, and goal-directed design research elements are added to form the methodology on micro level.

4.4. Deliverables

The Proof-Of-Concept is the main deliverable of the project. This should show whether the stated application could be a feasible solution for now, or in the near future. The goal is to make a proof-of-concept that can analyse camera footage of a remotely controlled bridge with sufficient accuracy and speed. What 'sufficient' means in this context, is to be decided in the analysis. The proof-of-concept will be applied to the actual footage of the CCTV system present on the bridges. Apart from this proof-of-concept, the literature study and user interviews should lead to a conceptual framework.

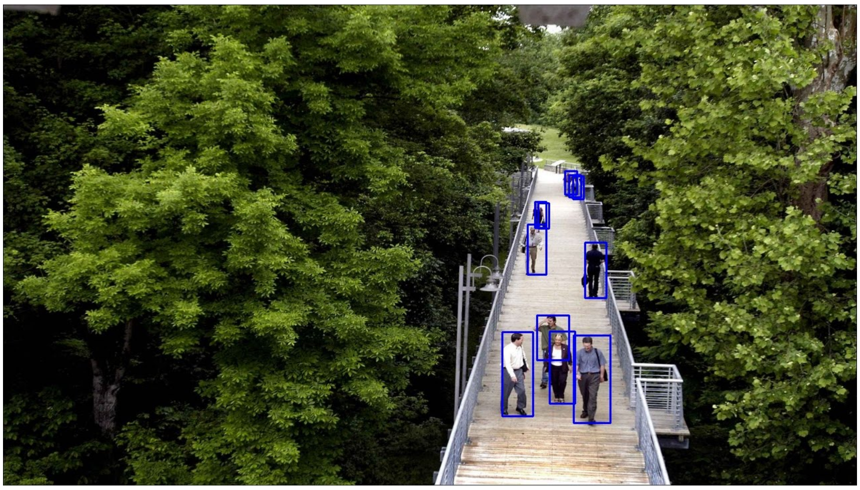


Figure 31.: *Human detection test on still image using Mask R-CNN (Todd, 2016)*

The proof-of-concept will be built in Python, using open-source pre-trained models and libraries. Probably this will be a combination of OpenCV(Open Source Computer Vision Library)(Bradski, 2000) and the Tensorflow detection model zoo(Tensorflow, 2019). In Figure 31, an early test is depicted using this combination on a still image, showing promising results. For this proof-of-concept, both a pre-trained model and a custom trained model will be used, to research the differences between a generalized model, and a model specifically built for this application.

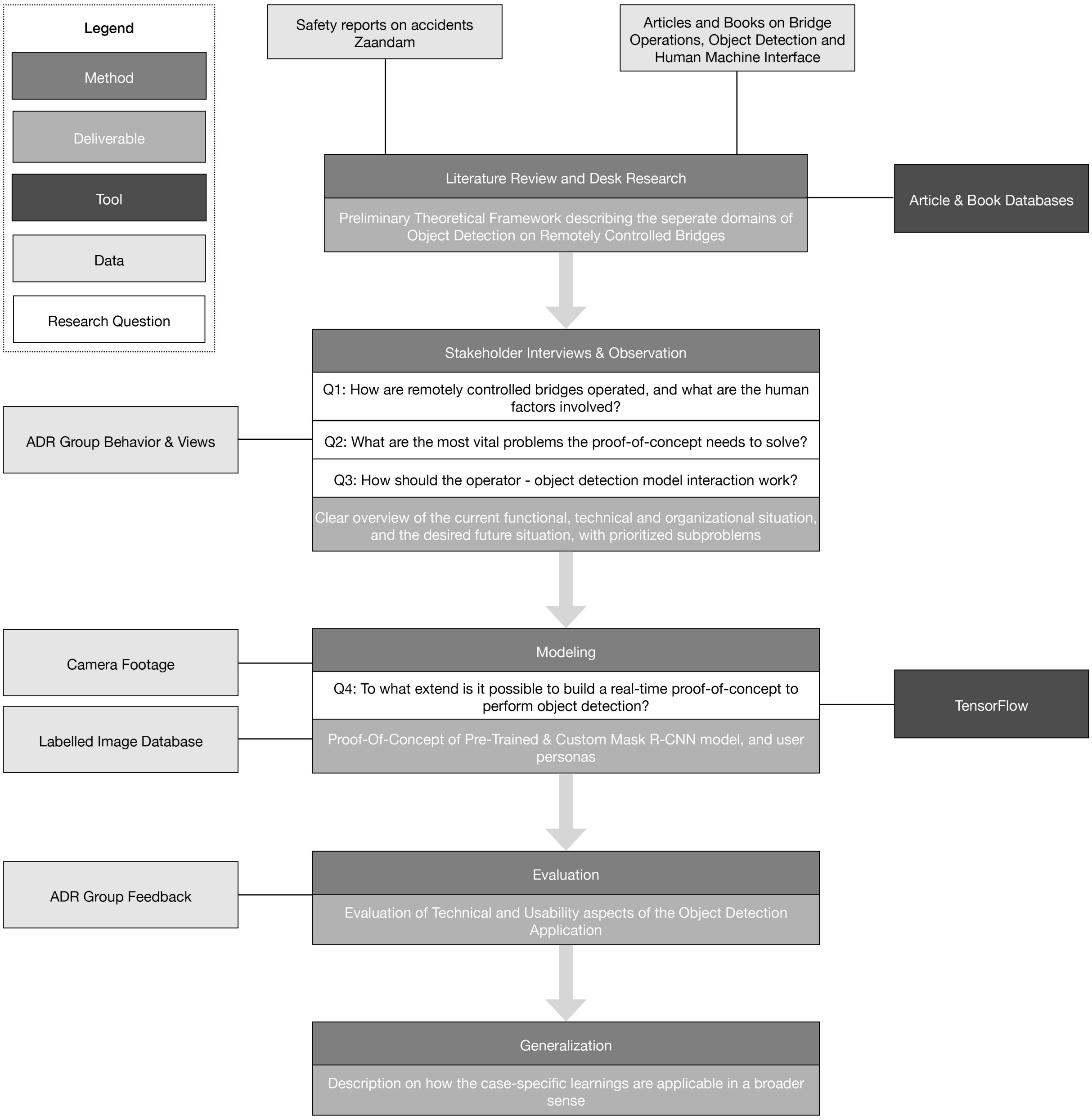

Legend
Method
Deliverable
Tool
Data
Research Question
Safety reports on accidents Zaandam
Articles and Books on Bridge Operations, Object Detection and Human Machine Interface
Literature Review and Desk Research
Preliminary Theoretical Framework describing the seperate domains of Object Detection on Remotely Controlled Bridges
Article & Book Databases
Stakeholder Interviews & Observation
ADR Group Behavior & Views
Q1: How are remotely controlled bridges operated, and what are the human factors involved?
Q2: What are the most vital problems the proof-of-concept needs to solve?
Q3: How should the operator - object detection model interaction work?
Clear overview of the current functional, technical and organizational situation, and the desired future situation, with prioritized subproblems
Camera Footage
Labelled Image Database
Modeling
Q4: To what extend is it possible to build a real-time proof-of-concept to perform object detection?
Proof-Of-Concept of Pre-Trained & Custom Mask R-CNN model, and user personas
TensorFlow
ADR Group Feedback
Evaluation
Evaluation of Technical and Usability aspects of the Object Detection Application
Generalization
Description on how the case-specific learnings are applicable in a broader sense

Figure 32.: *Research Flow Diagram*

5. Analysis & Design

5.1. Stakeholder Interviews & Observation

Gathering information from the reports on the accidents put together by the Dutch Safety Board is one thing, but validation is required to determine whether the identified problems are actually the most significant ones, and if the list of issues is complete.

To examine the behaviour, workflow, and views of the operators, first, a day of observation was conducted, to get a general look and feel of the working environment. By monitoring the procedures and talking to the operators, an initial list of issues was constructed, that later on was completed by problems proposed in the literature. These observations formed an important foundation for the survey.

However, this survey was not the first survey investigating the working conditions and challenges for bridge operators. In 2017, a national survey was conducted by VHP human performance to gain insight into the experiences of bridge and sluice operators, and to see where improvements could be achieved. The Survey was conducted between the 1st of November, and the 31st of December, and counted 175 respondents (van Veelen, 2018). Because of the limited number of respondents in the municipality, this survey is considered a valuable source for qualitative research. An important note is that this survey isn't only aimed at operators of remotely controlled bridges, but also for operators with direct sight. Only 22% operated solely on cameras. 13% operated only on direct sight, and 65% on the combination of both. The key findings that are relevant for this research, are described in the appendix, chapter C.

In the survey for the Municipality of Zaanstad, the current situation was investigated, and a possible future situation where an object detection system was applied. For the latter, the survey started off in a broader sense, where the general attitude toward new technologies and Object Detection in specific were questioned. Subsequently, the questions went more into detail on how the operators envisioned working with such a system. The following observations could be drawn when analysing the survey results.

Current Situation

Bridge operators are generally proud of their job and motivated to go to work. Here, the work environment is comfortable, the collaboration with the colleagues pleasant, and it is clear how to use the systems that are at their disposal.

However, for this research, the most interesting part is where the improvements could be made, and what to look out for when proposing a new system. The operators emphasized the importance of their experience and skills, and requested more attention for their suggestions, and their views on changes in the workplace. Related to these experiences and skills, the operators like to see more investments and appreciation when it comes to craftsmanship, both in the operators themselves, and the tenders. The last place for development mentioned concerning the operator's position in the organization, is the feeling that operators are blamed too quickly for incidents, when a technical problem can't be found. One of the respondents even mentioned that there is a certain fear to report incidents, because of the chance of backfiring to the operator. Discussing these situations and sharing best practices within the organization, could help to work on this problem.

When it comes to the operation procedure, there are some noteworthy remarks also. Mist, darkness, and blinding sunlight or artificial light can lead to unsafe situations. According to the operators of the municipality of Zaanstad, this is strengthened by rainy conditions, where raindrops stick to the camera lens, blurring the operator's view. Dust and insects on the lens are also considered annoying. This, combined with the overall image quality, are the most technical challenges found in the surveys.

The next step is to look at the human factors, an area that, according to the Dutch Safety Board, isn't covered sufficiently when operating remotely controlled bridges (Onderzoeksraad voor de Veiligheid, 2016). One of the statements made is that maritime traffic plays a guiding role in operating the bridge, and the operators feel pressured by the shippers. The national survey points out operators do have to deal with angry, aggressive shippers. However, according to the survey conducted in Zaanstad, this is less obvious. While talking to the employers in Zaanstad however, they did mention the awareness of most shippers being entrepreneurs, and that disruption of their planning, could affect the shipper's result, but also said that the relationship with most shippers is good, which partially comes from the fact that most shippers are frequent visitors, as they work in the area, so the operators and shippers are somewhat familiar with each other. Altogether, it is difficult to tell if the maritime traffic is actually the dominant actor in the bridge operation.

Other human factors concern dealing with the complex task of monitoring a busy situation. When looking at the national survey, the respondents somewhat agree with the statement of having sufficient visibility to operate safely, and are only a little on the disagreeing side of neutral when it comes to the statement that it is hard to monitor the situation accurately, when it is busy. Note, that in this survey, only 22% of the respondents operated bridges solely on camera screens. When looking at the survey at the municipality of Zaanstad, it is noticeable that operators experience difficulties monitoring the many camera streams. This is both the number of screens to focus on, and understanding the camera plans related to them. Also, operators deal with concentration problems. This can be caused by the many stimuli, but as respondents commented in the national survey, the change of day and night shifts make it more difficult to focus.

Future Situation

For implementing a detection system as described in this report, it is vital to know what the concerns and expectations are of the targetted future users, the operators. According to the national survey, the majority of operators is positive about the application of new technologies, and this was confirmed by the results of the survey of Zaanstad. According to the operators of Zaanstad, a good functioning detection system will increase the safety of operating remotely controlled bridges, where the current situation is already considered safe.

Because of the ADR-principle of user feedback from an early stage, and the request from the operator's of using their experience and being consulted when making changes at the workplace, the survey asked their opinions about the way the detection system should fit in the current situation. This was done both in a conceptual way, by asking what the operators though of the permissions and notifications, and a more detailed level, in the way of presenting the detections. These elements will be discussed more in detail in the next sections.

5.1.1. Identified Problems

After the last three steps, a table with prioritized sub-problems can be made. This table should be validated with the users and stakeholders to make sure there is a common understanding of the problems and the related priorities.

Problem	**Source**	**Priority**
No Uniformity SCADA Systems	Literature	Mid
Raindrops on Camera Lens	Interview	High
Dirt on Camera Lens	Observation	Mid
Blinding Headlights	Interview	High
Poor Image Quality	Observation	High
Angry Shippers	Literature	Low
Traffic Users Breaking Rules	Interview	Mid
Concentration Problems	Literature	High
Camera Overkill	Literature	High
Confusing Bridge Designs	Interview	Mid
Confusing Camera Angles	Literature	Mid

No Uniformity SCADA Systems

A bridge operator needs to operate at least three bridges during his shift, and needs to be able to operate all fifteen of them. Although all Bridges share more or less the same procedures, the way of controlling them differs. For some Bridges, Bernhardbrug for example, the operator needs to control each barrier individually. This was implemented after the accident, to improve the operator's focus. For most other bridges, the procedure of stopping the road traffic, is brought together into one push of a button. One could argue that having this automated, could provide the operator with more time to judge the situation and is less error-prone. However, having both approaches present distributed over several bridges, prevents the operator from working with a standardized workflow that can be used for each asset. According to the observed operators, they were not bothered by it, and had no preferred option. Literature cites that consistency in terms of visual appearance should not have to be an obstacle in operating the system, but the action language syntax is (Satzinger & Olfman, 1998).

Dirt & Moisture

Dirt and moisture on the camera lenses make it difficult to get a clear picture of the situation. As seen in figure 33, the formation of raindrops on the lens could distort the image severely. Also dirt, insects, and spiderwebs could block the view. There is a cleaning service that removes these elements periodically, and when it is not possible to wait for that, the operator could send out a request. The cameras can be cleaned by lowering them using the hinge that enables the top part to come down.

For dust, this approach is adequate, but with raindrops, it is not, as the problem is recurring whenever it rains and it is windy. According to the operators, there one has been a trial with water repellent coating. In practice, this coating only degraded the image quality, by attracting dust after a short period of usage.

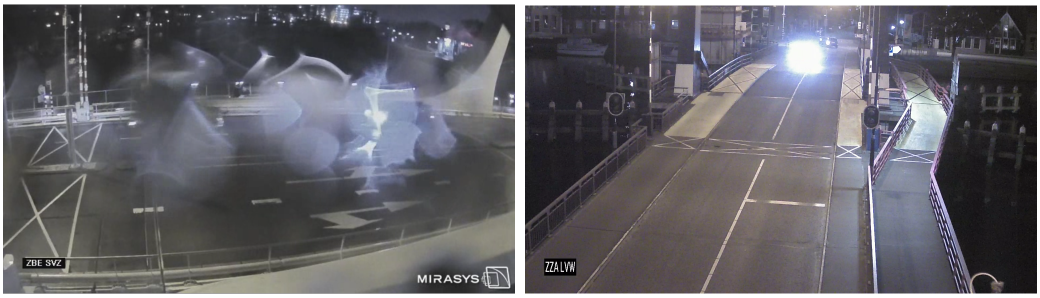


Figure 33.: *External Factors Affecting Image Quality*

Blinding Headlights

The combination of camera placement and the slope of the bridge, can lead to the problem of headlights shining straight into the cameras, blinding the operator, as seen in figure 33. Especially combined with the earlier mentioned raindrops, the image quality degrades quickly. Some bridges have the main cameras positioned at the long sides of the bridge, instead of the short side. This reduces the problem significantly. However, not all operators find this a pleasant angle when it comes to an overview of the entire situation, as it is harder to capture the entire bridge in one frame. A good alternative would be setting the cameras higher, but this is difficult because of regulations.

Poor image quality

Besides the image quality affected by external elements, as described in the last couple of problems, the standard image quality of cameras on particular bridges can be challenging on itself. These differences in resolution and low light quality, make some bridges harder to monitor than others. According to the operators, this doesn't always have to do with the cameras themselves. In some occasions, the connection between the asset and the central post is not up to the desired level, degrading the camera's image quality.

Angry Shippers

As highlighted in the interview section, the Dutch Safety Board and the national survey indicated the stressing event of dealing with angry, aggressive shippers, as commercial shippers have financial interests in a smooth throughput. Luckily, in Zaanstad, this is less of a problem, as most shippers are regular visitors, with a good relation with the bridge operators.

Traffic users breaking the rules

Frequently, road traffic negates red lights and lowering barriers. In figure 34, the total of incidents in week 4 of January 2019 are displayed. In this week, the Alexanderbrug opened 151 times, the Bernhardbrug 130 times, and the Zaanbrug 121 times. On average, this is about one red light negation by pedestrian/cyclist per two openings, and about one in ten openings when it comes to lowering/closed barrier negation by the same user group. The interviews showed that, although this happens frequently, the operators are not unanimously concerned by this. Possibly the stimulus perceived by the operator when detecting someone between closed barriers is decreased by the frequent negation, making it a significant problem, but that can not be proved by the current research. The types of users breaking the rules differ from people who deliberately ignoring the warning signs to make it to the other side of the water in time, to users, mostly tourists, who are not familiar with movable bridges, and therefore misunderstanding the situation.

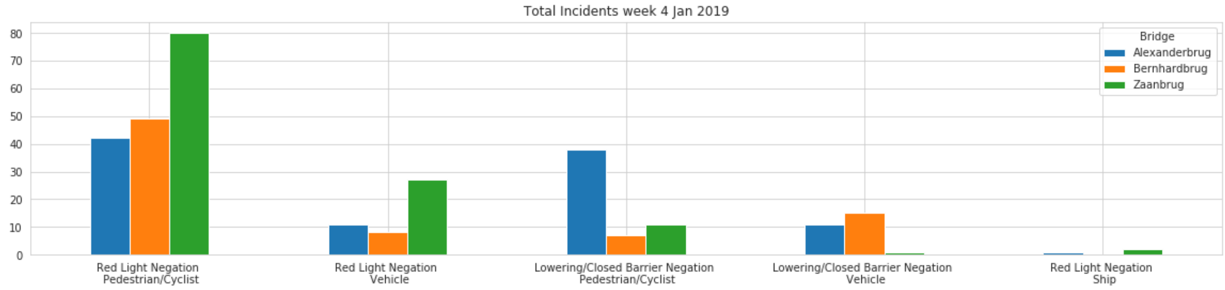


Figure 34.: *Total number of Incidents in week 4 Jan 2019*

Concentration Problems

A report published by vhp human performance on an incident that occurred on the Bosrandbrug, stated the difficulty and significance of remaining concentrated when operating a remotely controlled bridge (human performance, 2017). Although the municipality of Zaanstad has nothing to do with this bridge, the operations are the same, and so are the dangers. The interviewed operators of Zaanstad were divided when it comes to concentration problems, and the respondents of the national survey noted that changing day and night shifts cause concentration problems. So, it is difficult to say how frequent concentration difficulties occur, but when they do, they can pose an immediate threat to the safety on the bridge.

Camera Overkill

This is probably the remark that most operators mentioned during the observation. From a technical perspective, more cameras mean more information, which could lead to better judgement. From a human factor perspective, it works a bit differently. Every screen added to the operator's setup could be an added distraction, when not carefully chosen. Besides only being able to watch one screen at a time, it could also lead to confusion when it comes to linking the separate images, and getting a clear picture of the overall situation. This makes the mental model difficult to compile. In practice, each operator has his own preferred camera standpoints, instead of distributing their attention to all screens equally.

Confusing Bridge Designs

This is closely linked with the problem perceived with the camera overkill. Because of varying bridge designs, going from traditional structures to more organically shaped ones, it requires time to understand how the different cameras line up, also because sometimes additional cameras are needed to get everything in the frame, and what the situation on the bridge is like. For example, the free-formed structure of the Bernhardbrug blocks the view of particular cameras, making it necessary to place extra equipment. Also, because of the design of the bridge, it is challenging to spot approaching pedestrians, because they're coming from underneath the bridge, and enter the bridge right in front of the barriers.

Confusing Camera Angles

The problem with confusing camera angles, is that the traffic flow isn't logically depicted when displayed to the operator. The traffic can exit an camera angle on one side of the frame, to go in the opposite direction in another frame. This makes it difficult to form an accurate mental model of the situation at the bridge.

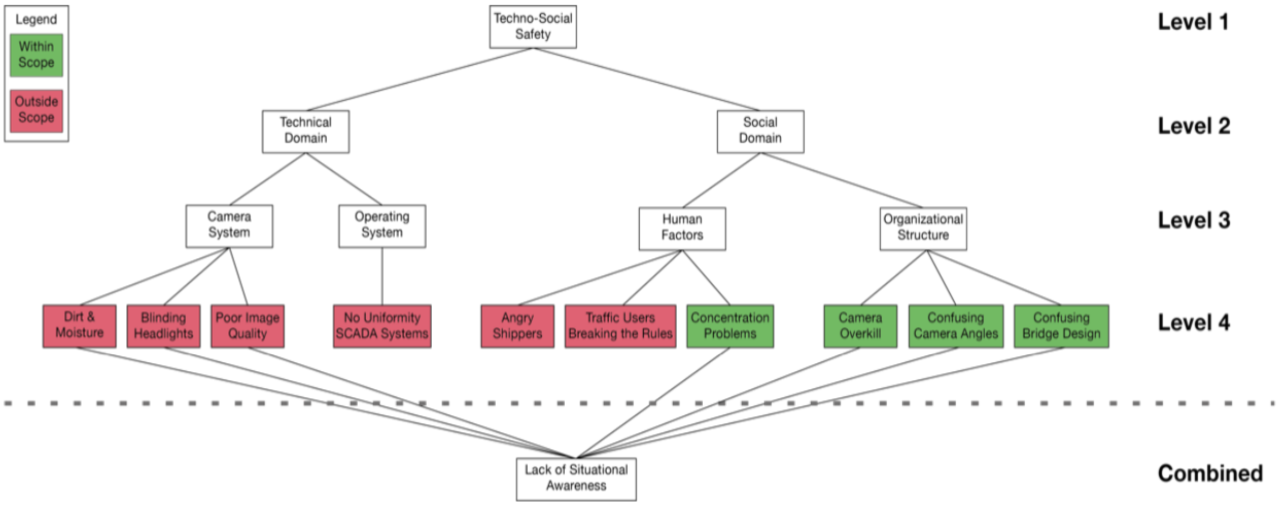


Figure 35.: *Problem Analysis Tree*

5.1.2. Scope

Certain problems will be addressed in this research, and others will be mentioned in the discussion and further research recommendations, as solving these issues, will cause the object detection solution to yield better results. This is mainly the case for the challenges that can be linked to the image quality of the incoming camera footage. The rubbish in, rubbish out principle, a concept that is well known in computer science, also plays its part in computer vision. If the images that form the input for the computer model are too distorted by the moisture, dirt and other factors, the model's output will be less accurate. This is why it is important to research the possibilities to optimize image quality.

The problems that this research tries to solve or mitigate, are the challenges related to the lack of situational awareness, caused by human limitations. As seen in figure 35.

The camera overkill problem, focussing on the number of simultaneous camera streams the operator has to monitor, and the concentration problems, are mainly focussed at the first step of the Endsley's situational model, the perception of elements in the situation . The multiple screens can cause the operator to watch one, when his attention on another screen is required. Also, by constantly switching between screens, the chances of missing motion signals will increase, hereby increasing the chance of change blindness, possibly leading to missing a fully-visible person on the bridge deck. Concentration problems also lead to difficulties in perceiving elements, as the operator's ability to focus his attention on the situation on the bridge will decrease.

The two other problems highlighted in the problem tree, are the issues with the confusing camera placement and bridge design. This has to do with the second step of Endsley's situational awareness model, the cognitive stage. The operator may perceive elements on the camera streams, but because of illogical presentation of camera footages, have difficulties to get a mental picture of how these elements are translated to the placement on the bridge. Local knowledge of the asset's design and experience in mentally connecting these images, may help to decide wether a perceived element is in a safe place or not. If the region of interest for the detection system is well-positioned during the installation of the application, the coverage of the desired area is guaranteed. Still, for understanding user flows, and optimizing the operator's performance before the object detection kicks in, it would still be convenient for the operator to have a more intuitive design to his disposal.

5.1.3. Personas

For making design choices and evaluating the human-machine interface, personas are used. The more detailed reason for this is explained in sub-section 3.2.2 The two personas are based on interviews, and assumed profiles.

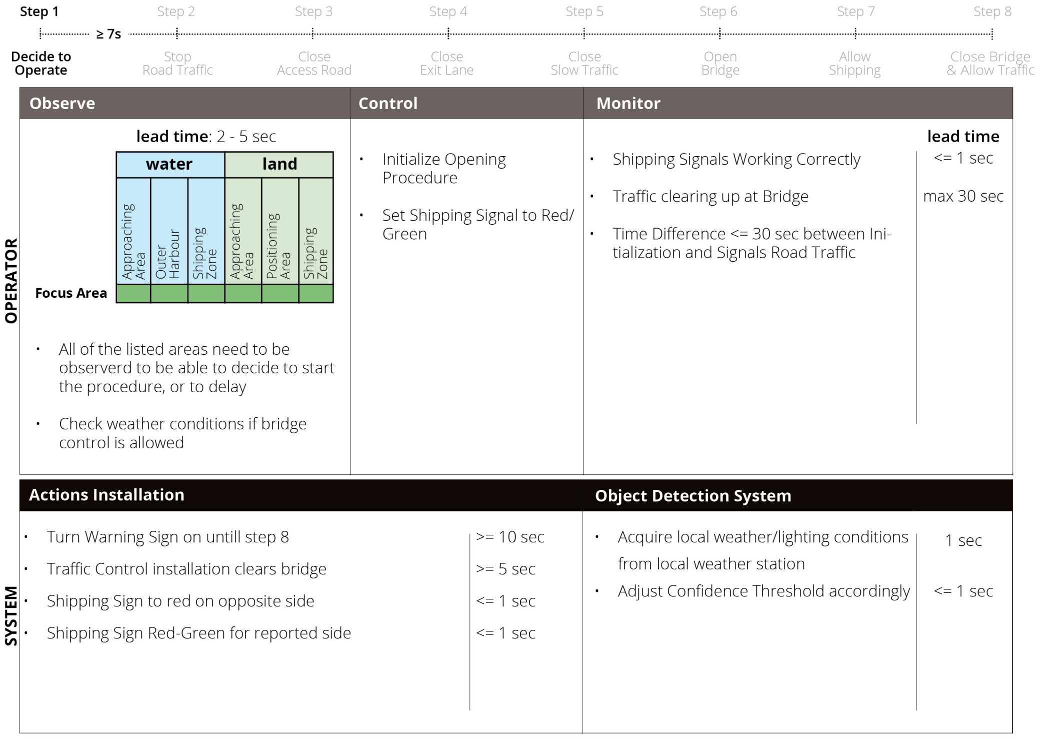
Persona 1: Emilie van der Laan
Emilie is 26 years old, and she has been working at the central post for almost a year now. In her free time she likes playing video games with her friends, and she follows the latest technology trends closely. Her father told her stories about when he was a bridge operator, and although things have changed, and she's not physically sitting next to a bridge anymore, she gets the excitement to work on this logistically challenging task.
Her goal is to do her job as well as possible. Maximizing the safety, and minimizing delays, is a challenge, but she is up for it! She listens to her experienced co-workers to get to know the tricks of the trade, but she is also curious to see whether these proven workflows, could be improved.

Persona 2: Jop de Groot
Jop is 59 years old, and he's been working at the central post for as long as he can remember. He started working there in his early twenties, and has always enjoyed working on his responsible job of controlling the situation on and around the bridge. He starts early, drinks a cup of coffee with his co-workers, and chitchats about last night's football scores. He is an experienced and very capable operator, and is rightfully proud of that. He showcases his knowledge by advising the inexperienced operators, like Emilie, in what the best way of working with the system is. His main goal at work is to operate the assets

in a safe manner, by using all his experience and skills, as he is convinced that is the best way to do it. He is a couple of years from retirement, and although he likes to work at the central post, he is already looking forward to it.

5.1.4. Tech & Organizational Workflows

The next step is to see how the object detection should be integrated into the current workflow. The actions performed by the system, are noted in the workflow steps, listed in appendix D. Here, the differences between the steps are highlighted in red, to keep track of the changes during the operational procedures. In these diagrams, both the system and operator workflows are depicted.



Step 1 ≥ 7s Step 2 Step 3 Step 4 Step 5 Step 6 Step 7 Step 8

Decide to Operate | Stop Road Traffic | Close Access Road | Close Exit Lane | Close Slow Traffic | Open Bridge | Allow Shipping | Close Bridge & Allow Traffic

OPERATOR

Observe

lead time: 2 - 5 sec

water			land		
Approaching Area	Outer Harbour	Shipping Zone	Approaching Area	Positioning Area	Shipping Zone

Focus Area

- All of the listed areas need to be observerd to be able to decide to start the procedure, or to delay
- Check weather conditions if bridge control is allowed

Control

- Initialize Opening Procedure
- Set Shipping Signal to Red/ Green

Monitor

	lead time
Shipping Signals Working Correctly	<= 1 sec
Traffic clearing up at Bridge	max 30 sec
Time Difference <= 30 sec between Initialization and Signals Road Traffic	

SYSTEM

Actions Installation

Turn Warning Sign on untill step 8	>= 10 sec
Traffic Control installation clears bridge	>= 5 sec
Shipping Sign to red on opposite side	<= 1 sec
Shipping Sign Red-Green for reported side	<= 1 sec

Object Detection System

Acquire local weather/lighting conditions from local weather station	1 sec
Adjust Confidence Threshold accordingly	<= 1 sec

Figure 36.: *Step 1: Decide to Operate (Intergo, 2019)*

5.2. Modeling

Before building the model, the system requirements should be stated, needed to solve the aimed problems. These are separated in technical and functional needs, subsequently, the constructed framework is discussed, and how the two personas interact with this framework. Lastly, the construction of the model is explained, starting with the assembly of the dataset, followed by selecting the model and training it.

5.2.1. Technical Needs

Object Detection

As the goal is to construct an object detection support system, the main functional need is object detection. The system should be able to detect the elements it is designed to. In this case, the goal is to detect bridge users present on the bridge between the closed barriers. This results in a system that should be able to detect pedestrians, cyclists, and motorized vehicles. In later stages, it could be possible to implement other types of elements, like ships, to decrease the chances of collisions, but in this stage, this is not the primary safety concern.

For the tests later in this chapter, the coordinates of the restricted area are hard-coded into the model. As the cameras are static, these coordinates will remain the same when the bridge is closed. Eventually, the goal is to also detect objects on the bridge during the opening procedure. An option would be to change these coordinates over time by linking these values to a gyroscope, or the time after initiation. However, a more flexible way would be by using the recently applied yellow paint. The model can be adjusted in a way that it automatically uses the coordinates of the yellow painted region, which is the movable deck, as it is capable of subtracting this area based on RGB values.

Detection Speed

The required speed of an object detection system, is fully dependant upon the task at hand. For his occasion, the system is of no help when it takes five minutes to run the model on an image, as the bridge would have been opened by then, making the prediction useless. In essence, for the system to be of added value, it should analyze a frame, and then provide the operator with enough time, to perceive the detection, observe the area of interest, and then react to it based on the situation. It is dependant on when the object detection is initiated, how much time this will give the model to give a right prediction. When looking at the workflow in appendix D, there is around 40 seconds between lowering the entry barriers, and the actual bridge opening. As users frequently neglect the closing barriers to reach the other end, the systems should take this into account, by starting the detection 5 seconds later. To ensure enough time to react, the aim is to detect within 10 seconds.

Detection Accuracy

It is difficult to give a definite answer on the question: "When is the model accurate enough?" In an ideal world, the goal is to achieve an accuracy of 100 percent in each frame, but in practice, this is not feasible. It is also dangerous to go with the assumption that even if the model has a hit rate of only 1 or 2 percent, there is added value, as it is a support system that works parallel with the monitoring operator without affecting his workflow. This approach assumes that there is absolutely no decreased awareness of the operator, knowing there is a backup system in place, let alone taking the distractions caused by false detections into account.

Based on the situation at hand, the system should detect a user within 10 seconds, in every lighting and weather condition. The goal of this system is to increase the safety instead of trying to get a perfect accuracy per frame. As the system will be working on multiple frames per seconds, on multiple cameras, it should be sufficient if at least one frame during this period is correctly analyzed. However, to ensure this, a high hit-rate per frame should indicate the models effectiveness.

In this research, the hit rate per input clip will be measured, randomly taking one or two minute clips of CCTV footage of the bridge. In these test, the effect of different weather conditions will be considered, and will focus mainly on situations with just a couple of persons on the bridge, as in real-life, the operator will probably have less trouble with detecting groups of people on the bridge.

Data transmission

When it comes to data transmission, not much should change. The research aims to work mostly with the current infrastructure. The detection system works with the standard video signal already transmitted to the central post. Currently, this is displayed on the operator's system and it is saved on a media server. On the latter it is stored for approximately a week to be able to re-watch it in case of incidents, and will be deleted when not needed. The detection system can be used as a cloud service, but a local multi-GPU computer will work as well. In this report, the system is composed with the latter in mind, as it is easier to upscale the model in a later stage than downscaling it, and as the current system is mainly running locally, it would be a smaller change to the current infrastructure.

5.2.2. Functional Needs

Notification Types

After a detection has been done, it should be communicated to the operator in the form of a notification, in order for the operator to act on it. Before going into the actual graphical interpretation of the communicated information, first, a decision should be made on a higher design level. Should there be different notification types? To what aspects are they linked? Too many different notification stages could make it difficult to distinguish severity.

For this reason, two notification types have been chosen. One for detection when the barriers are lowered, but the 'open bridge' command, hasn't been given yet, and another one for when the 'open bridge' command has been activated. Because the urgency of the latter is bigger than the former, the notification for this occasion should be more prominent. Therefore, the choice is made to only give a visual notification for the times where the 'open bridge' command hasn't been given, and an audio-visual + visual command for when the bridge deck is moving.

For visualizing the detections, the 'when' is decided upon, but two questions remain: Where should they be presented, and how? The two main options for where, are directly on the camera streams or on the SCADA interface. To be able to perceive the detection as fast and accurately as possible, the annotation on the camera stream is the most important option of the two. However, because the SCADA system shows command logs, and the operator can be focussed on this system, a warning symbol should be shown that informs the operator of a detected instance.

To answer the 'how' questions, the options are shown in figure 12, whereas explained in the perceptual stage subchapter, a speedy way of communicating the detected instance, could mean the difference between a risk firing, or not. Both options 3 and 4 help the operator to focus on the instance, and then it comes down to personal preference. According to the survey, 60% of the respondents preferred the third option. The remaining 40% preferred the second option, where a red border is drawn around the entire screen, only indicating which camera captured the detection. The main reason such an option isn't preferred from an HMI point of view is, as mentioned in an earlier section, the added confusion in case of a false positive, decreasing the operator's understanding of the system, and with that, the trust in it.

For the audio-visual notification, the main objective is to make it easily distinguishable from the ongoing audio-visual information in the work environment, by adjusting the frequency and audio level for example.

Initiation Moment

An important design choice that has to be made, is when to start analyzing the input images. An early detection, could increase the safety of the asset. This means starting the object detection model as soon as the operator decides to act upon a shipper's request. This way, the operator could closely follow the detections when clearing the bridge, continuing the procedure without sudden cancellations or delays of commands. At first glance, this would increase both the safety and the straight flow in case of a misplaced object/user.

However, this choice could lead to the problem of operators completely trusting the system to detect users, which could decrease their awareness. This shouldn't have to be a problem in case of an object detection system that is a hundred percent accurate, but unfortunately, it doesn't work that way in reality. The model will come up with false positives(FP) and false negatives(FN). False Positive being an occasion where the system wrongfully notifies that a user is detected. On this occasion, that wouldn't be that big of a problem, because the operator would look at it, ignore it, and no harm is done. A False Negative on the contrary, would mean the system fails to detect a user, which in this case is much more dangerous. Because of the reduced concentration of the operator, chances are that he would also miss the person in the prohibited area, possibly leading to a severe accident.

Given this insight, it may be wise to wait for a couple of seconds after the barriers are all closed, before initializing the detection model. This way, the operator must still concentrate on the procedure, but has the Object Detection model as a safeguard. This way probability of error at the operator side isn't negatively affected by the detection system, and so the overall probability of error is smaller.

False Detections Tolerance

Where in the previous section, False Negatives were an important element to consider, in this one it is about False Positives; the occasions where the system wrongfully detects an element of interest on the bridge. The problem here has to do with the confidence threshold of the system, so the certainty of a detection being a right call. By setting the confidence threshold on the low side, say 30 percent, a lot of dimly lit users or uncommonly dressed people will be detected, that otherwise maybe wouldn't be detected, because the images in the training set of the model, aren't too similar to the detected subjects. However, because of the low threshold, forms and shapes that show similarities with the intended category may also be reported. Dealing too often with such a false alarm, the trustworthiness of the system may decrease, making the operator pay less attention to notification. This effect is called the cry-wolf syndrome or alarm-fatigue, as explained in the theoretical framework. Different kinds of research show different effects, but most do agree that a good understanding of the system, reduce the potential negative effect of these false alarms.

This phenomenon, was verified in practice when the Bosch system was piloted for a week at the CP. According to one of the interviewed operators, a false positive occurred once every bridge opening. Also, the system only gave the notification that something was detected, without communicating where the system thought the element of interest was . The operator subsequently needed to recheck all cameras in detail to find out that there were no users in the prohibited area. There was only one moment where there was a user on the bridge when it wasn't allowed to. For the operator, not having experienced the potential of having a system functioning as a safeguard for missed bridge users, the system went from being a support tool, to being a distracting extra task. The system was subsequently turned off.

Thus, setting the confidence level too low could lead to troubles working with the system. The other side of the spectrum is setting the confidence threshold on the higher side, say 95 percent. This has the benefit of reducing the false positives, as the system only notifies the operator when there is little doubt the detected object is of interest. Unfortunately, such a setting has the downside of missing a lot of instances, because every dimly lit subject or uncommonly dressed person, will not be above this threshold in terms of detection confidence, leading to them being filtered out. These False Negatives, or simply missed objects of interest, also affect the reliability and usefulness of the system.

There are several ways of dealing with such a problem. An apparent one is to use different confidence thresholds for different weather/lighting conditions. When it is broad daylight, the image quality is good, and the people well lit. In this occasion, the threshold could be on the high side, as the instances where the system is in doubt, are either not the objects of interest, or they will probably be detected with high confidence in one of the surrounding video frames. When it is misty, the image quality is poorer, and therefore the threshold should be lowered. Synchronizing the detection model with a local weather station could put these adjustments into work.

Besides this option, there is also the fact that multiple cameras are covering the same area. Comparing the values of the different viewpoints, and setting thresholds for multiple cameras combined, could also help in reducing false positives. When doing this, time could also be used as an element to work with the system and confidence rates in a more integral setting. If an instance has a certain confidence below the threshold, but is detected for a certain amount of time, this could also be regarded as a way of making a more accurate prediction.

Lastly, which is a more complex solution, but feasible with static cameras as present on bridges, is the element of distance. With multiple cameras it is possible to make a 3D stereo reproduction of the situation on the bridge. Here, elements as size and speed can be estimated, to accordingly check whether these match the characteristics of the objects of interest. Combining this with condition-based confidence thresholds could boost the accuracy to a level where the balance of the number of detections and the number of false positives become less relevant.

Requirements Summarized

- Detect bicycles, motorized vehicles and pedestrians
- Should detect users in different lighting and weather conditions
- Detect users between the barriers within 15 seconds after closing the entry barrier, starting after 5 seconds.
- Have a high hit-rate
- Reduce false detections to 1/10 openings or less.
- Make the false detections easy to interpret.
- Avoid operator's reliance

5.2.3. Framework Construction & HMI Validation Scenarios

For constructing the model, the functional requirements as described in the previous section should be met, while making use of the HMI learnings from literature, and the constructed personas based on interviews and observations. In this section, the working of the conceptual framework will be described, and the way the proposed personas, Jop and Emilie, interact with the system. The confidence rates mentioned in the model's logic gates, are indicative, and may vary based on weather conditions and testing.

Main Application

For the main application, the goal is to produce as few false positives and false negatives as possible, while keeping the operator's workflow as similar as possible. According to the surveys, operators are proud of their work and want to have their experience and skills valued. This should be reflected in the working of the model.

The state diagram of the framework is depicted in figure 37, where the upper box outlines the data processing that happens in the background, and the lower box describes the way the detections are communicated to the operator. After the diagram, some of the functions will be highlighted in more detail.

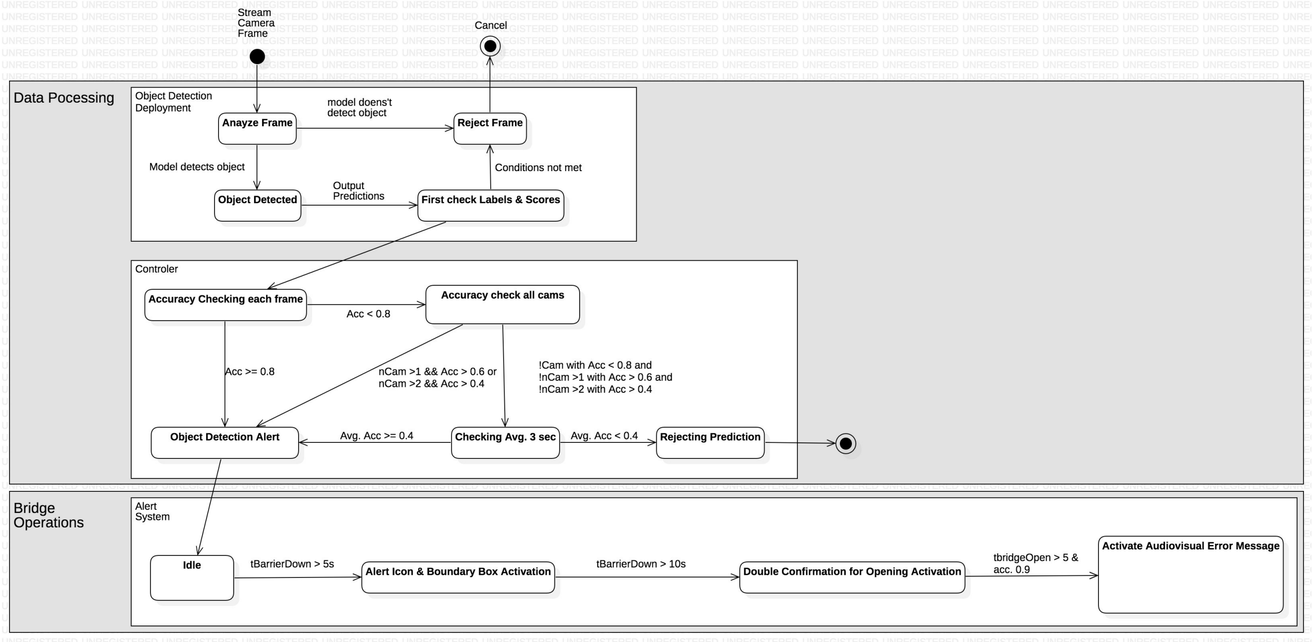


Figure 37.: *State Diagram Framework*

The first step is to run the object detection model on the available static cameras. For each frame, the framework will check for detected instances, and passes it through a first condition check. If a frame meets the conditions, it will go to the next stage, and otherwise, it will be rejected.

In the next stage, the controller, the labels and scores will go through another gate. If the score is above a certain value, 0.8 in this case, it will go directly to the alert system, where it is compared to other conditions, as described in the next step. If not, it will be put together with the frames of other cameras that passed the first condition check, with the same timestamp. Upon these average conditions, the system will decide whether to push the detection to the alert system, or compare these results with the frames of the last three seconds. This is the last stage where detections can make it to the alert system, otherwise, they will be rejected.

The last step, is the alert system. Here, it is dependant on the stage of the bridge opening process, whether a detection will be displayed, and how it will happen. When the barriers haven't been lowered, or it is within 5 seconds after giving the command, the detection will not be displayed. If an instance is detected after these 5 seconds, the system will show an alert icon on the SCADA system, and a boundary box on the related camera screen. If the barriers have been closed for longer than 10 seconds, and an instance has been detected, a double confirmation will be asked for opening the bridge, asking the operator to monitor one more time. If the 'open bridge' command has been given, and the detection system is convinced that an instance has been missed, an audio-visual notification will be given.

Thus, how is persona Jop, who is the most conservative and tech-averse of the two personas, taken into account in this design? As Jop is convinced using his eyes, experience, and skills is the most reliable system, he wants to stick to normal as much as possible. The proposed system tries to reduce the false positives as much as possible, by going through several logic gates. While doing so, it tries to reduce the false negatives as much as possible, by not directly rejecting frames with a lower score. It first combines them with other frames from that moment or around that period, before it finally decides whether to reject or show them.

Then, when an instance is detected, it will show on both the SCADA display, and the camera stream, by drawing a boundary box, adding to the understanding of the system. It will not start of with audio-visual notifications, but it allows Jop to see the detection, and act upon it. Subsequently, it asks for a double confirmation. The survey and observations, showed that the bridge operators did not want the support system to have any operational permissions. By asking a double confirmation, the system will not stop the procedure itself, but keeps the decision-making for the human operator, showing that his experience and skills are valued.

Lastly, when Jop does his job perfectly, and the false positives of the support system are brought to a minimum, Jop will barely notice the existence of the system at all. This way, he can perform his work the way he is used to, while an additional safety system is running in the background.

Model Trainer

As described in the theoretical framework, one of the main downsides of using a convolutional neural network, is the need for a big dataset to train on. Putting an inclusive labelled training set together, takes time and dedication, but will improve the model's performance.

This is where the persona Emilie comes in. Being a tech enthousiast and eager to learn, she can ask for permission to contribute to the ongoing development of the model. The object detection support system she will be working with, will have a lower confidence threshold than the normal setting. This way, the model will produce more false detections, as it will communicate detections at a lower confidence level. Emilie is not hindered by these false detections, as the understands the reasoning behind it, and by telling the system whether a doubtful detection is a true positive or a false positive, the system can add this annotation to its training set that is running in the background. Periodically, the current object detection model will be replaced by the model that is retrained by Emilie's help, improving the overall working model.

By implementing this approach, Emilie will have an increased job satisfaction by feeling more valuable in a way that in closely related to her interests, and the entire organization will benefit from it.

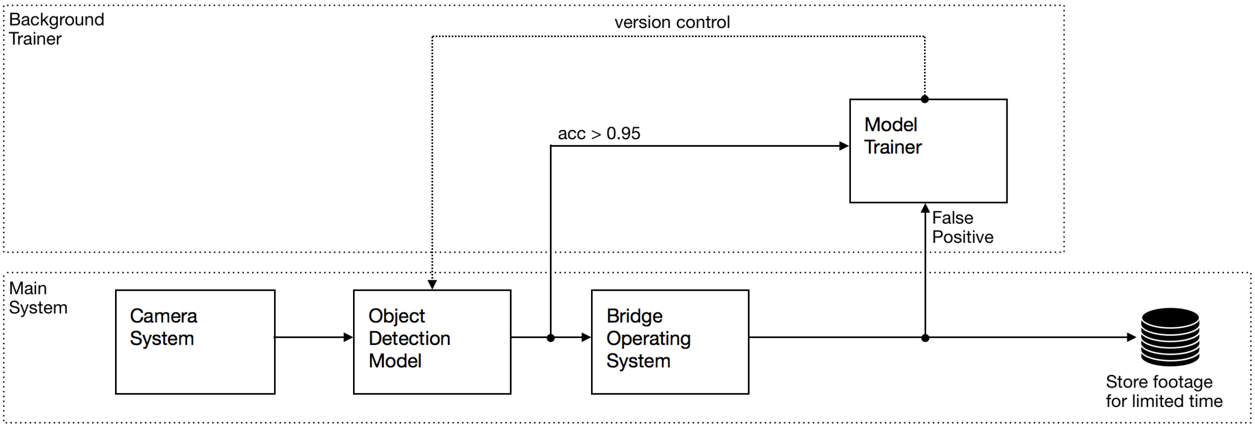


Figure 38.: *Model Trainer*

Permissions

For bringing this all together, the proposed concept for permissions within the organization are depicted in table 5.1. The way the operator and operator(trainer) work, has already been described in this section. The team supervisor, will get access to a dashboard where he can monitor the interruptions the object detection system had to make. Hereby, he can monitor the operator's performance, by seeing how often the support system had to intervene. The supervisor should take false positives into account, and anomalous behaviour of bridge users, but big differences in operator's scores, could indicate performance issues. This could then be used to coach or retrain the concerning operators.

The functional manager may have some additional insights in the model performance. As not all camera systems on the bridges are similar, he could see how the different cameras and angles affect the model performance. Also, tracking could be included in a later stage, and this way he could monitor the usage of the bridge.

The technical manager should be able to adjust the model when needed, and keeps an eye on the version control. When it becomes apparent that the sensitivity of the model needs to be adjusted, he should be able to modify the confidence thresholds.

Table 5.1.: *System Permissions*

Function	**Detections**	**Model Trainer**	**Detection Statistics**	**Model Performance**	**Trainer Preferences**	**Version Control**	**Platform**
Bridge Operator	x						Operating System
Bridge Operator(Model Trainer)	x	x					Operating System
Team Supervisor	x	x	x				Operating System & Dashboard
Functional Manager			x	x			Dashboard
Technical Manager			x	x	x	x	Dashboard

5.2.4. Training the model: In Practice

In chapter 3.3, a theoretical explanation has been given on training a CNN. In this section the steps to train an actual working model are discussed in short. This process is decomposed into two steps. Importing and parsing the dataset(1), and selecting and training the model(2).

Importing and parsing the dataset

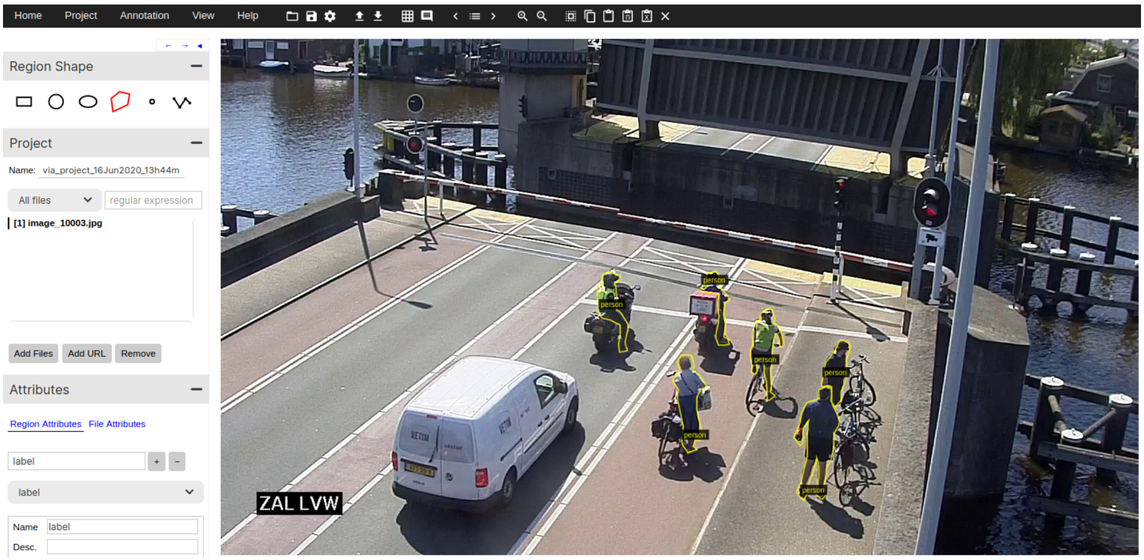
Before the actual importing and parsing of the dataset, it is vital to have a clear goal for the model to achieve, and this results in the kinds of data that are needed to gather for training and test purposes. In this research, the goal is to make the model detect vulnerable users of the bridge, based on the existing infrastructure present at the assets. This suggests that the input images should be of comparable characteristics of the camera footage that it is planned to work on. This means the inclusion of different kinds of image qualities, as they vary from one bridge to another, different weather conditions, bridge users in several shapes and forms, and so on. Also, the viewing angle should be comparable to the video stream in the desired end product.

It is possible to mimic these characteristics by setting up a camera system in an environment that is about the same as the bridge setting, but it is easiest to just use the actual camera footage from the bridges themselves. As one of the benefits of working with a CNN model is the transferability to other environments, it is not needed to have images in the training set of every bridge the system is designed for. As long as the characteristics that are to be expected on the bridges that aren't analysed, are represented in the training set. Luckily, the footages of the bridges are stored on a network drive, and retained for about a week. This is meant for safety purposes, to be able to analyse the footage when needed.

For training purposes, the availability of this footage is of great importance, as is makes it possible to make a big training set, a prerequisite of building an accurate CNN-based model. However, because the footage is only retained for about a week, and the access to the database was granted in May, the database didn't contain footage of challenging conditions as to be expected during winter and autumn. Also, it made it more difficult to find camera footage of dim lighting conditions with a lot of activity of pedestrians, as in May the sun is down between 21:30 and 6:00, and in December between 16:30 and 08:30 (KNMI, 2020). In the Netherlands, rush hour is approximately between 06:30 - 09:30 and 15:30 - 19:00 (ANWB, 2020), which makes it easier to gather capture training and test footage in December.

Having gathered the video footage needed to train the model, the next step is to subtract the individual frames, and annotate them. The goal here is to add the labels(classes) to the image, so the model can either train on them by calculating the loss by comparing the labels with the model's predictions, or evaluate the training results by looking at the loss of the validation/test set.

Annotating these labels can be a tedious job, as the developer of the model has to draw polygons(for masks) or rectangles(for bounding boxes) around the intended instances, in this model persons, to feed into the network. These geometries should then be labeled with the right class.



Name: via_project_16Jun2020_13h44m

[1] image_10003.jpg

label

label

Name label

Desc.

Figure 39.: *VGG Image Annotator (VIA)*

As stated before, CNN benefits from having a big training set, but as the process of annotating is time-consuming, the training set used for the model has little over 100 images, where most of the images contain multiple persons. For an end product, this is too small of a dataset, but for this proof of concept it seems sufficient. As a comparison, the COCO dataset contains over 200.000 labeled images, divided over 80 categories.

The next step is to format the annotations into a format that is processable by the model. The exact way of formatting defers with different frameworks, but the file format and the key ingredients are mostly the same. The annotations are often saved in a JSON(JavaScript Object Notation) formatted file, and has for every annotated image the file path and size, the pixel coordinates of the mask/bounding box and the corresponding labels.

Selecting the model

Different types of CNN-based models can be used for the task at hand, and the goal is to select the model that is best suitable for the job. The main trade-off that has to be made is between speed and accuracy. The more accurate a model is, the slower it tends to be, because these models are computationally expensive. So, the combination of the available computational resources and the acceptable detection duration should lead to a model of choice. Then, there is a difference between object detection and object segmentation. For object detection the goal is to detect objects in an image, and draw a bounding box around it, and with segmentation the goal is to draw a pixel-wise mask for the desired objects. The difference can be seen in figure 12, where option 3 is detection, and 4 is segmentation.

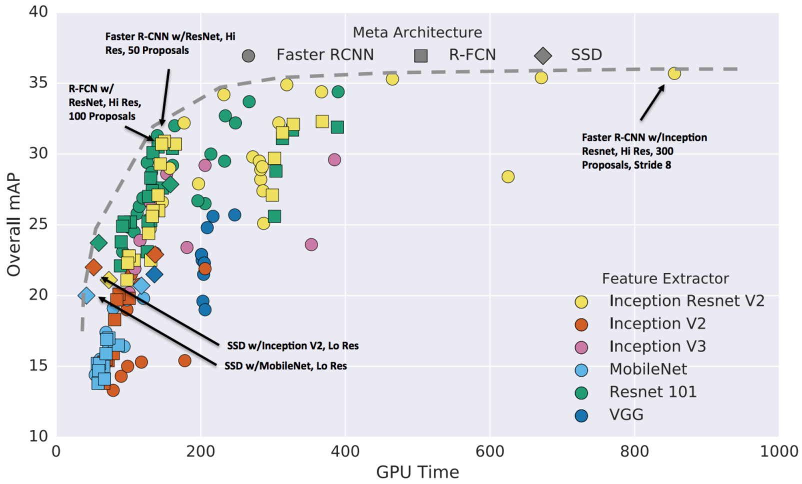


Figure 40.: *Speed/Accuracy Trade-off (Huang et al., 2017)*

The model picked for this research, is a Mask RCNN model, with a Resnet-101 backbone. In figure 40 is displayed, how the Resnet-101 feature extractor, manages to combine accuracy with processing speed. In this figure, Faster RCNN is displayed, but Mask RCNN is chosen, because it is the evolved version of this architecture, adding object segmentation functionality, and improving the accuracy and speed of the model. Where the Mask RCNN predominantly takes care of the training, the input for the feature extraction phase, and the interpretation and display of the object detection results, the heart of the object detection model is the Resnet-101 model, as it performs the convolutions and classification. So how does the Resnet-101 compare to the basic CNN architecture as described in the theoretical framework?

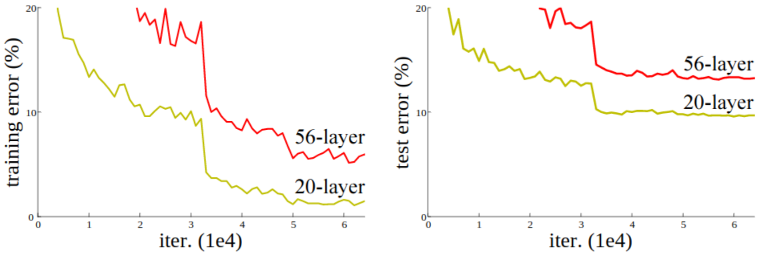


Figure 41.: *Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. (He et al., 2016)*

As explained, adding more layers to the model, the complexity increases, and higher accuracy can be reached. A logical assumption would be to take the basic CNN architecture, and add numerous convolutional layers to improve the model. Unfortunately, in practice adding 'plain' layers to the model, decreases the model's performance from a certain point, as accuracy gets saturated, and then degrades rapidly(fig 41). Although it is possible to significantly deepen a network without affecting the performance by adding identity mapping layers to a shallow network. Unfortunately, experiments show that traditional solvers are unable to find solutions that are comparably good or better than this.

In 2015, this degradation problem was tackled by (He et al., 2016), by introducing deep residual learning for image recognition. Using residual learning blocks with shortcuts, also known as skip connections, in the convolutional layers, activations from previous layers could be used, resulting in identity mapping. Also, by skipping layers, the problem of vanishing gradients can be mitigated, and the training time shortened. This way, Resnet-101 achieved a depth of 101 layers, where VGG only has 19.

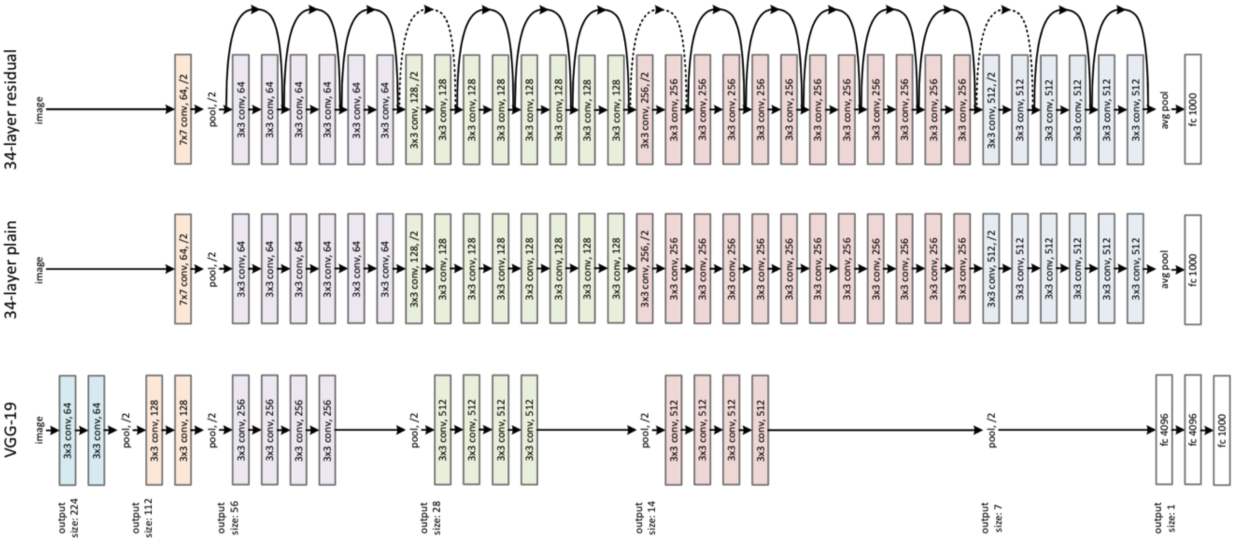


Figure 42.: *Resnet-34, 34-layer plain and VGG-19 visualized (He et al., 2016)*

Model Specifications

Custom Model	*Pre-trained Model*
Model	**Model**
Mask R-CNN	Mask R-CNN
Resnet-101	Resnet-101
Dataset	**Dataset**
Custom Dataset	COCO Dataset
Dataset Features	**Dataset Features**
150 images	330.000 images
1 object category	80 object categories
Object Segmentation	Object Segmentation
CCTV Quality	Different Image Sources
Bird-eye Perspective	Varying Perspectives
Data Augmentation	**Data Augmentation**
Rotation 20 degrees L/R	None
Flip L/R	
Speed(GTX1060 6GB)	**Speed(GTX1060 6GB)**
4 fps	4 fps

5.3. Evaluation

For the evaluation of the model, three bridges in Zaandam are considered. The three bridges have differences in camera quality and bridge users, so working with these will provide a representative overview of the application of object detection on remotely controlled bridges.

	Camera Quality	**Environmental Characteristic**	**nOpenings 4th week of Jan 2019**
Alexanderbrug	+	Near Public Transport Station	151
Bernhardbrug	++	Near School	130
Zaanbrug	-	Public Transport and Connecting Municipalities	121

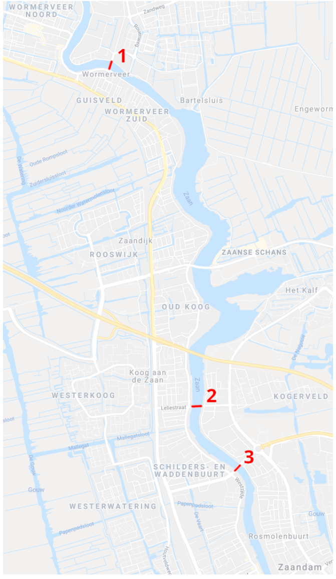


Figure 43.: *1.Zaanbrug 2.Alexanderbrug 3.Bernhardbrug*

5.3.1. Model Performance

Having trained the Mask R-CNN from scratch, it's interesting to see how the model compares to the model that was pre-trained on the COCO dataset. This is done by looking at the hit rate, also known as the true positive rate or sensitivity of the model. This is calculated $HitRate = TP/(TP + FN)$. Because of a limited test set, it is difficult to measure the performance of the system concretely. The results presented in this section will provide experimental insights, which will indicate the model performance, and the overall effectiveness of the model. The tests were done only with the object detection model on one camera stream, not going through the logic gates of the controller and alert system as depicted in the state diagram of figure 37

Test case 1: Alexanderbrug

In this test, two ways of counting the false negatives(misses), are applied. In the first way of counting, every detectable person will be counted, regardless of the portion of the person's body that is in the frame. The second way of counting is by ignoring the instances where the person is for more than approximately 50% covered. In figure 44, screenshots are depicted with the types of detections that were left out in the last way of counting.

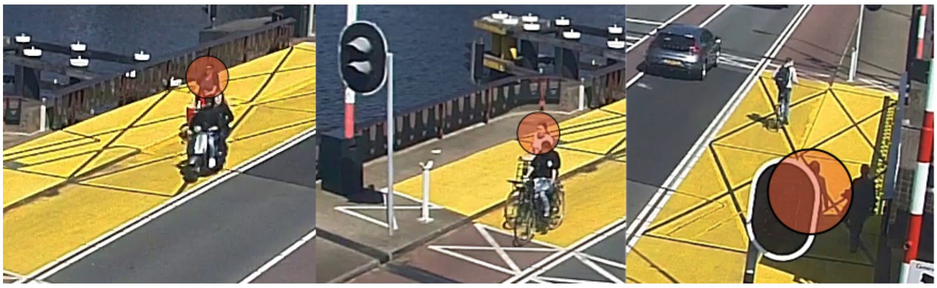


Figure 44.: *False Negative through overlap*

The Reason to do both is that it is valuable to have insights to what extent the model is capable of detection only based on certain body parts. This could show the model's usability in case a person is only (partially) visible on one camera, because of a poor camera plan, or malfunctioning cameras. On most occasions, a person will be visible on multiple cameras, so when the person is behind a traffic light, another camera can still detect this user. Also, when a person is missed because it is behind another person, this would not be an issue in the application this research is after, as long as the person in front is detected.

This test was done in broad daylight, and with a confidence rate of 0.8. The conditions and angle can be seen in figure 45.

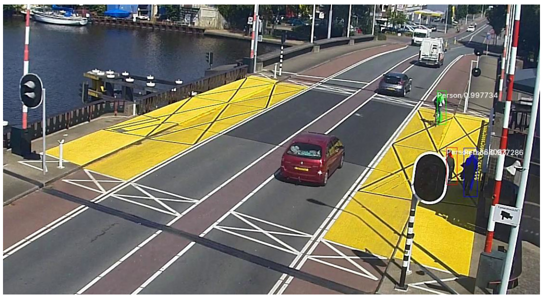


Figure 45.: *Alexanderbrug Screenshot*

Custom Model				
Confidence Threshold: 0.8	**True Positives**	**False Positives**	**False Negatives**	**Hit Rate**
False negative through overlap counted	293	0	49	0.86
False negative through overlap not counted	293	0	27	0.92

Pre-Trained Model				
Confidence Threshold: 0.8	**True Positives**	**False Positives**	**False Negative**	**Hit Rate**
False negative through overlap counted	238	0	104	0.7
False negative through overlap not counted	238	0	82	0.74

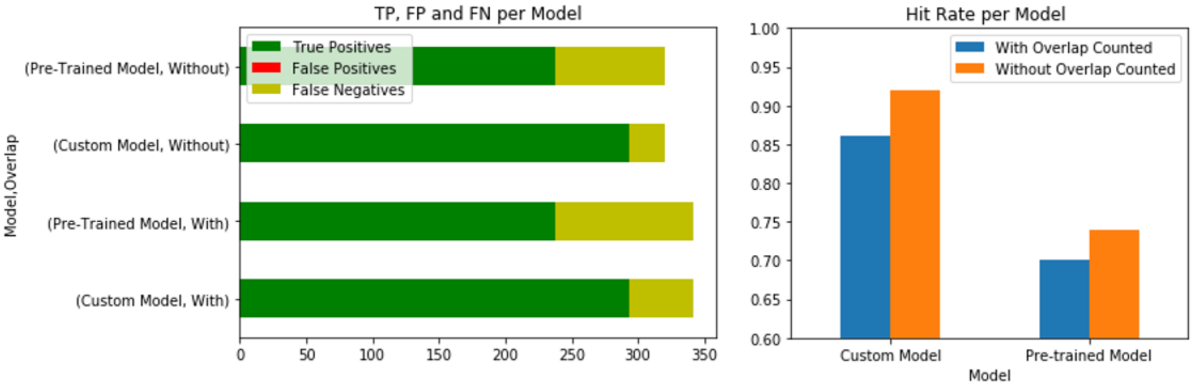


Figure 46.: *Results Visualized*

The model has most difficulties with persons walking in the shadow, filmed from behind. Indicating a high possibility that the missed detections would have been detected on another camera. Apart from this, no person was left undetected longer dan 2 seconds.

The performance of the custom model was better than the performance of the pre-trained model by a fair margin.

Test case 2: Bernhardbrug

In this test case, also two different scores were measured. For about 1.5 minutes, a scooter stood on the bridge, hardly moving at all. For this research, that was quite interesting, as both accidents in Zaandam happened with persons standing still, and for the operators, these are the most difficult to spot.

For this test, evening conditions were chosen, as they similar to the angle and conditions of the accident, with an confidence rate of 0.5. The conditions and angles can be seen in figure 47.

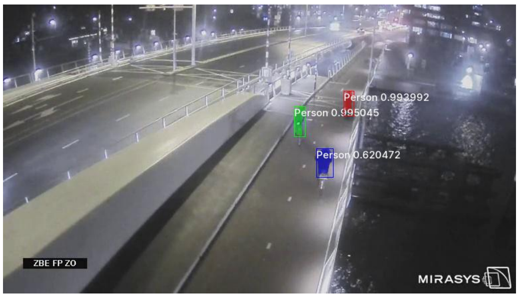


Figure 47.: *Bernhardbrug Screenshot*

Custom Model				
Confidence Threshold: 0.5	**True Positives**	**False Positives**	**False Negatives**	**Hit Rate**
Non-moving scooter not taken into account	312	0	35	0.9
Non-moving scooter taken into account	836	0	46	0.95

Pre-Trained Model				
Confidence Threshold: 0.5	**True Positives**	**False Positives**	**False Negative**	**Hit Rate**
Non-moving scooter not taken into account	199	0	148	0.57
Non-moving scooter taken into account	538	0	344	0.61

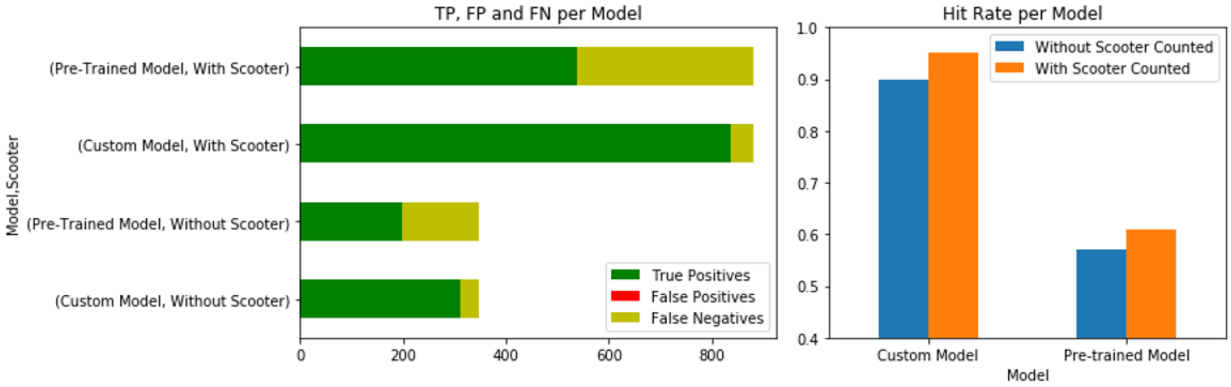


Figure 48.: *Results Visualized*

Here, the custom trained model made the biggest difference, achieving a hit rate of 0.95 with the scooter counted, where the COCO model achieved only a 0.61 percent hit rate in this occasion.

Test case 3: Zaanbrug

This test was done in the evening, and a confidence rate of 0.5 was used. The conditions and angles can be seen in figure 49.

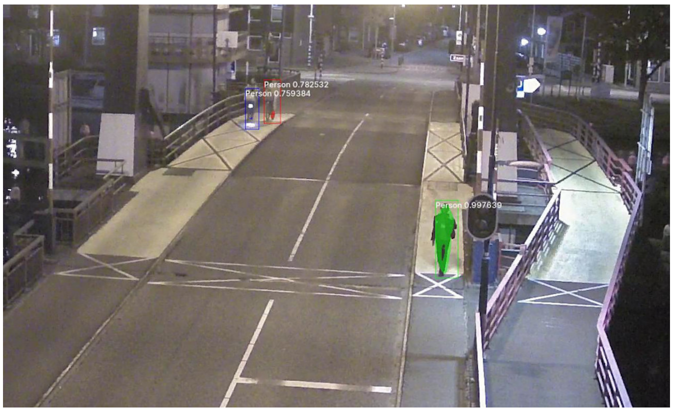


Figure 49.: *Zaanbrug Screenshot*

Custom & Pre-Trained Model				
Confidence Threshold: 0.5	**True Positives**	**False Positives**	**False Negatives**	**Hit Rate**
Custom Model	40	2	9	0.82
Pre-Trained Model	28	0	21	0.57

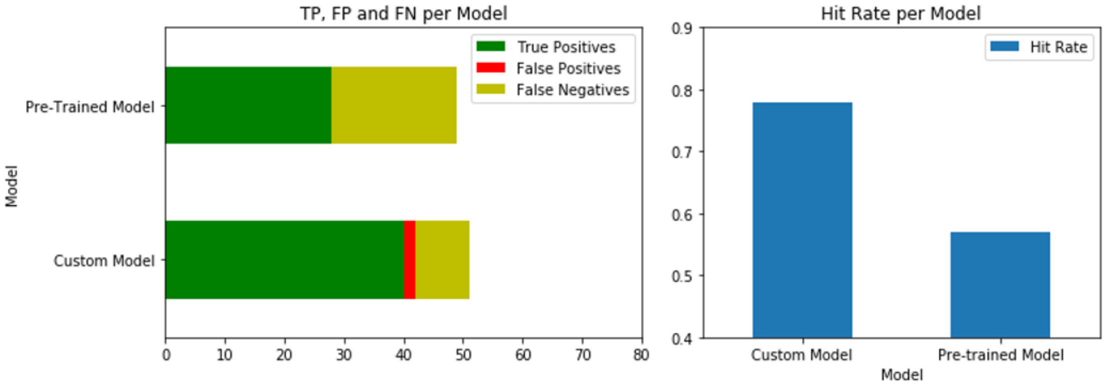


Figure 50.: *Results Visualized*

In this occasion, the model had most problems, possibly caused by the image quality that was worse than the other bridges. Also, this is the only test where the model produced false detections. This was twice on the same spot, being part of the railing of the bridge. The coco-model, although scoring a lower hit rate, did not show any false positives. For this test, the number of bridge users was the smallest.

Incident evaluation

The stimulus for this research was the occurrence of the tragic accidents, on the Den Uylbrug, and the Bernhardbrug. Those are the types of occasions where the support system should show its value, by increasing the safety. It is impossible to say whether the accidents could have been prevented by the system, as running test cases in a controlled environment is different from running the system in the real world.

However, by running the model on the image subtracted from the report of the Dutch Safety Board, it is possible to get a sense of the technical feasibility of detecting the elderly couple on the Bernhardbrug in this case. Both wearing dark clothes, and carrying an umbrella, the model managed to detect the couple in this particular image, as shown in figure 51.

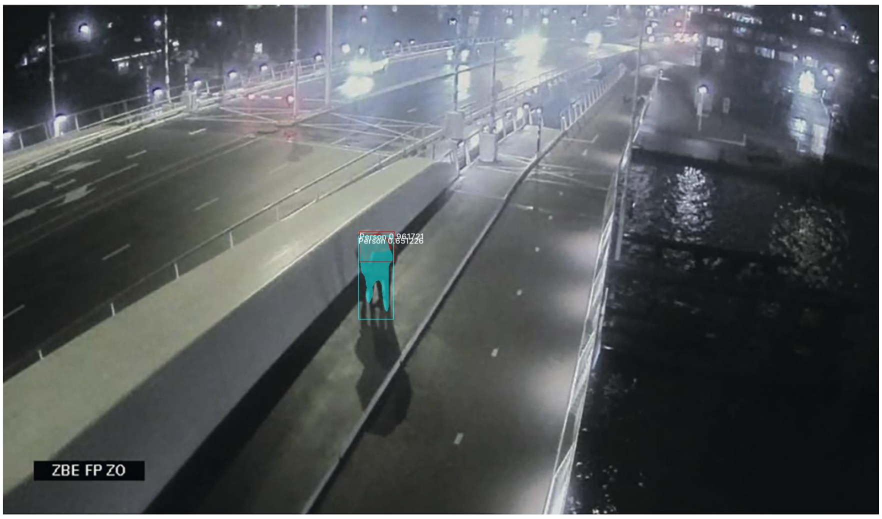


Figure 51.: *Victims Bernhardbrug Accident Detected (input image retrieved from (Onderzoeksraad voor de Veiligheid, 2019))*

For the accident on the Den Uylbrug, there is no representative screenshot available from that occasion to run the model on. However, the conditions of the masked image(fig 52 do show that the lighting and camera angle is comparable to the footage from the first test case of this report, giving the impression that the model could have detected the victim in a controlled test environment. Further development and implementation could in the end show whether the theoretically capable model could contribute in the real world.

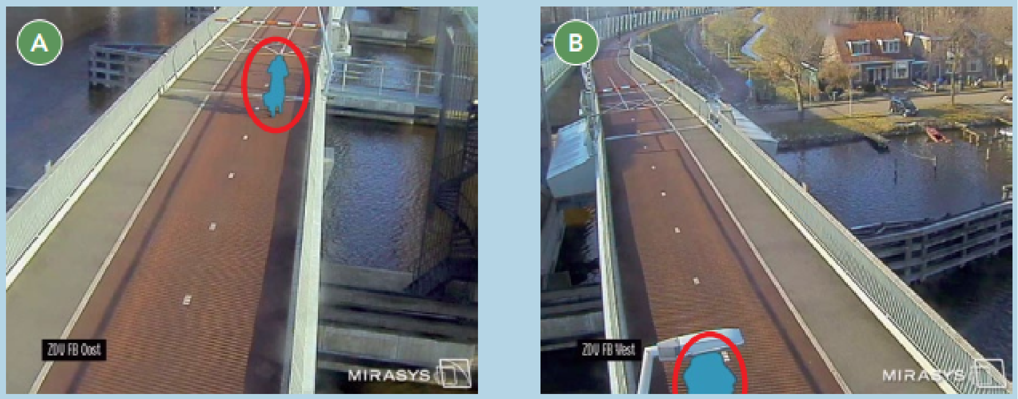


Figure 52.: *Masked Image Den Uylbrug (Onderzoeksraad voor de Veiligheid, 2016)*

6. Discussion

In this discussion the methodology and the results will be discussed, elaborating on their limitations and considerations. After going over these two elements, these findings will be translated into key points for further research.

6.1. Methodology

The methodology that was used in this research, was mainly based on action design research, with concrete steps subtracted from goal-directed research. The main reason for choosing these research methodologies was because they were invented for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting, where traditional design science does not fully recognise the role of organizational context. The reason why involving the organizational context was so important in this report, was that the Dutch Safety Board identified the lack thereof as one of the key findings in the accident investigations. They stated, that the safety of the asset was mainly seen as a technical problem, instead of approaching safety in a holistic way, where the interaction between human, machine and environment should be considered.

Fundamental to this methodology was involving operators from an early stage by confirming assumptions through questioning, or by observing their current challenges and worries. Apart from being valuable in terms of getting the specifications and setting right, it is also a good way of easing the operators into this new technology, by providing a sense of ownership, and through tackling worries in an early stage. It is perfectly imaginable that operators fear that a (semi-)automated system such as object detection will eventually make their jobs redundant, as this fear has been present since the introduction of computers. By developing the system in collaboration with the operators, they will likely understand the motivations behind the system, and how it is used to complement and support the operator, instead of aiming at replacing them. The understanding of both the reason for existence, and the working of the system, is very important for successful integration, as shown in the theoretical framework. The national survey that was discussed in this research shows that many operators miss this kind of involvement in their jobs. It gets the message across that their skills and experiences are indeed valued, which could subsequently lead to better job satisfaction.

To conclude, in this research, a gradual development with close collaboration was chosen. However, during this period, a pilot study was conducted where operators had to try out working with the object detection system already installed on the camera system. This system was mainly developed for security purposes, and there wasn't a clear implementation plan on how it would work with the current system. As a result, the operators that were observed at the central post were not familiar with the system. According to them, the detection system warned for a false positive almost every bridge opening without showing the area of interest, leaving the operator guessing what it could have been.

In itself, this is not an ideal pilot study. However, what will happen if the feedback of the operators is written down, and the models get adjusted accordingly? Will it function worse compared to a situation where the operators are involved from the first drawing? What is the difference in expenses and time between using this approach for developing and implementing such an application compared to working together intensively from an early stage? Would, after using the system for a certain amount of time, the understanding of the operators and acceptance towards such a system be compatible with a situation where these were built from the ground up? It is difficult to answer all these questions, and therefore to justify one approach over another. Moreover, it is difficult to validate how this pilot affects the outcomes and potential acceptance of the system that could be build based on the finding in this report. Negatively, operators can approach a new system less open-minded because of the disappointing experiences of working with a similar system. Regardless of a completely different system architecture, the goal and mental model is overlapping.

On a positive note, after using the system in their work, the feedback can be more valuable than it could have been when working with a proof-of-concept in a controlled environment, or working with hypothetical situations. As the time window for this research was too short for actually producing a production-ready system, it was not possible to generate these real-life experiences. Also, when a CNN-based object detection model is introduced, and it outperforms the pilot project application, the operators could notice this increased effectiveness, feeling more comfortable working with the system.

Moving away from the pilot project, and the way it was implemented, brings up a different challenge experienced with the used methodology. When working together with committed practitioners, building on their skills and experiences, when should the research theory overrule personal comfort and preferences? In an ideal world, the preferred way of the practitioners complies with the best practices from research. But what if this is not the case? What is the effect of asking for input and making the operators feeling heard, but yet subsequently impel the project to a different direction?

Take for example the moment of object detection, and the way of presenting the results. The survey could indicate that operators feel most comfortable by running the object detection system from the very beginning of the process. This way, the system starts detecting and communicating these detections, even before the barriers are lowered. However, research may indicate that this builds such a reliance on the system, that the operator isn't trying to perceive human-shaped figures anymore. But he might solely look for bounding boxes generated by the detection system instead. Chances are that the operator will miss a crucial detected instance as he may be less concentrated, since the importance of utmost concentration is reduced through the implementation of a semi-automated system. Moreover he might miss the instance caused by inattentional blindness, as he's expecting a boundary box instead of a person.

In this scenario, it may be safer not to go for the operator's preferred way. If the same effect counts for the way a detection is displayed(the operator may want a red border across the entire display, instead of a marking of the specific detected instance), and for many other decisions where his input was asked, it is not sure if it was wise to involve him in the first place. Not only does the operator have to work with a system that is different from what he would have liked, but it can also negatively affect his job satisfaction and overall trust in the organization since his views and experiences were not implemented.

For this research, the survey results and information gathered from the literature study are compliant. However, not all respondents will feel heard. In this stage, that hasn't been a problem, but it might in a later stadium.

6.2. Results

When looking at the results presented in this research, several things need to be taken into account with respect to the research questions and the generalization of the results.

When it comes to generalizing the results in this report, it should be kept in mind that this research is based on the remotely controlled bridges of Zaandam, and the context related to these assets. This means that the literature on bridge operations, the most significant user interviews and observations, and the data used for testing the object detection model, were all context-oriented. Although there will probably be quite some overlap between the operations in Zaanstad and the way it is done in other organizations, the extent to which these processes are similar is not clear. Also, as the camera systems used in other organizations could be different from those in Zaanstad, making it unsure how to translate the indicated potential for object detection on remotely controlled bridges from the setting in this report, to other organizations. However, in this report different image qualities have been used successfully in the test cases, increasing the generalisation chances.

Besides the operational and camera aspects, this also counts for the organizational setting. From an organizational point of view for example, many organisations use temporary workers to operate the bridges. It is not apparent whether this leads to different choices in the design process. It could be possible that this temporary worker has less experience working with the operating system than someone with years of experience with the same system. Besides, the temporary operator can be less familiar with the bridge design, and the way the monitors display the situation on it. This could have a negative influence on the response time in case of an unexpected occasion.

Notifying an experienced operator in case of an urgent occasion makes him perceive, decide and respond on that signal in a very quick and effective way, as his situational awareness is high. In case of working with a temporary worker as described before, the response mechanism could be slower and less effective. It may be worth reconsidering the support system's permissions as it could be safer to let the support system interrupt the process, instead of waiting for the operator to respond. In case of an occasional wrong call. Again, this scenario is based on assumptions, but displays how generalizing the result without further research could be dangerous.

Considering the results as obtained in this research, possibly one of the most important things to keep in mind when going through them, is that they are the outcome of experiments in a controlled environment. These results are experimental insights, indicating the possibilities of using an object detection support system. The hit rates achieved while running the model on the chosen video clips can provide no guarantees for the actual accuracy of the model in a real-life scenario, when implemented. Also, the model is trained on a very limited training set, of approximately 100 images. For a production-ready model, this should be way bigger. Theoretically, this would also improve the generalization abilities of the model, although those are not examined in this research.

The goal of this report was to cover both the human-machine interaction side, and the object detection field. The latter is evaluated by making use of experimental insights, yet it is difficult to evaluate the decisions based on human-machine interaction, besides using the validation scenarios. Testing a working system with actual operators is needed to show if the interaction between the system and the operator works as intended.

From a safety point of view, in the hypothetical situation of implementing the object detection support system in Zaandam, it is difficult to evaluate its effectiveness on the asset's safety. Because of the attention both bridges have gotten after the accidents, operators could be more concentrated than before. Furthermore, the extra measures that were taken by painting the bridge decks yellow, and by installing emergency buttons at some bridges for bystanders to notify on urgent occasions, make it difficult to compare the effectiveness of one single measure, in this case being the object detection system.

What also should be kept in mind when reading this report, is the fact that this analysis took the current situation and infrastructure into account. Possibly, adding night vision cameras or radar technologies could improve the safety at the bridge further. This should be considered in further research.

Finally, another element that could be further researched, is the implementation of a background trainer for the CNN. The CNN architecture should improve with a bigger and more complex dataset, and these images could be gathered throughout the usage of the system. Training the model on the newly captures images, can improve the system's performance significantly.

6.3. Future Recommendations

Taken the limitations and the problems that were left outside of the scope, these future recommendations should be considered, in no particular order of significance.

- Investigate 3D object tracking model, using the existing static cameras.
- Investigate different technologies like radar, night vision, and geo-tracking.
- Investigate measures to reduce image degradation caused by external factors.
- Investigate the human behaviour of road and maritime traffic with respect to remotely controlled bridges.

7. Conclusion

This research is done in reaction to the reports published by the Dutch Safety Board on the accidents that happened in 2015 en 2018 in Zaandam. These reports indicated that in both accidents, the victims were visible on the camera systems, yet through errors caused by human factors, were not detected. Based on these researches, the Dutch Safety Board concluded that the safety of the remotely controlled bridges was not sufficient, as the safety policy was mainly focussed on the technical aspects, instead of approaching it as an integral challenge where the interaction between human, machine, and environment should be considered. In this research, the possibility of mitigating these human factor base by using object detection through deep learning was investigated, with the main research question being:

How can Object Detection provide Decision Support for Mitigating Human Factors for Operating Remotely Controlled Bridges?

When looking at the current situation, the operators follow a strict operational workflow as depicted in figure 6, with more detailed information on what the operator should observe, control and monitor per stage in appendix D. Also, the operators manual describes which forms of traffic should be prioritized, and in which conditions safe operation can't be guaranteed, and how to act upon that. From observations and interviews, it becomes clear that the operators are well aware of these procedures and guidelines, in both common and anomalous situations, indicating the human factors are not connected to a lack of knowledge.

However, in practice it can be challenging to judge situations correctly, and with that choosing the appropriate reaction. Reduced image quality caused by external factors like rain and dirt, concentration problems and the complexity to watch multiple monitors simultaneously, all decrease the operator's situational awareness by limiting his capacity of perceiving bridge users. A phenomenon strongly related to this, is change blindness, as the operator could easily miss out on the motion signals in case someone enters the screen unnoticed, but stops moving to await the bridge opening.

These perceptual difficulties are the most important problems the proof-of-concept needs to mitigate, as the operators are capable of making the right decisions as long as the understanding of the situation at the bridge is correct. In practice this means that the model should accurately detect bridge users that cause a threat to the safe operation of the asset, while leaving enough time for the operators to react on these detections, and avert the risk.

The risk of adding the object detection application is that it could further increase the complexity of the system, resulting in it being another distraction instead of a support tool. A well designed Human-machine interface should avoid this, by maximizing the operator's understanding of the system. This translates into clearly annotating the detected elements, which speeds up the operator's comprehension, but also helps to understand the system's confusion, in case of false detection. 5.2.2.

For the model itself, a Mask R-CNN Resnet-101 was trained on a custom dataset containing images of CCTV footage from bridges in Zaanstad. Experimental insights showed that the custom trained model outperforms the pre-trained model, as depicted in subsection 5.3.1. In all experiments, no user was undetected in any 5 seconds period. This, together with a high hit-rate per image, shows the promising possibilities of using object detection for mitigating human factors, as the increased perception of elements will enhance the operator's situational awareness, and therefore increase the safety of bridge operation. Although the interviewed operators indicated that a situation with object detection instinctively feels safer, further research is needed to get a more quantitative insight into the safety improvements.

Appendices

A. Methodology

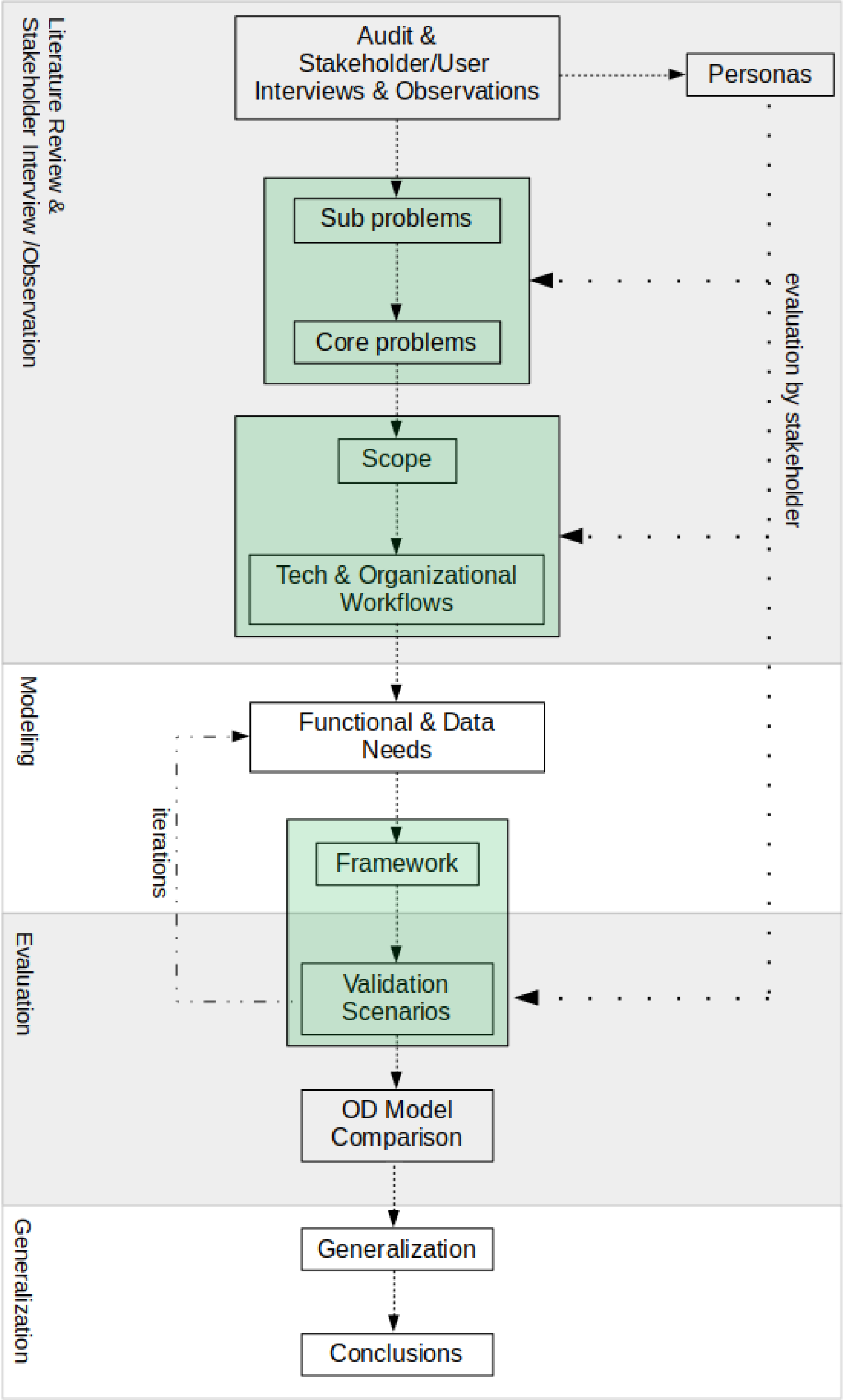


Figure 53.: *Actions per Research Phase*

A.1. Action Design Research

Stage 1: Problem Formulation

Stage Description
A problem perceived in practice or anticipated by the researchers is the trigger for the problem formulation stage, and is the general starting point for the research endeavour. The input for this stage can come from practitioners, end-users, researchers, existing technologies, and/or review of prior research. This stage contains of an initial empirical investigation of the problem, determining the inital scope, deciding the roles and scope for the practitioner participation, and the initial research question are formulated(Sein et al., 2011).

Principle 1: Practice-Inspired Research
This principle focuses on ADR's aim to view field problems as knowledge-creation opportunities, opposed to theoretical challenges. ADR seeks these opportunities at the intersection of technological and organizational domains, with possible emphasis on one of the two domains. The researcher should obtain enough knowledge about each domain, to be able to come up with a holistic solution(Sein et al., 2011).
In the research that will be conducted in this report, this principle is represented by the combination of the technical infrastructure of the camera and operating system, with the organizational challenge of working with bridge operators who have to be able to work focussed in stressful conditions for long periods of time. The solution for the perceived problem should take both domains into account to work well.

Principle 2: Theory-Ingrained Artifact
This principle emphasis that ensembled artifacts created and evaluated using the ADR methodology, are informed by theory(Sein et al., 2011). In this process, theory is regarded as systems of statements that allow generalization and abstraction(Gregor, 2006). Three uses of prior theories are acknowledged in this principle:(i) To structure the problem, (ii) to identify solutions, (iii) and to guide design. The theoretical framework in the prior chapter serves to get a clear overview of the different subdomains, to decompose the perceived problem, and to model an ensembled artefact to solve the core obstacles. This knowledge is subsequently complemented and validated by stakeholder interviews and observations.

Stage 2: Building, Intervention, and Evaluation

Stage Description
The second stage builds on the problem framing and theoretical premises from first stage, by using it as a platform for generating the initial design of the IT artifact. Through organizational use and subsequent design cycles, the artifact will be modified and updated. The process is iterative, and it is carried out in the target environment. Here, the building of the artifact takes place, the intervention in the organizaton, and the evaluation of the ensemble. The outcome of this BIE stage should be the realized design of the artifact(Sein et al., 2011).

For the research on safety of remotely controlled bridges, this translates into the phase of building prototypes and mock-ups of the object detection system, and testing them on both the technical level by analyzing the accuracy in different circumstances, and the way of interacting with the artifact by the organization. Starting off with an unpolished version in an early stage, the feedback of a subset of the organization will gradually mould the system into a well integrated solution for mitigating human factors in he bridge operation process.

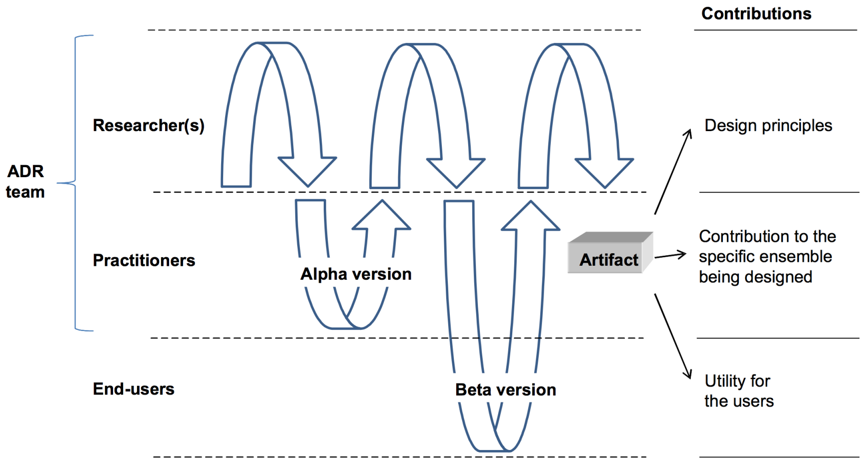


Figure 54.: *The Generic Schema for IT-Dominant BIE (Sein et al., 2011)*

Principle 3: Reciprocal Shaping

This principle focusses on the inseparable influences mutually exerted by both the IT artifact and the organization context. The ADR team performs an iterative process of going into the fine details of each domain to get a better understanding, to go back to the overview to see how that details affects the whole (Sein et al., 2011).

Principle 4: Mutually Influential Roles

This focusses on the importance of mutual learning among the different project participants. One one hand, the researcher brings knowledge of theory an technological advances to the table, and on the other hand are the practitioners that bring knowledge of organizational work practices. (Sein et al., 2011)

Principle 5: Authentic and Concurrent Evaluation

This principle emphasizes a key characteristic of Action Design Reseach; evaluation is not just a final stage after developing a product, but is interwoven into the design process(Sein et al., 2011). By constantly reflecting with the bridge operators and managers, design choices will be evaluated during the process, and taken into account by reshaping the object detection system.

Stage 3: Reflection and Learning

Stage Description

The third phase moves conceptually from building a solution for one particular instance, to see if the learnings can be applied to a broader class of problems. This is a continuous effort, which parallels the first two stages, as can be seen in figure 29 . Conscious reflection on the problem formulation, the chosen theories and the emerging ensemble is critical to ensure the identification of contributions to knowledge (Sein et al., 2011).

Principle 6: Guided Emergence

This principle captures the interplay between two seemingly conflicting perspectives. On the one hand there is design, which implies external, intentional intervention, and the other hand there is emergence, which hints at organic evolution. Guided emergence emphasizes that the ensemble artefact will not only reflect the preliminary design, as designed by the researchers (Principle 2), but also its ongoing shaping by organizational use, participants, perspectives and outcomes of authentic, concurrent evaluation (Principle 5) (Sein et al., 2011).

Stage 4: Formalization of Learning

Stage Description

The last stage is about formalizing the learning. The ongoing learnings gathered in the last steps, should be further developed into general solution concepts for a class of field problems. Reseachers outline the accomplishments realized in the artifact, and describe the organizational outcomes. These outcomes can be seen as the design principles learned in the project, as refinements to the theories that contributed to the initial design (Sein et al., 2011).

Principle 7: Generalized Outcomes

Generalization is challenging because the outcomes of the ADR embody both organizational change, along with implementation of an artifact. The ensembled artifact represents a solution that addresses a problem, and both of these elements, can be generalized. Three steps of conceptual generalization are proposed: (1) Generalizaton of the problem instance, (2) generalization of the solution instance, and (3) derivation of design principles form the design research outcomes (Sein et al., 2011).

A.2. Goal-Directed Research

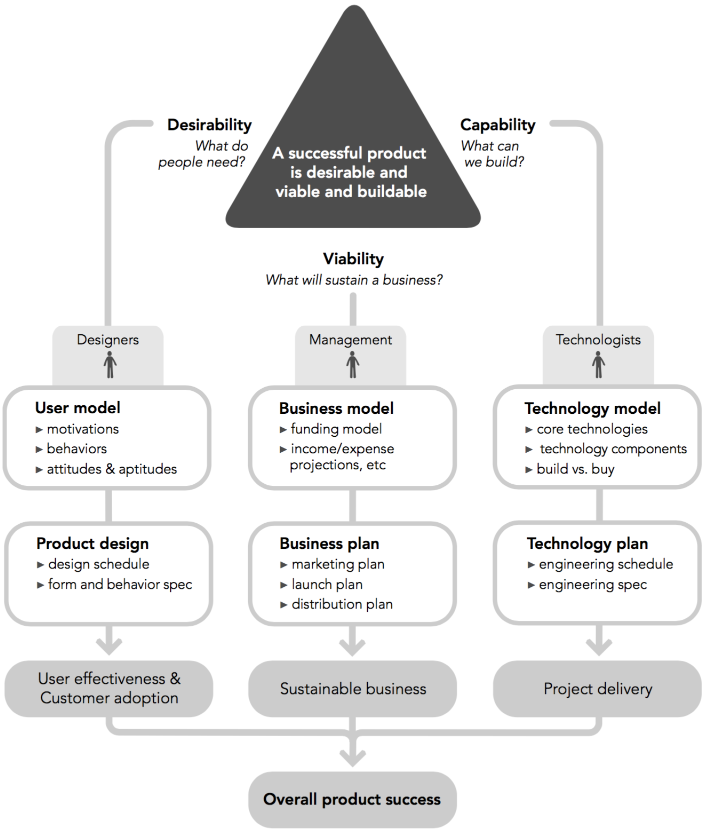
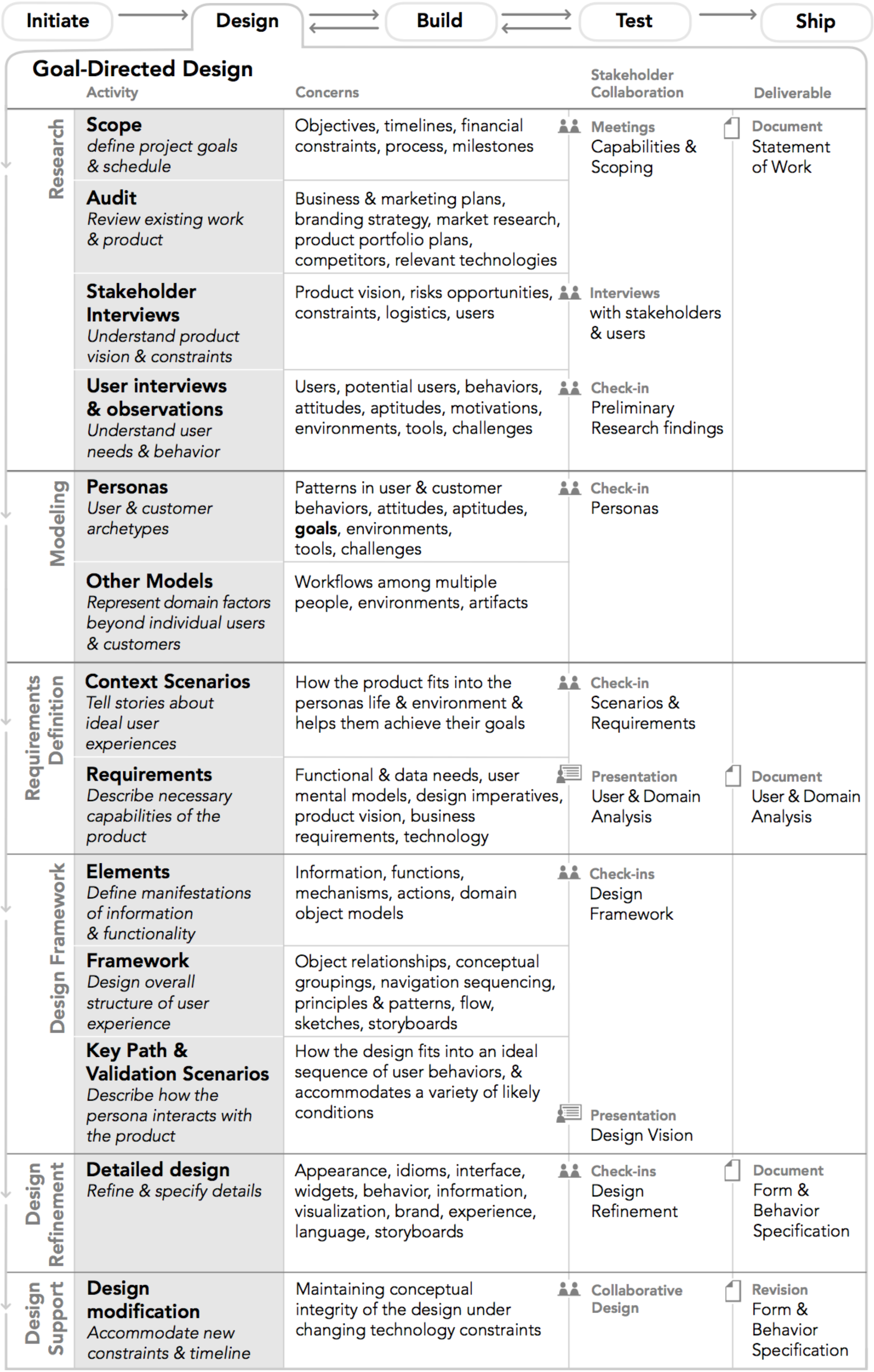


Figure 55.: *Modern Triad of Product Development (Cooper et al., 2007, pg.12)*



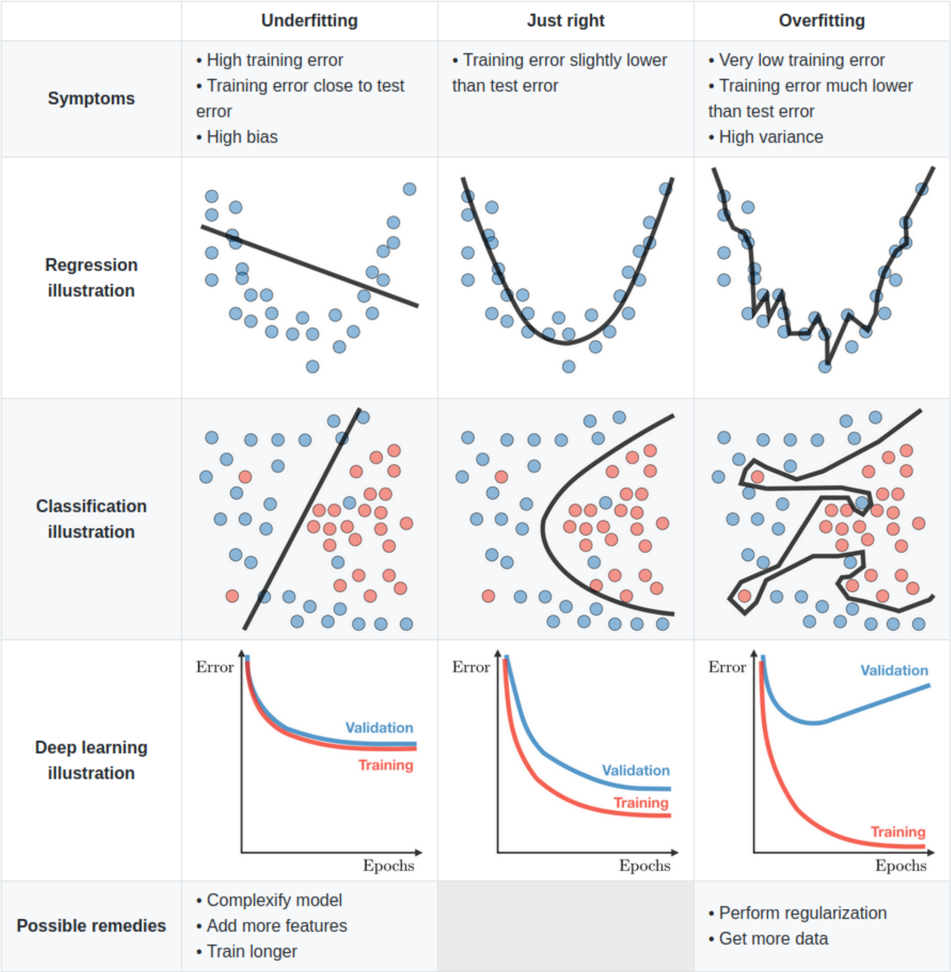
Initiate → Design ⇄ Build ⇄ Test → Ship

Goal-Directed Design

	Activity	Concerns	Stakeholder Collaboration	Deliverable
Research	**Scope** *define project goals & schedule*	Objectives, timelines, financial constraints, process, milestones	Meetings Capabilities & Scoping	Document Statement of Work
	Audit *Review existing work & product*	Business & marketing plans, branding strategy, market research, product portfolio plans, competitors, relevant technologies		
	Stakeholder Interviews *Understand product vision & constraints*	Product vision, risks opportunities, constraints, logistics, users	Interviews with stakeholders & users	
	User interviews & observations *Understand user needs & behavior*	Users, potential users, behaviors, attitudes, aptitudes, motivations, environments, tools, challenges	Check-in Preliminary Research findings	
Modeling	**Personas** *User & customer archetypes*	Patterns in user & customer behaviors, attitudes, aptitudes, **goals**, environments, tools, challenges	Check-in Personas	
	Other Models *Represent domain factors beyond individual users & customers*	Workflows among multiple people, environments, artifacts		
Requirements Definition	**Context Scenarios** *Tell stories about ideal user experiences*	How the product fits into the personas life & environment & helps them achieve their goals	Check-in Scenarios & Requirements	
	Requirements *Describe necessary capabilities of the product*	Functional & data needs, user mental models, design imperatives, product vision, business requirements, technology	Presentation User & Domain Analysis	Document User & Domain Analysis
Design Framework	**Elements** *Define manifestations of information & functionality*	Information, functions, mechanisms, actions, domain object models	Check-ins Design Framework	
	Framework *Design overall structure of user experience*	Object relationships, conceptual groupings, navigation sequencing, principles & patterns, flow, sketches, storyboards		
	Key Path & Validation Scenarios *Describe how the persona interacts with the product*	How the design fits into an ideal sequence of user behaviors, & accommodates a variety of likely conditions	Presentation Design Vision	
Design Refinement	**Detailed design** *Refine & specify details*	Appearance, idioms, interface, widgets, behavior, information, visualization, brand, experience, language, storyboards	Check-ins Design Refinement	Document Form & Behavior Specification
Design Support	**Design modification** *Accommodate new constraints & timeline*	Maintaining conceptual integrity of the design under changing technology constraints	Collaborative Design	Revision Form & Behavior Specification

Figure 56.: *Goal Oriented Design Process (Cooper et al., 2007, pg.24)*

B. Object Detection

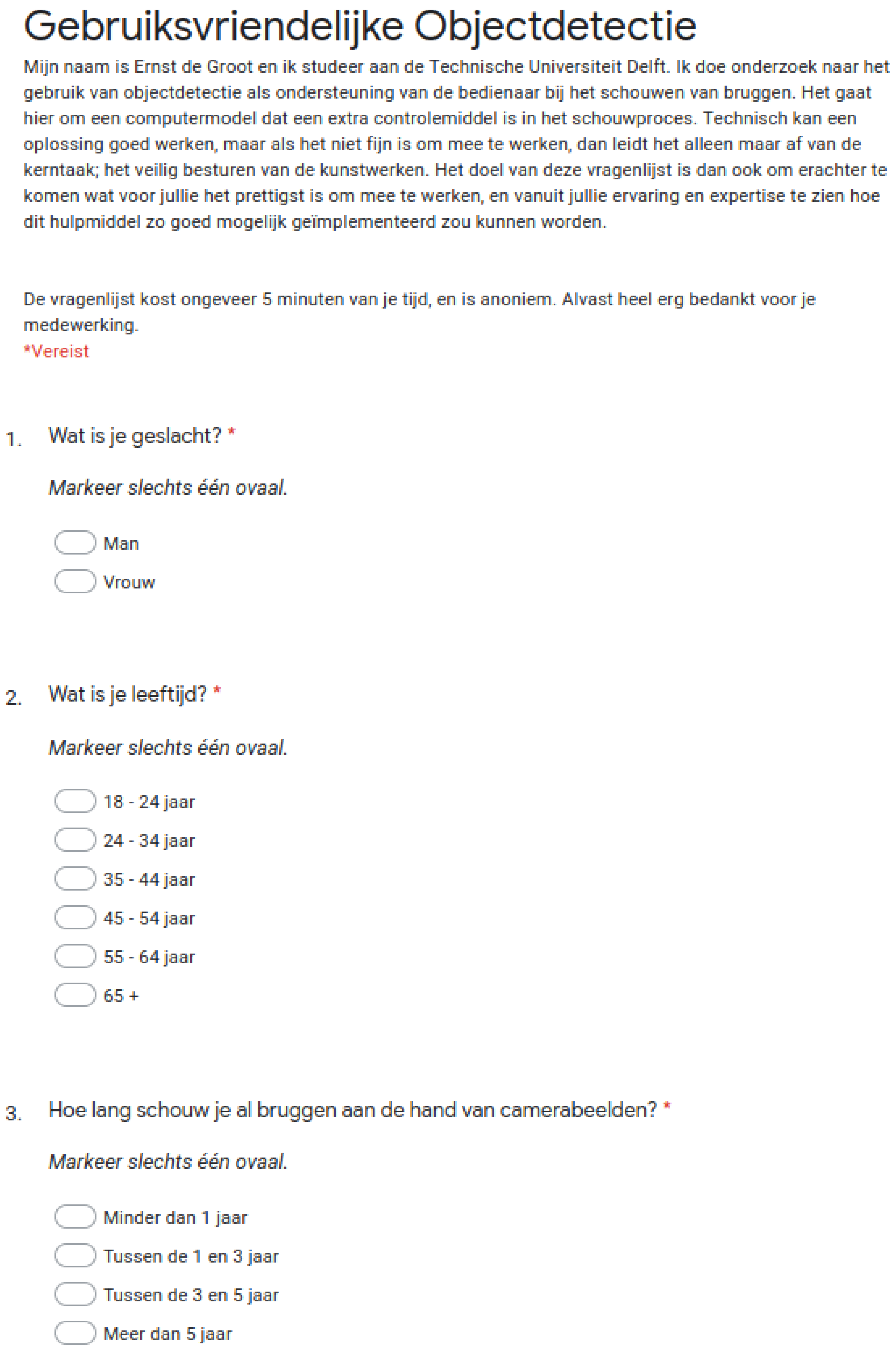


	Underfitting	Just right	Overfitting
Symptoms	• High training error • Training error close to test error • High bias	• Training error slightly lower than test error	• Very low training error • Training error much lower than test error • High variance
Regression illustration			
Classification illustration			
Deep learning illustration			
Possible remedies	• Complexify model • Add more features • Train longer		• Perform regularization • Get more data

Figure 57.: *Overfitting and Underfitting Explained (Amidi & Amidi, 2018)*

C. Survey

C.1. Questions



Gebruiksvriendelijke Objectdetectie

Mijn naam is Ernst de Groot en ik studeer aan de Technische Universiteit Delft. Ik doe onderzoek naar het gebruik van objectdetectie als ondersteuning van de bedienaar bij het schouwen van bruggen. Het gaat hier om een computermodel dat een extra controlemiddel is in het schouwproces. Technisch kan een oplossing goed werken, maar als het niet fijn is om mee te werken, dan leidt het alleen maar af van de kerntaak; het veilig besturen van de kunstwerken. Het doel van deze vragenlijst is dan ook om erachter te komen wat voor jullie het prettigst is om mee te werken, en vanuit jullie ervaring en expertise te zien hoe dit hulpmiddel zo goed mogelijk geïmplementeerd zou kunnen worden.

De vragenlijst kost ongeveer 5 minuten van je tijd, en is anoniem. Alvast heel erg bedankt voor je medewerking.

*Vereist

1. Wat is je geslacht? *

Markeer slechts één ovaal.

○ Man
○ Vrouw

2. Wat is je leeftijd? *

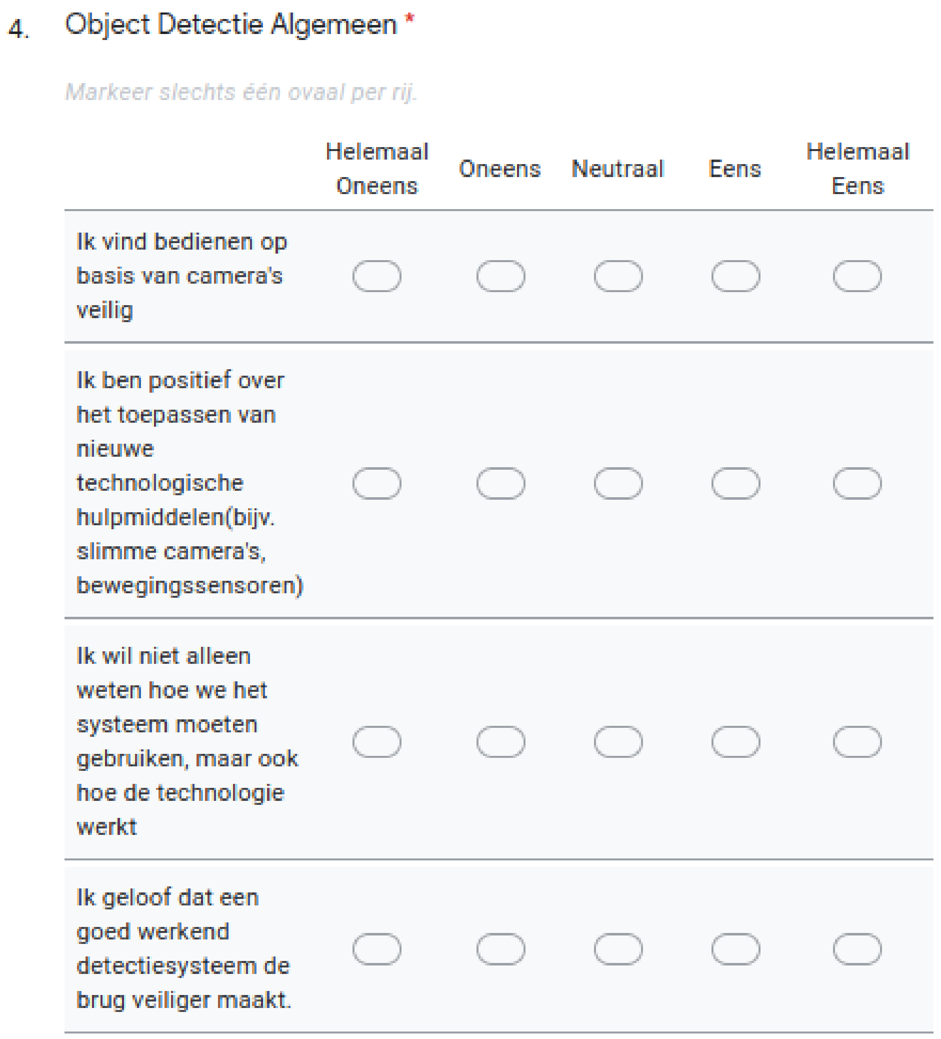
Markeer slechts één ovaal.

○ 18 - 24 jaar
○ 24 - 34 jaar
○ 35 - 44 jaar
○ 45 - 54 jaar
○ 55 - 64 jaar
○ 65 +

3. Hoe lang schouw je al bruggen aan de hand van camerabeelden? *

Markeer slechts één ovaal.

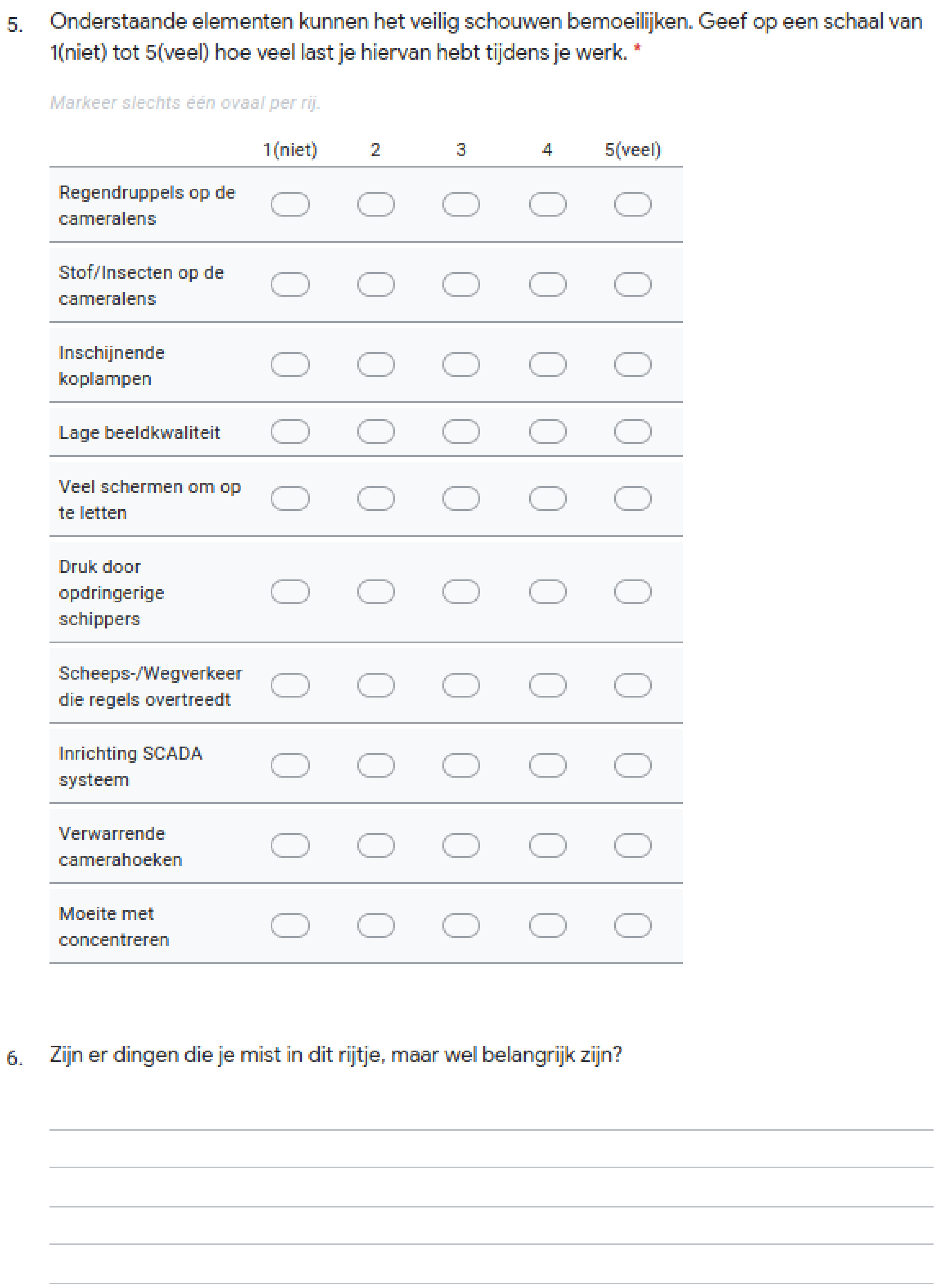
○ Minder dan 1 jaar
○ Tussen de 1 en 3 jaar
○ Tussen de 3 en 5 jaar
○ Meer dan 5 jaar



4. Object Detectie Algemeen *

Markeer slechts één ovaal per rij.

	Helemaal Oneens	Oneens	Neutraal	Eens	Helemaal Eens
Ik vind bedienen op basis van camera's veilig	○	○	○	○	○
Ik ben positief over het toepassen van nieuwe technologische hulpmiddelen(bijv. slimme camera's, bewegingssensoren)	○	○	○	○	○
Ik wil niet alleen weten hoe we het systeem moeten gebruiken, maar ook hoe de technologie werkt	○	○	○	○	○
Ik geloof dat een goed werkend detectiesysteem de brug veiliger maakt.	○	○	○	○	○



5. Onderstaande elementen kunnen het veilig schouwen bemoeilijken. Geef op een schaal van 1(niet) tot 5(veel) hoe veel last je hiervan hebt tijdens je werk. *

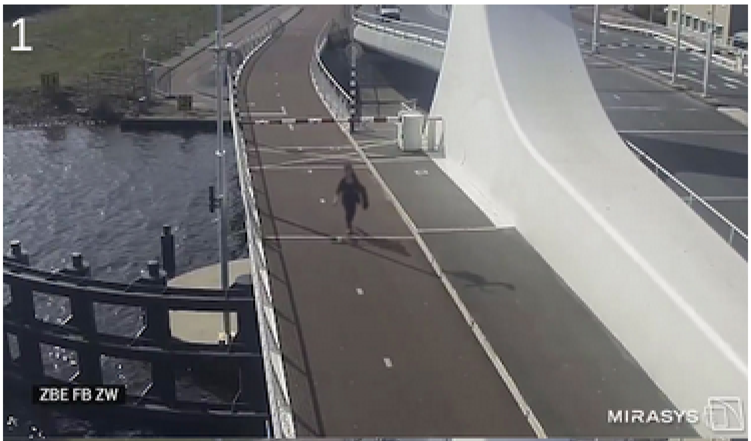
Markeer slechts één ovaal per rij.

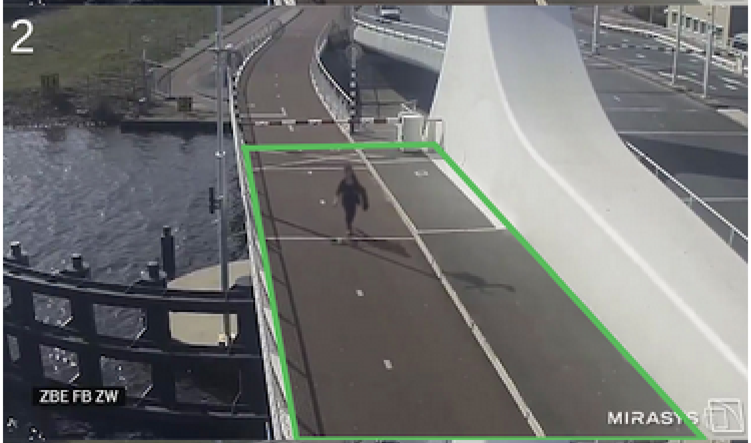
	1(niet)	2	3	4	5(veel)
Regendruppels op de cameralens	○	○	○	○	○
Stof/Insecten op de cameralens	○	○	○	○	○
Inschijnende koplampen	○	○	○	○	○
Lage beeldkwaliteit	○	○	○	○	○
Veel schermen om op te letten	○	○	○	○	○
Druk door opdringerige schippers	○	○	○	○	○
Scheeps-/Wegverkeer die regels overtreedt	○	○	○	○	○
Inrichting SCADA systeem	○	○	○	○	○
Verwarrende camerahoeken	○	○	○	○	○
Moeite met concentreren	○	○	○	○	○

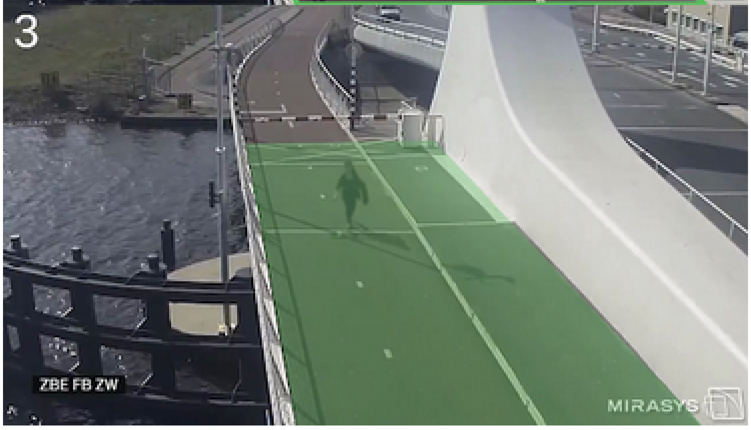
6. Zijn er dingen die je mist in dit rijtje, maar wel belangrijk zijn?

Detectiegebied

Vooraf wordt ingesteld welk gebied het detectiesysteem analyseert. Dit is op meerdere manieren weer te geven op de cameraschermen. Er kan gekozen worden het gebied niet zichtbaar te maken, te omlijnen, of er een gekleurd vlak over te leggen.







7. Welke methode heeft je voorkeur? *

Markeer slechts één ovaal.

○ Optie 1
○ Optie 2
○ Optie 3

Detectie Weergave

8. Waar zou je willen zien als er een detectie plaatsvindt? *

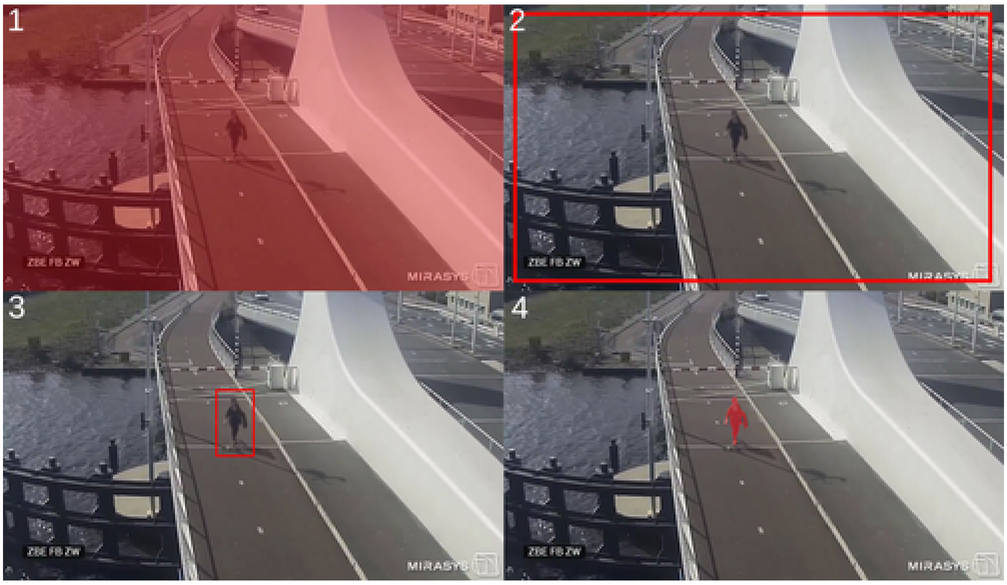
Markeer slechts één ovaal.

○ SCADA Systeem

○ Camerabeelden

○ Beide

Weergavemogelijkheden



9. Op welke manier zou je de detectie het liefst gepresenteerd zien op de camerabeelden? *

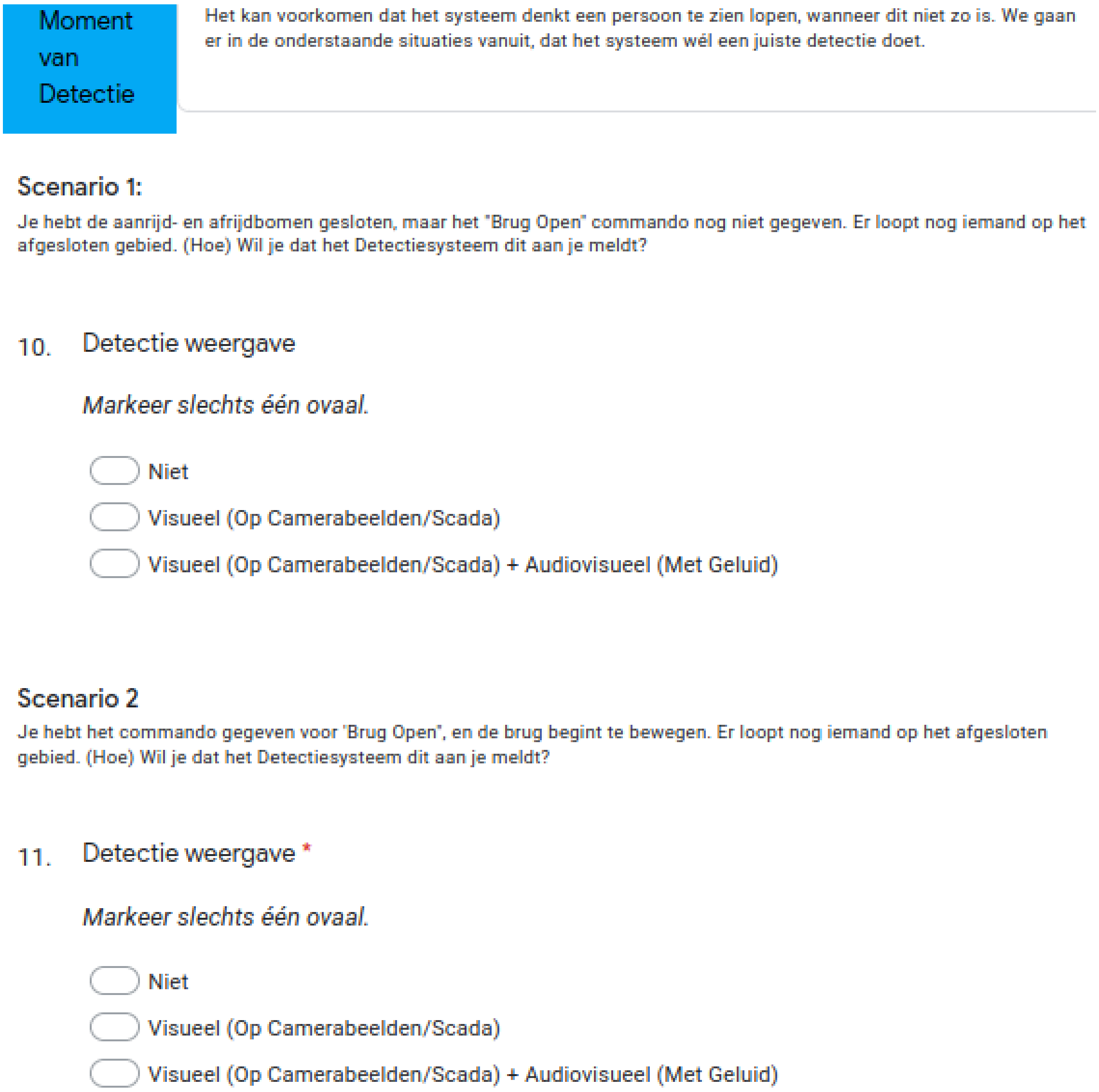
Markeer slechts één ovaal.

○ Optie 1

○ Optie 2

○ Optie 3

○ Optie 4



Moment van Detectie

Het kan voorkomen dat het systeem denkt een persoon te zien lopen, wanneer dit niet zo is. We gaan er in de onderstaande situaties vanuit, dat het systeem wél een juiste detectie doet.

Scenario 1:

Je hebt de aanrijd- en afrijdbomen gesloten, maar het "Brug Open" commando nog niet gegeven. Er loopt nog iemand op het afgesloten gebied. (Hoe) Wil je dat het Detectiesysteem dit aan je meldt?

10. Detectie weergave

Markeer slechts één ovaal.

○ Niet

○ Visueel (Op Camerabeelden/Scada)

○ Visueel (Op Camerabeelden/Scada) + Audiovisueel (Met Geluid)

Scenario 2

Je hebt het commando gegeven voor 'Brug Open", en de brug begint te bewegen. Er loopt nog iemand op het afgesloten gebied. (Hoe) Wil je dat het Detectiesysteem dit aan je meldt?

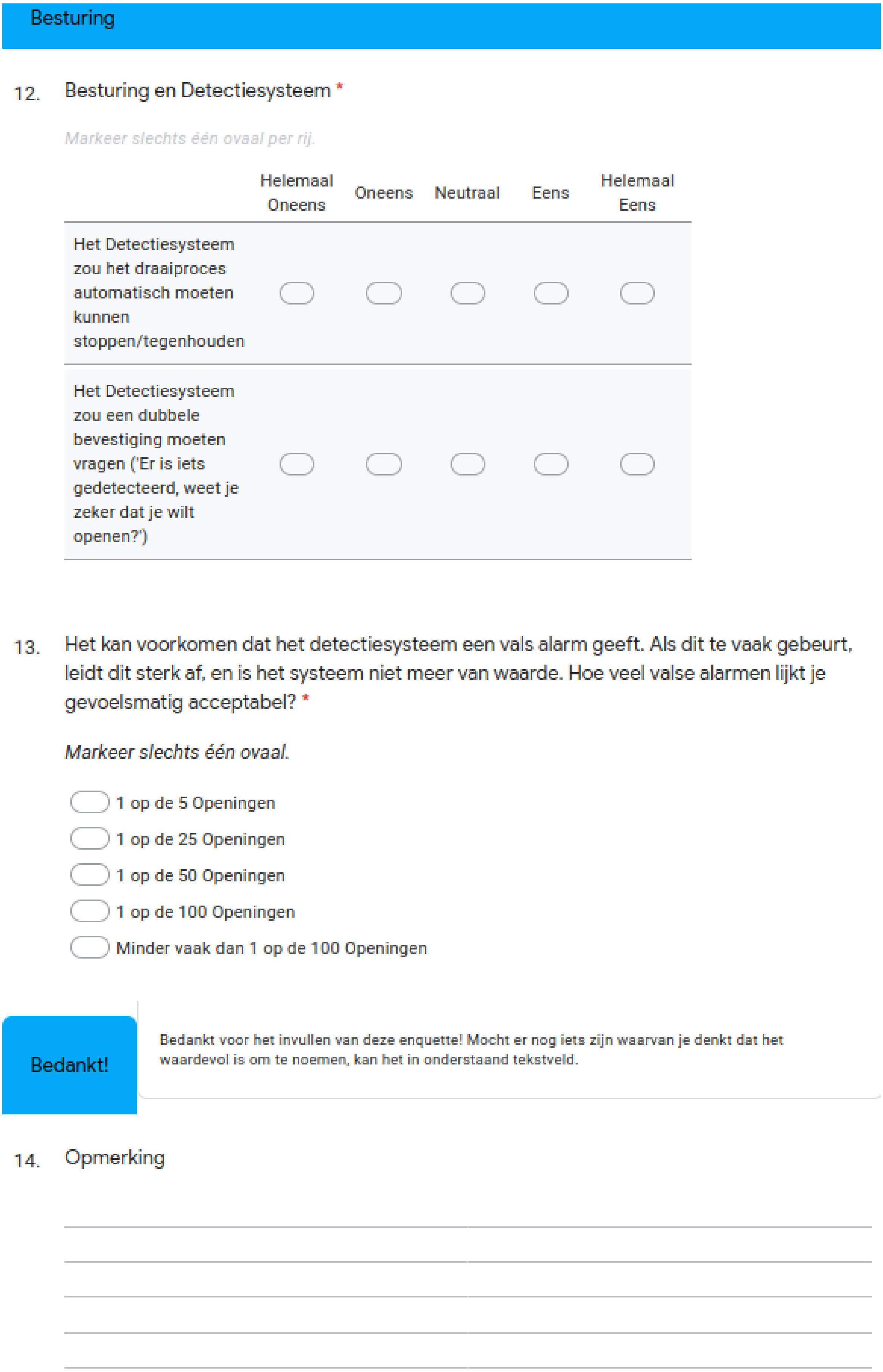
11. Detectie weergave *

Markeer slechts één ovaal.

○ Niet

○ Visueel (Op Camerabeelden/Scada)

○ Visueel (Op Camerabeelden/Scada) + Audiovisueel (Met Geluid)



Besturing

12. Besturing en Detectiesysteem *

Markeer slechts één ovaal per rij.

	Helemaal Oneens	Oneens	Neutraal	Eens	Helemaal Eens
Het Detectiesysteem zou het draaiproces automatisch moeten kunnen stoppen/tegenhouden	○	○	○	○	○
Het Detectiesysteem zou een dubbele bevestiging moeten vragen ('Er is iets gedetecteerd, weet je zeker dat je wilt openen?')	○	○	○	○	○

13. Het kan voorkomen dat het detectiesysteem een vals alarm geeft. Als dit te vaak gebeurt, leidt dit sterk af, en is het systeem niet meer van waarde. Hoe veel valse alarmen lijkt je gevoelsmatig acceptabel? *

Markeer slechts één ovaal.

○ 1 op de 5 Openingen

○ 1 op de 25 Openingen

○ 1 op de 50 Openingen

○ 1 op de 100 Openingen

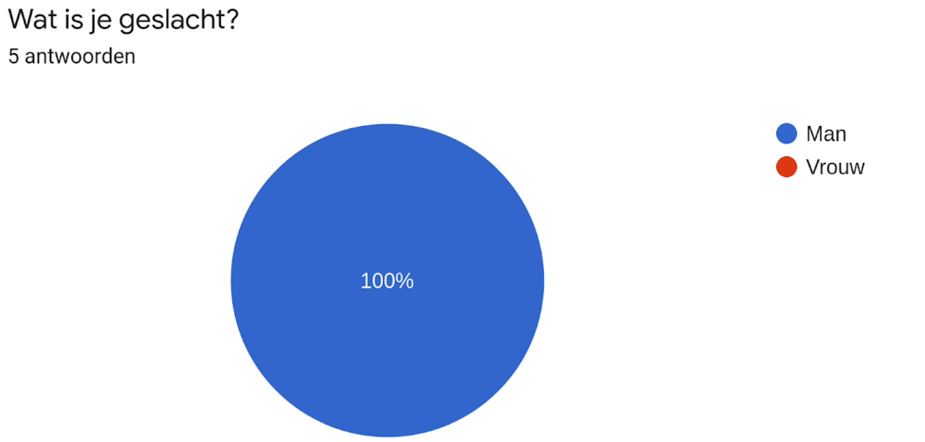
○ Minder vaak dan 1 op de 100 Openingen

Bedankt!

Bedankt voor het invullen van deze enquette! Mocht er nog iets zijn waarvan je denkt dat het waardevol is om te noemen, kan het in onderstaand tekstveld.

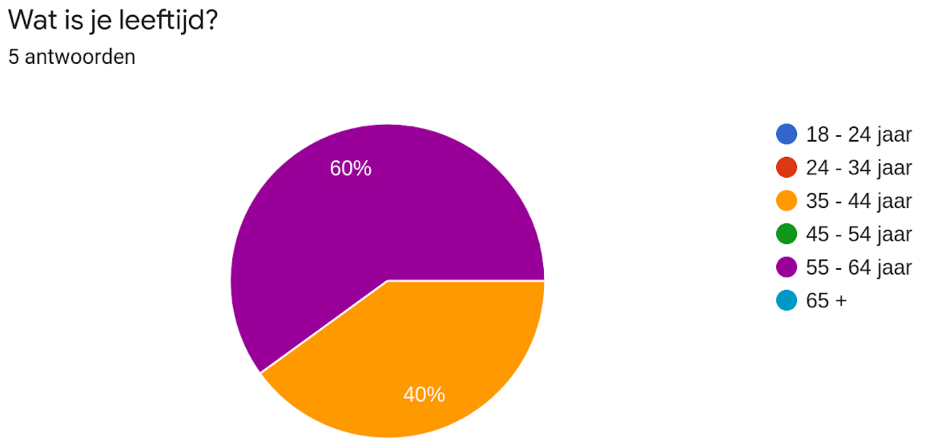
14. Opmerking

C.2. Answers



Wat is je geslacht?

5 antwoorden

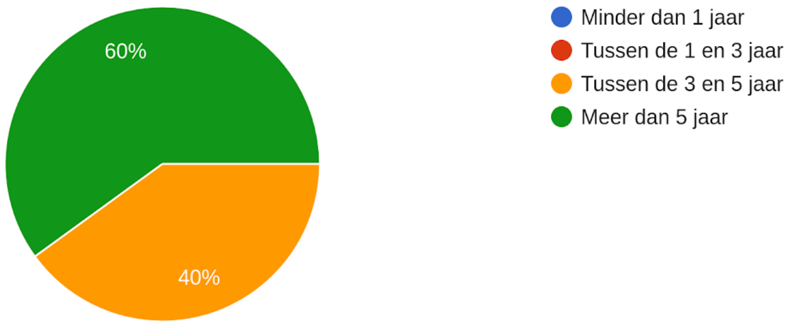


Wat is je leeftijd?

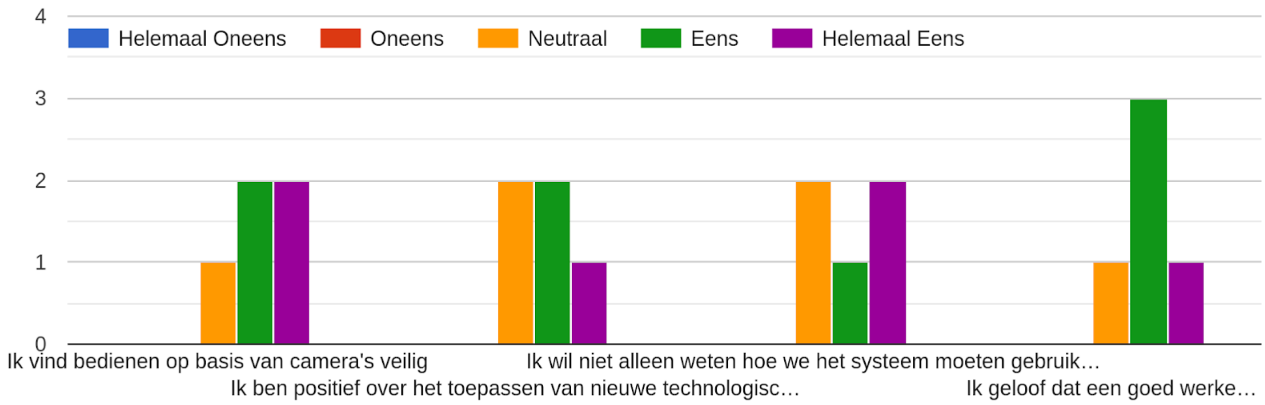
5 antwoorden

Hoe lang schouw je al bruggen aan de hand van camerabeelden?

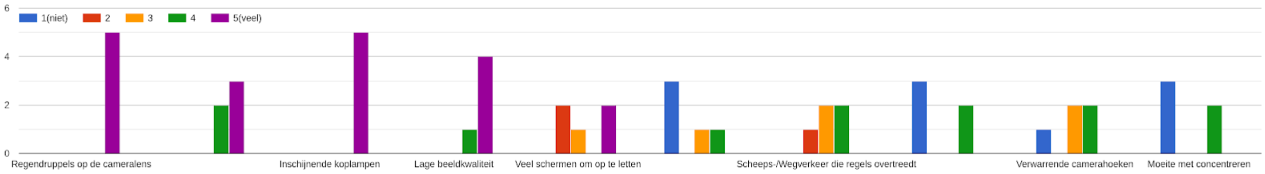
5 antwoorden



Object Detectie Algemeen

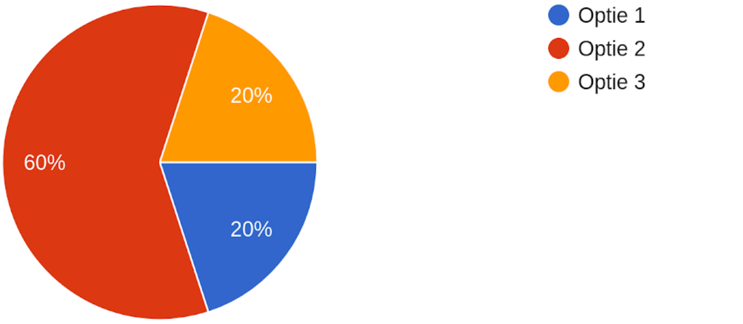


Onderstaande elementen kunnen het veilig schouwen bemoeilijken. Geef op een schaal van 1(niet) tot 5(veel) hoe veel last je hiervan hebt tijdens je werk.



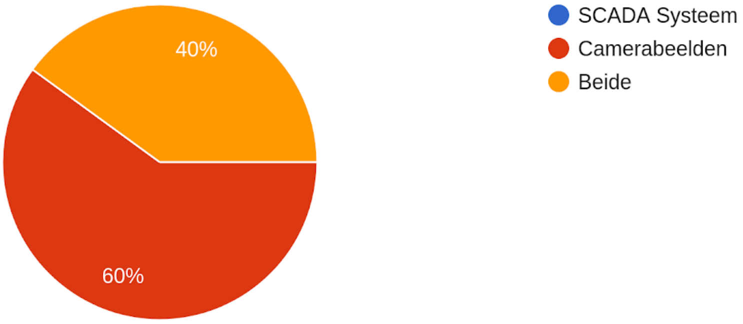
Welke methode heeft je voorkeur?

5 antwoorden



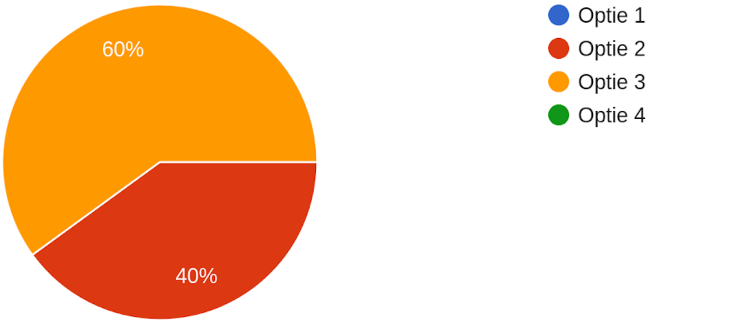
Waar zou je willen zien als er een detectie plaatsvindt?

5 antwoorden



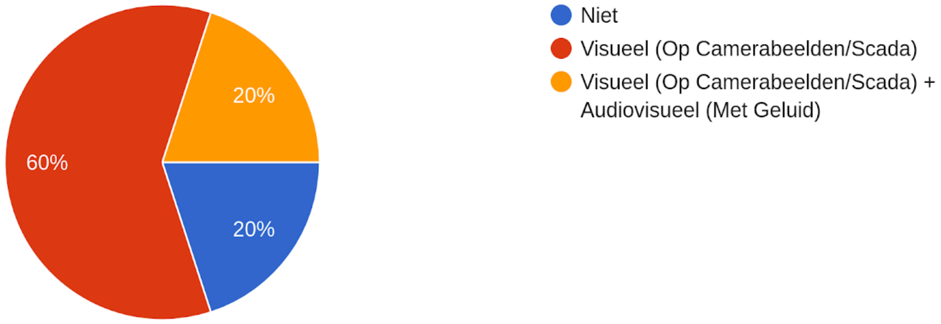
Op welke manier zou je de detectie het liefst gepresenteerd zien op de camerabeelden?

5 antwoorden



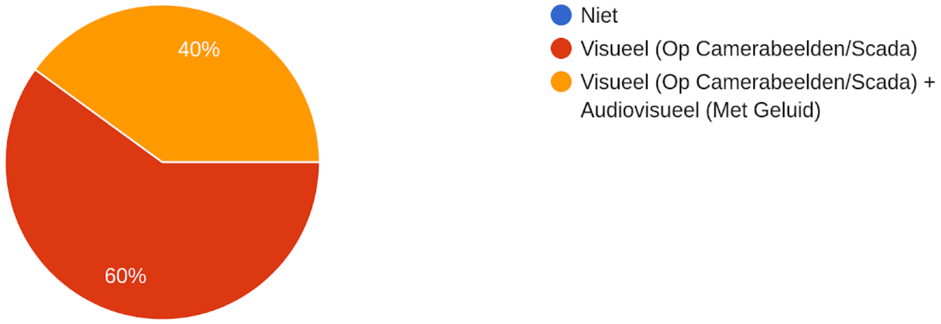
Detectie weergave

5 antwoorden

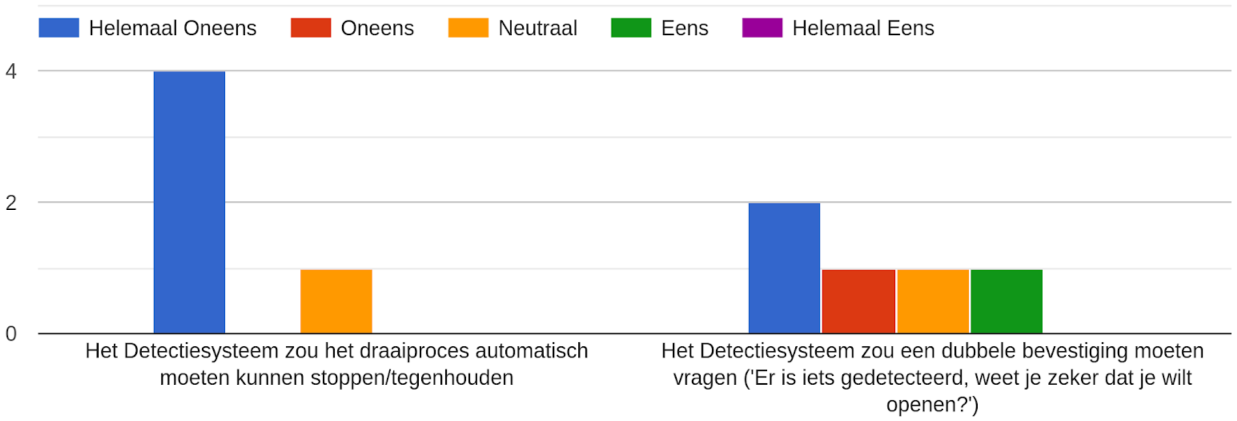


Detectie weergave

5 antwoorden

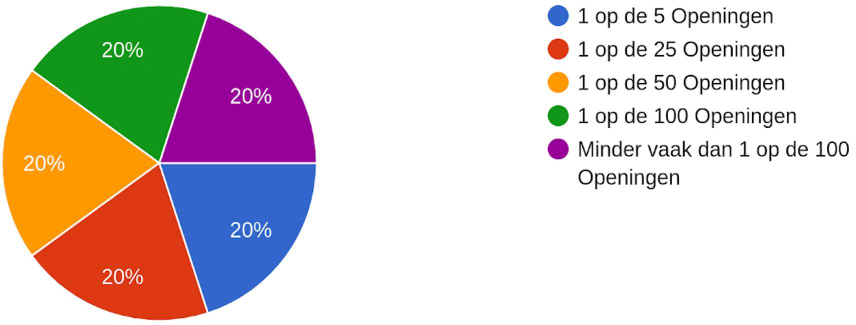


Besturing en Detectiesysteem



Het kan voorkomen dat het detectiesysteem een vals alarm geeft. Als dit te vaak gebeurt, leidt dit sterk af, en is het systeem niet meer van waarde. H...el valse alarmen lijkt je gevoelsmatig acceptabel?

5 antwoorden



Working Environment

When it comes down to the general comfort of the working environment, the average operator was pleased, with a small variance.

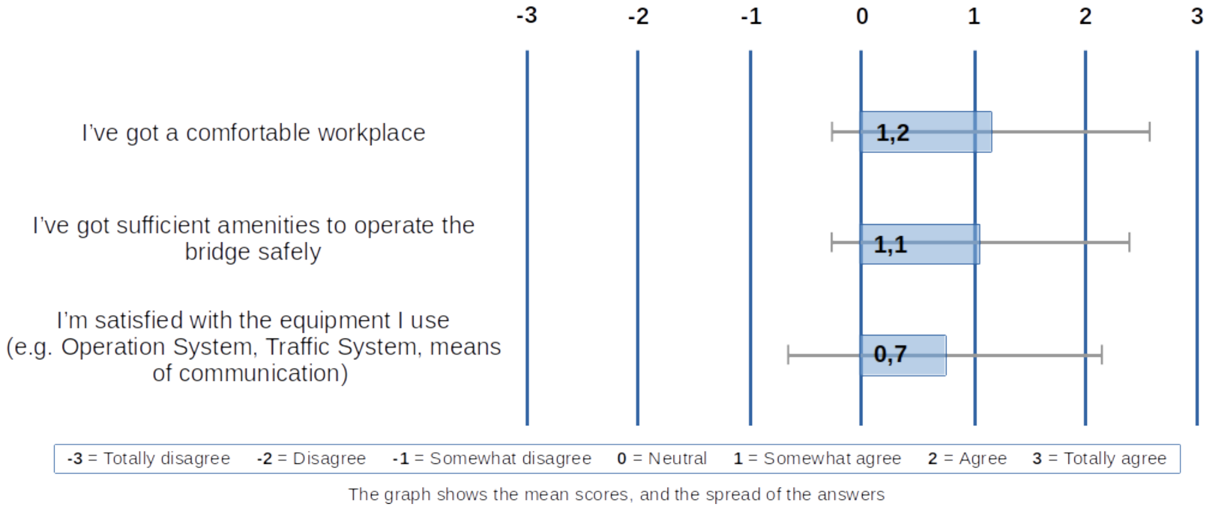


Figure 58.: *Survey Results: Workplace (van Veelen, 2018)*

Education and Competences

The second field is about education and competences.

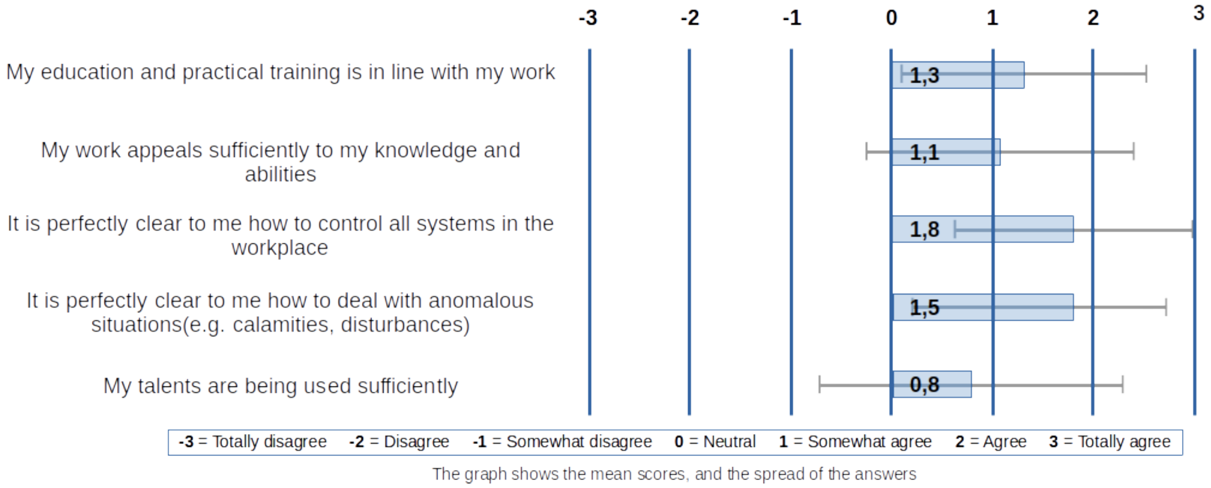


Figure 59.: *Survey Results: Education (van Veelen, 2018)*

Relevant Comments:

- Organization could pay more attention to the suggestions of the operators
- Use the operator's experience and skills. Don't limit him to button presser.
- Try to develop a simulator to train on incidents.

Cameras

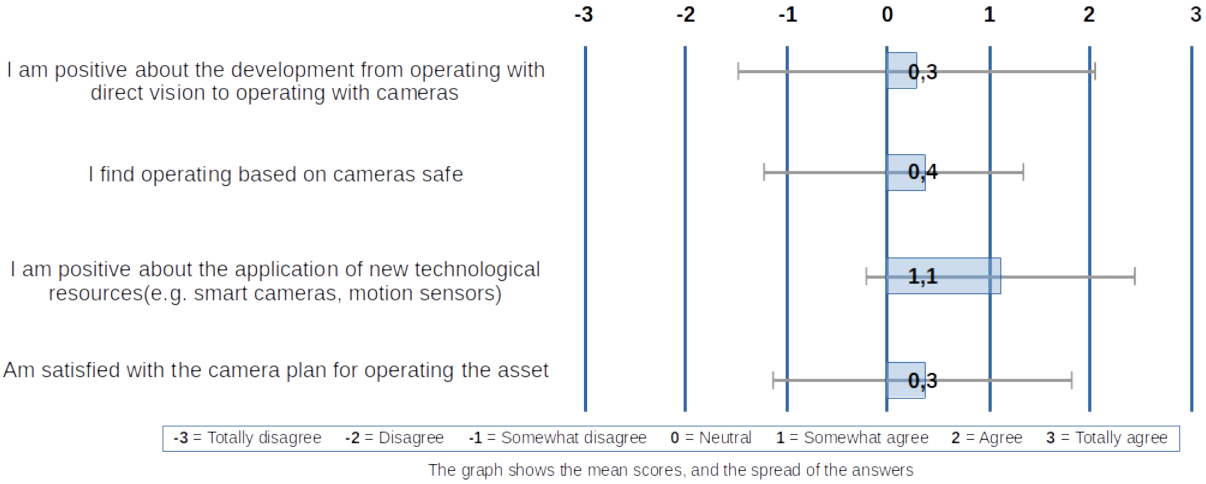


Figure 60.: *Survey Results: Camera Setup (van Veelen, 2018)*

Relevant Comments:

- Mist, darkness, and blinding sunlight/artificial light cause problems for the cameras.
- The application of new camera systems is slow.
- Involve experienced operators in setting up the camera plans.
- Knowledge about the local situation is important for controlling remotely.

Incidents

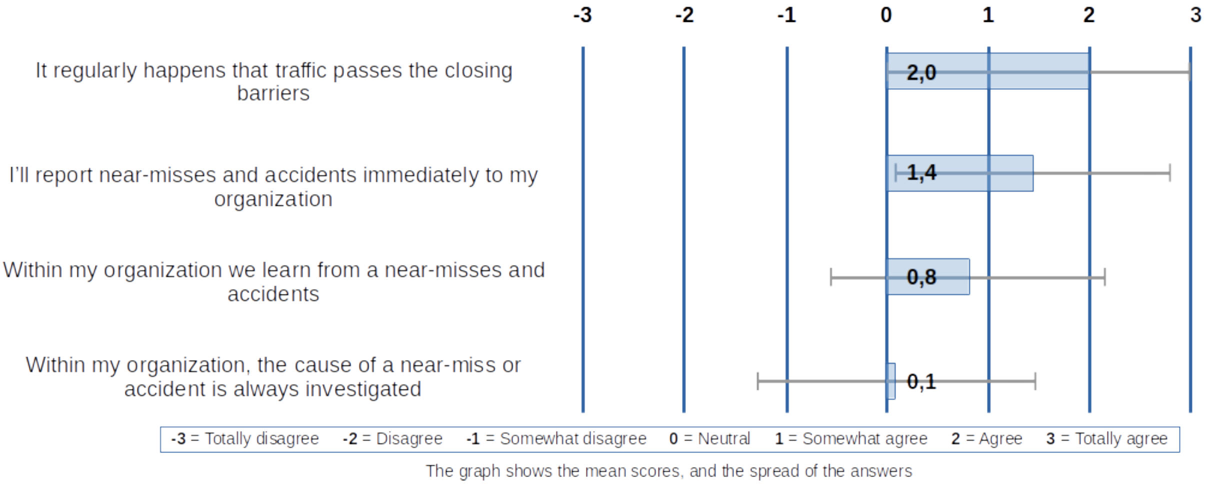


Figure 61.: *Survey Results: Incidents (van Veelen, 2018)*

Relevant Comments:

- Operators are blamed too quickly when an incident occurs, and no technical malfunction can be found.
- Controlling remotely misses social control at times of red light negation.

- There is fear of reporting incidents, because it can backfire on the operator.
- The analysis of the incidents is too seldom used, to prevent future incidents.
- Near-misses happen so frequently, it is impossible to keep up with them.

Organization

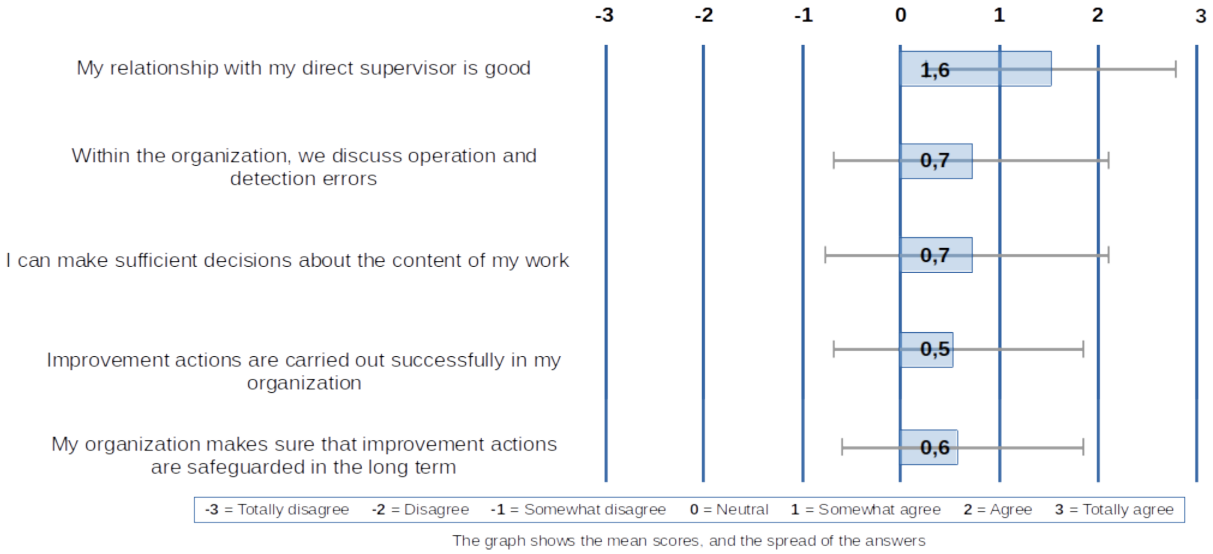


Figure 62.: *Survey Results: Organization (van Veelen, 2018)*

Relevant Comments:

- The workload is too high due to staff shortage.
- There operator gets limited feedback.
- More investments are done in technologies, than in people.

Safety

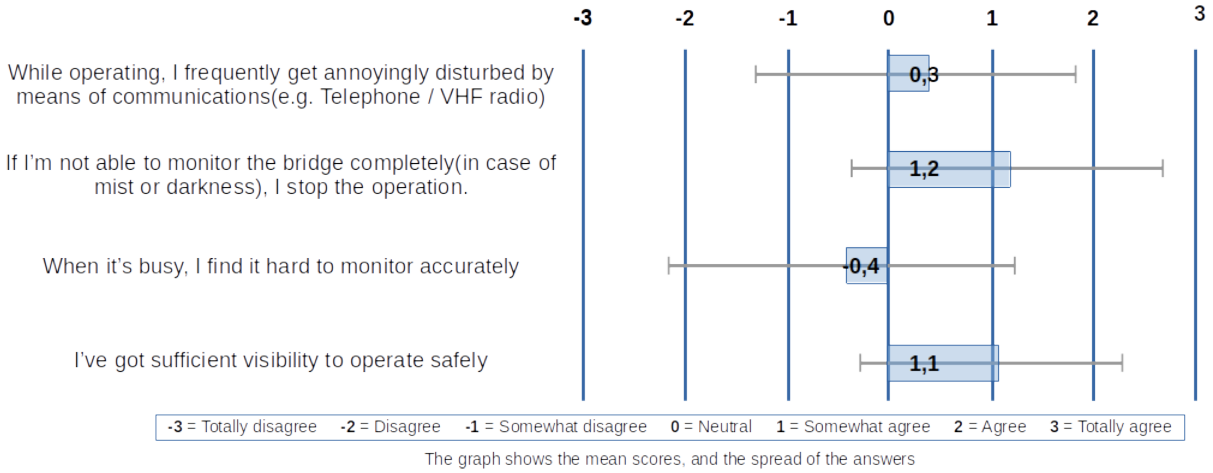


Figure 63.: *Survey Results: Safety (van Veelen, 2018)*

Relevant Comments:

- Operators feel pressured to continue the opening procedure, even if they actually don't think it's safe(e.g. poor visibility, malfunctioning cameras).
- Mist, darkness and blinding (sun)light cause unsafe situations, as it is difficult to monitor accurately.

Motivation & Commitment

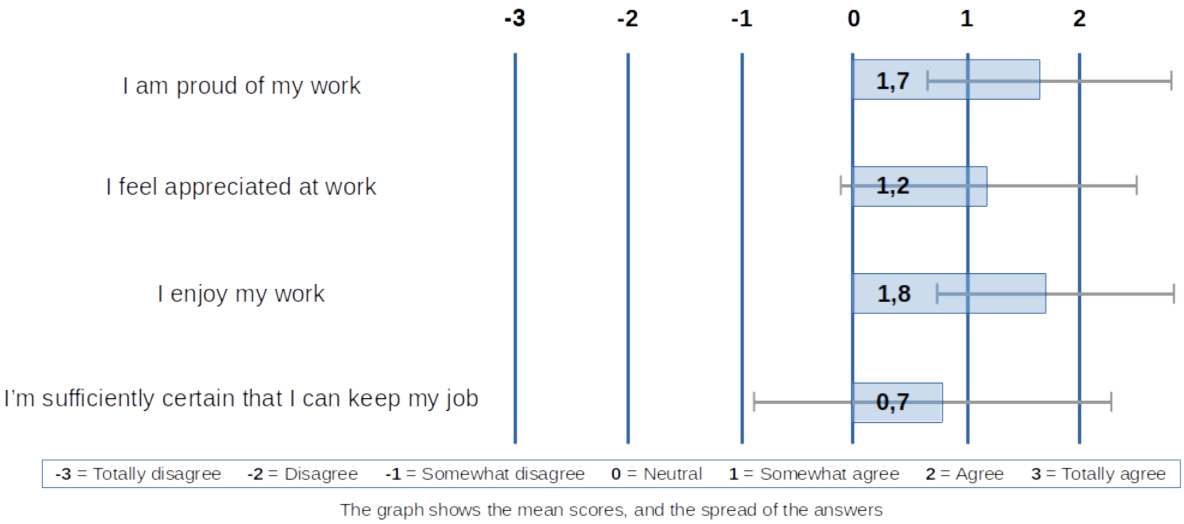


Figure 64.: *Survey Results: Motivation (van Veelen, 2018)*

Relevant Comments:

- Often, temporary workers are used.
- The costs are often leading in tenders, not the craftsmanship

Workload

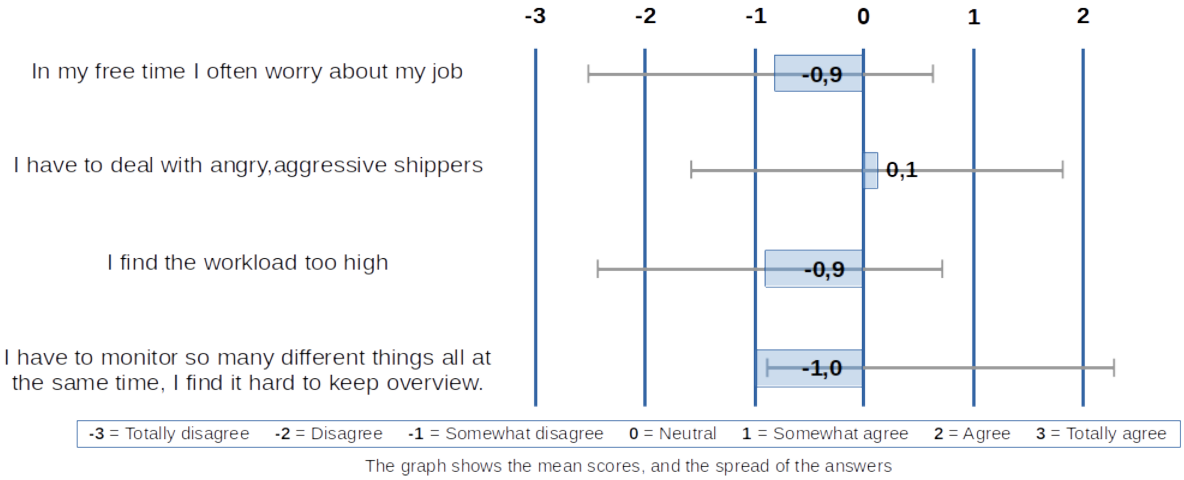
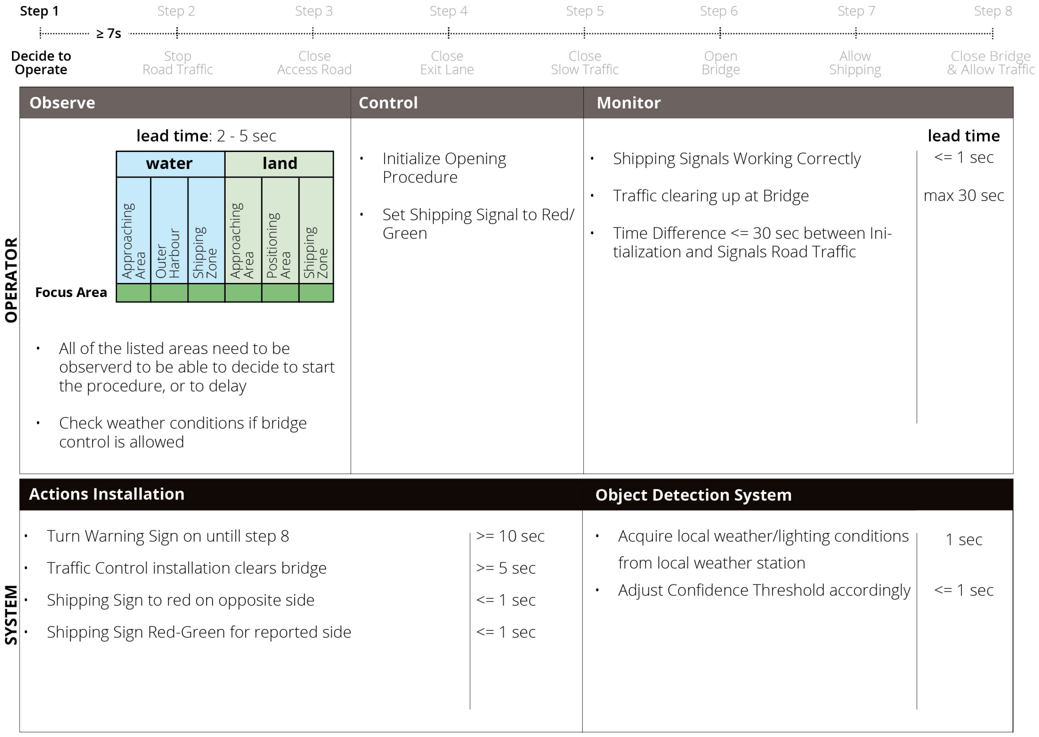


Figure 65.: *Survey Results: Workload (van Veelen, 2018)*

Relevant Comments:

- Changes in day/night shifts cause concentration problems.
- Many respondents mentioned the changing workload(e.g. summer- and winter period), leading to boringness when it is silent, and too much pressure when it is busy.

D. Workflow



Observe — **lead time**: 2 - 5 sec

- All of the listed areas need to be observerd to be able to decide to start the procedure, or to delay
- Check weather conditions if bridge control is allowed

Control

- Initialize Opening Procedure
- Set Shipping Signal to Red/ Green

Monitor

	lead time
• Shipping Signals Working Correctly	<= 1 sec
• Traffic clearing up at Bridge	max 30 sec
• Time Difference <= 30 sec between Initialization and Signals Road Traffic	

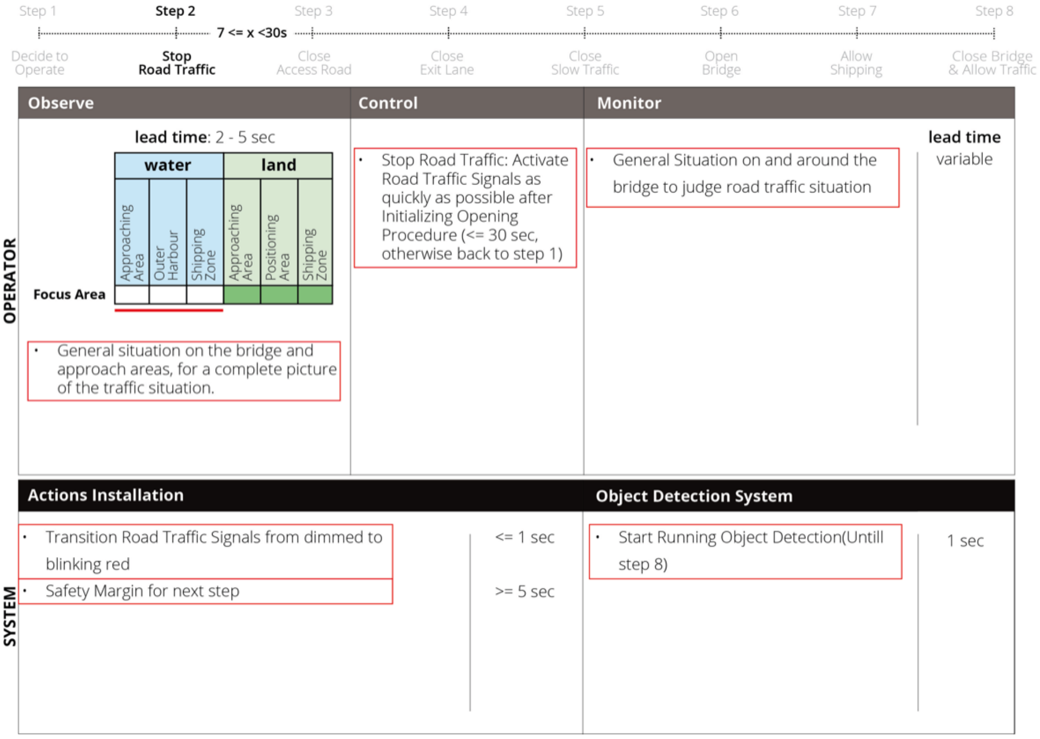
Actions Installation

• Turn Warning Sign on untill step 8	>= 10 sec
• Traffic Control installation clears bridge	>= 5 sec
• Shipping Sign to red on opposite side	<= 1 sec
• Shipping Sign Red-Green for reported side	<= 1 sec

Object Detection System

• Acquire local weather/lighting conditions from local weather station	1 sec
• Adjust Confidence Threshold accordingly	<= 1 sec

Figure 66.: *Step 1: Decide to Operate (Intergo, 2019)*



Observe

lead time: 2 - 5 sec

- General situation on the bridge and approach areas, for a complete picture of the traffic situation.

Control

- Stop Road Traffic: Activate Road Traffic Signals as quickly as possible after Initializing Opening Procedure (<= 30 sec, otherwise back to step 1)

Monitor

- General Situation on and around the bridge to judge road traffic situation

lead time variable

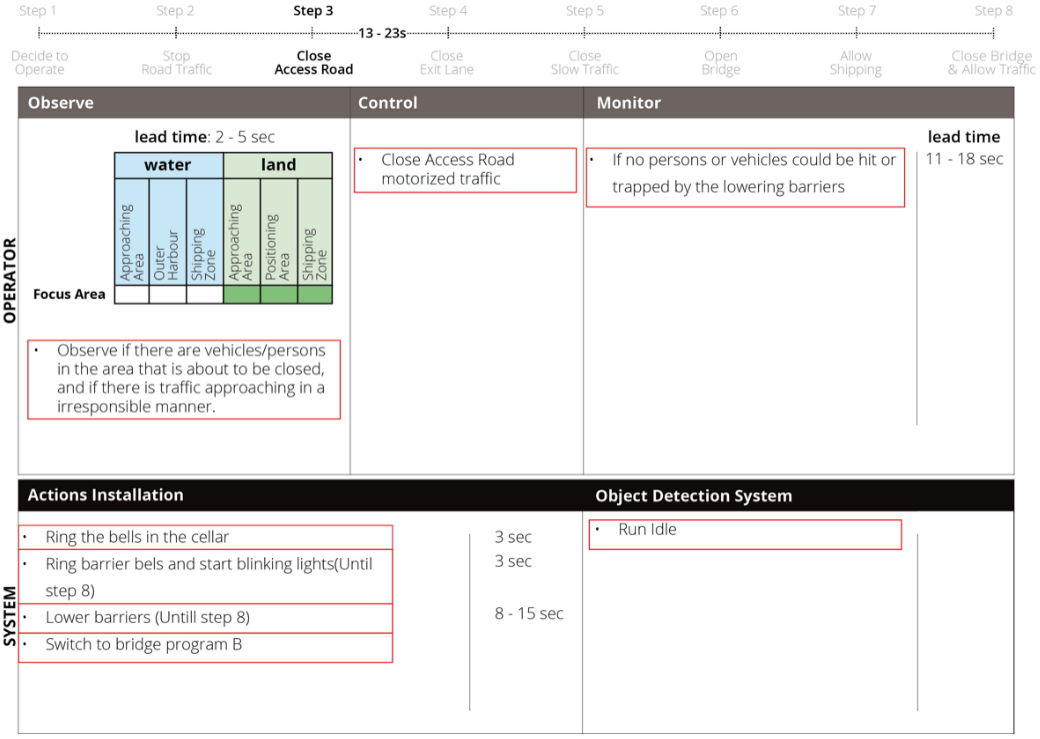
Actions Installation

- Transition Road Traffic Signals from dimmed to blinking red — <= 1 sec
- Safety Margin for next step — >= 5 sec

Object Detection System

- Start Running Object Detection(Untill step 8) — 1 sec

Figure 67.: *Step 2: Stop Road Traffic (Intergo, 2019)*



Observe

lead time: 2 - 5 sec

- Observe if there are vehicles/persons in the area that is about to be closed, and if there is traffic approaching in a irresponsible manner.

Control

- Close Access Road motorized traffic

Monitor

- If no persons or vehicles could be hit or trapped by the lowering barriers

lead time 11 - 18 sec

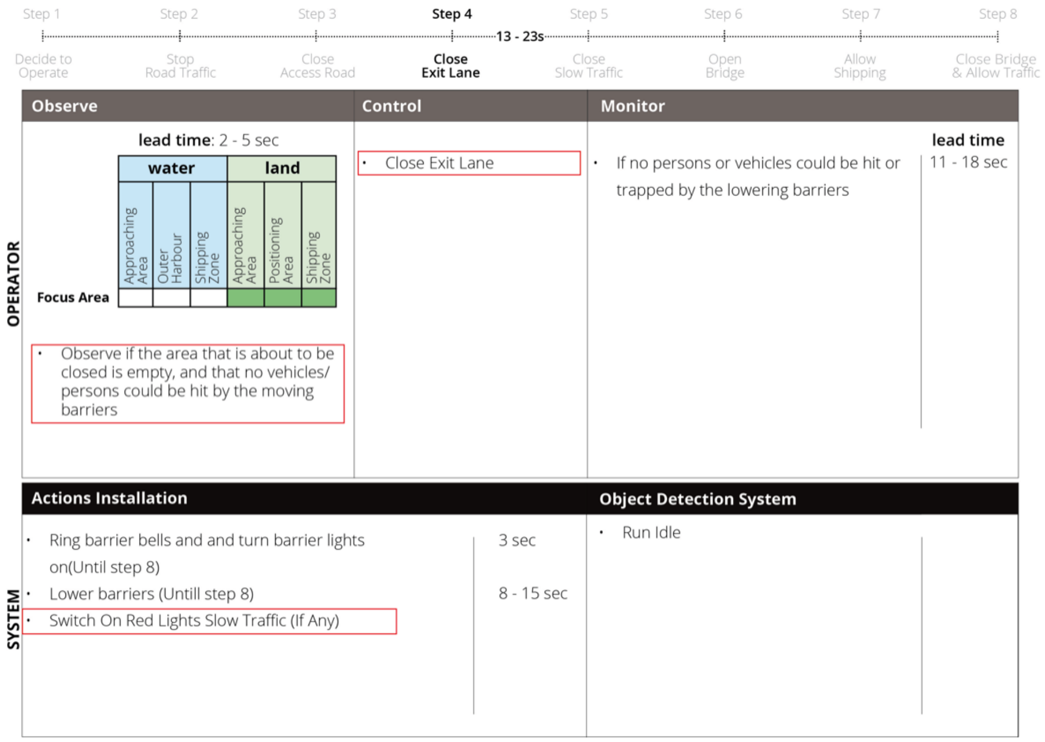
Actions Installation

- Ring the bells in the cellar — 3 sec
- Ring barrier bels and start blinking lights(Until step 8) — 3 sec
- Lower barriers (Untill step 8) — 8 - 15 sec
- Switch to bridge program B

Object Detection System

- Run Idle

Figure 68.: *Step 3: Close Access Road (Intergo, 2019)*



Step 1 · Step 2 · Step 3 · **Step 4** · Step 5 · Step 6 · Step 7 · Step 8

13 - 23s

Decide to Operate · Stop Road Traffic · Close Access Road · **Close Exit Lane** · Close Slow Traffic · Open Bridge · Allow Shipping · Close Bridge & Allow Traffic

OPERATOR

Observe

lead time: 2 - 5 sec

	water			land		
	Approaching Area	Outer Harbour	Shipping Zone	Approaching Area	Positioning Area	Shipping Zone
Focus Area				🟩	🟩	🟩

- Observe if the area that is about to be closed is empty, and that no vehicles/ persons could be hit by the moving barriers

Control

- Close Exit Lane

Monitor

- If no persons or vehicles could be hit or trapped by the lowering barriers

lead time
11 - 18 sec

SYSTEM

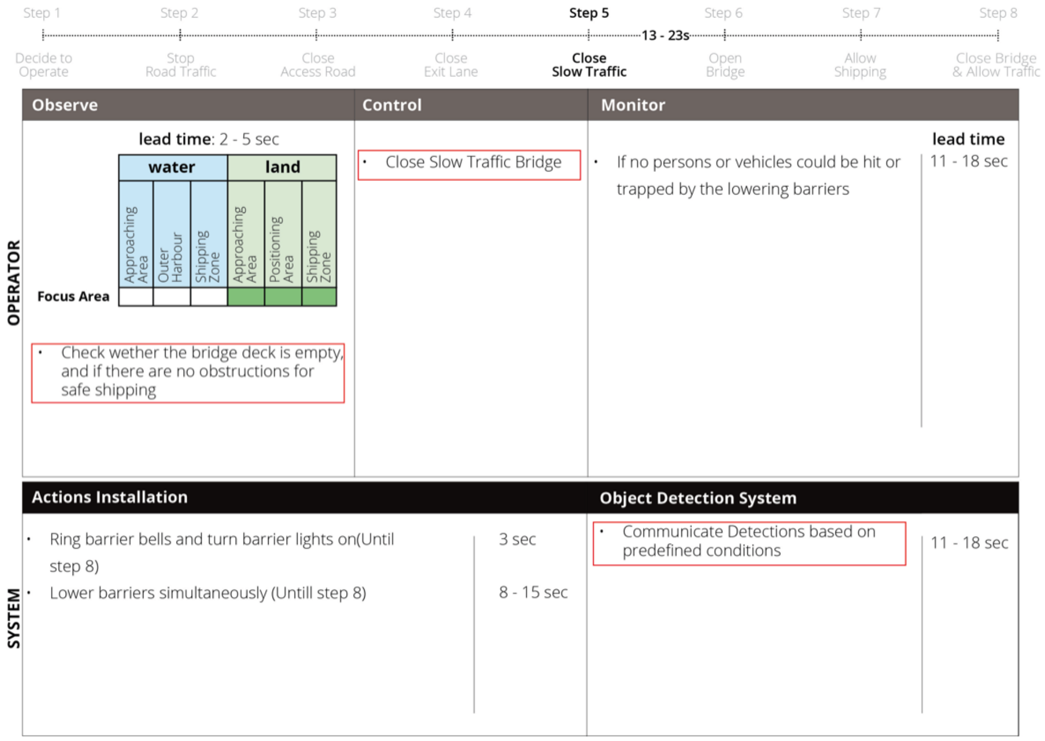
Actions Installation

- Ring barrier bells and and turn barrier lights on(Until step 8) — 3 sec
- Lower barriers (Untill step 8) — 8 - 15 sec
- Switch On Red Lights Slow Traffic (If Any)

Object Detection System

- Run Idle

Figure 69.: *Step 4: Close Exit Lane (Intergo, 2019)*



Step 1 · Step 2 · Step 3 · Step 4 · **Step 5** · Step 6 · Step 7 · Step 8

13 - 23s

Decide to Operate · Stop Road Traffic · Close Access Road · Close Exit Lane · **Close Slow Traffic** · Open Bridge · Allow Shipping · Close Bridge & Allow Traffic

OPERATOR

Observe

lead time: 2 - 5 sec

	water			land		
	Approaching Area	Outer Harbour	Shipping Zone	Approaching Area	Positioning Area	Shipping Zone
Focus Area				🟩	🟩	🟩

- Check wether the bridge deck is empty, and if there are no obstructions for safe shipping

Control

- Close Slow Traffic Bridge

Monitor

- If no persons or vehicles could be hit or trapped by the lowering barriers

lead time
11 - 18 sec

SYSTEM

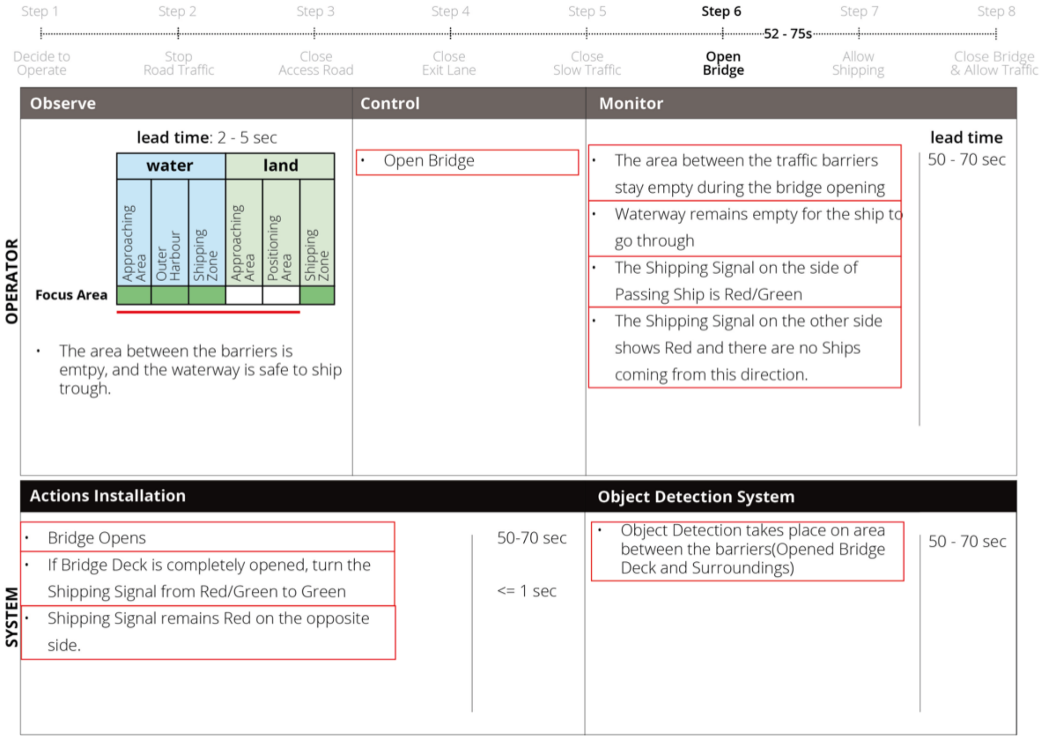
Actions Installation

- Ring barrier bells and turn barrier lights on(Until step 8) — 3 sec
- Lower barriers simultaneously (Untill step 8) — 8 - 15 sec

Object Detection System

- Communicate Detections based on predefined conditions — 11 - 18 sec

Figure 70.: *Step 5: Close Slow Traffic (Intergo, 2019)*



Observe

lead time: 2 - 5 sec

- The area between the barriers is emtpy, and the waterway is safe to ship trough.

Control

- Open Bridge

Monitor

lead time 50 - 70 sec

- The area between the traffic barriers stay empty during the bridge opening
- Waterway remains empty for the ship to go through
- The Shipping Signal on the side of Passing Ship is Red/Green
- The Shipping Signal on the other side shows Red and there are no Ships coming from this direction.

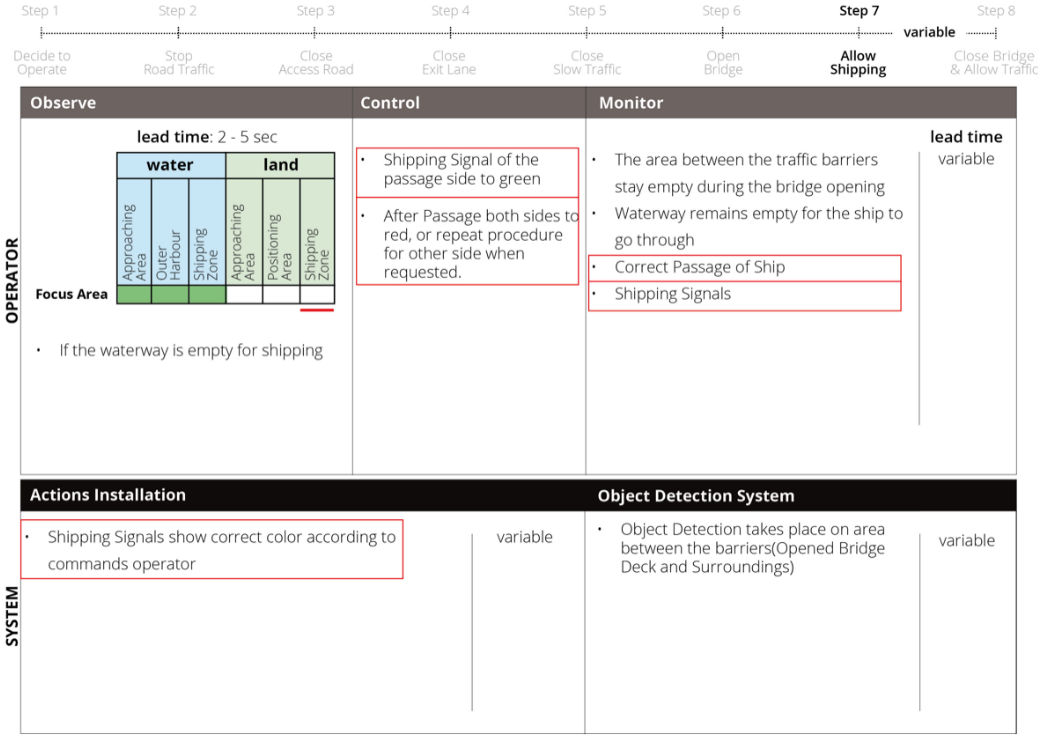
Actions Installation

- Bridge Opens — 50-70 sec
- If Bridge Deck is completely opened, turn the Shipping Signal from Red/Green to Green — <= 1 sec
- Shipping Signal remains Red on the opposite side.

Object Detection System

- Object Detection takes place on area between the barriers(Opened Bridge Deck and Surroundings) — 50 - 70 sec

Figure 71.: *Step 6: Open Bridge (Intergo, 2019)*



Observe

lead time: 2 - 5 sec

- If the waterway is empty for shipping

Control

- Shipping Signal of the passage side to green
- After Passage both sides to red, or repeat procedure for other side when requested.

Monitor

lead time variable

- The area between the traffic barriers stay empty during the bridge opening
- Waterway remains empty for the ship to go through
- Correct Passage of Ship
- Shipping Signals

Actions Installation

- Shipping Signals show correct color according to commands operator — variable

Object Detection System

- Object Detection takes place on area between the barriers(Opened Bridge Deck and Surroundings) — variable

Figure 72.: *Step 7: Allow Shipping (Intergo, 2019)*

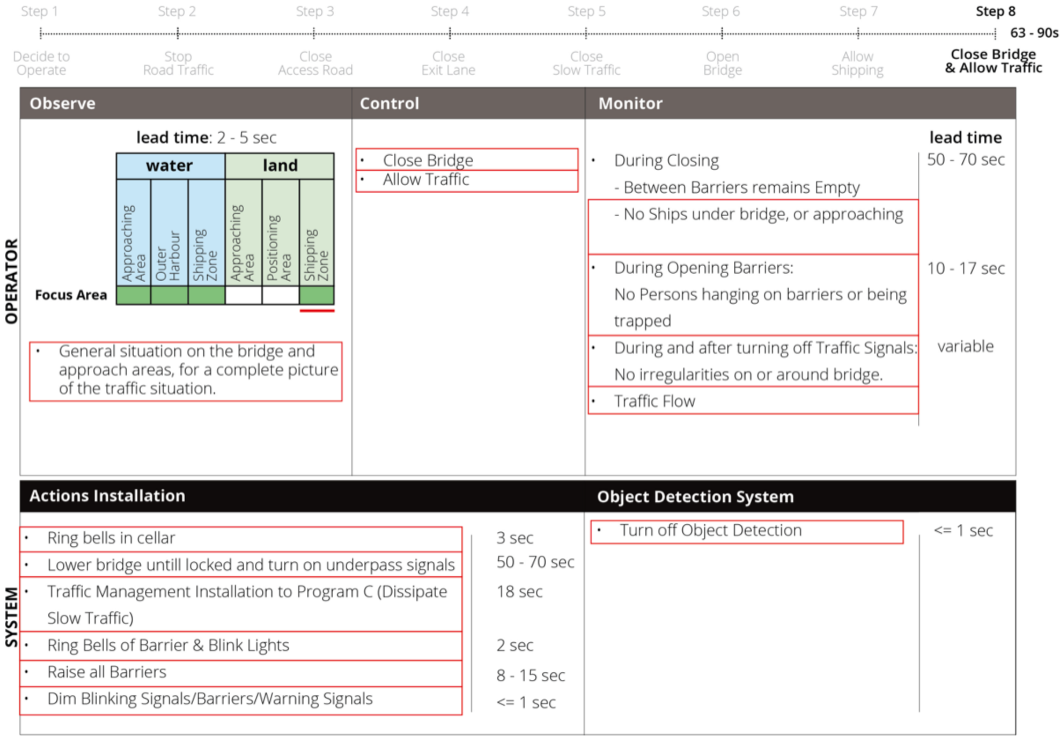
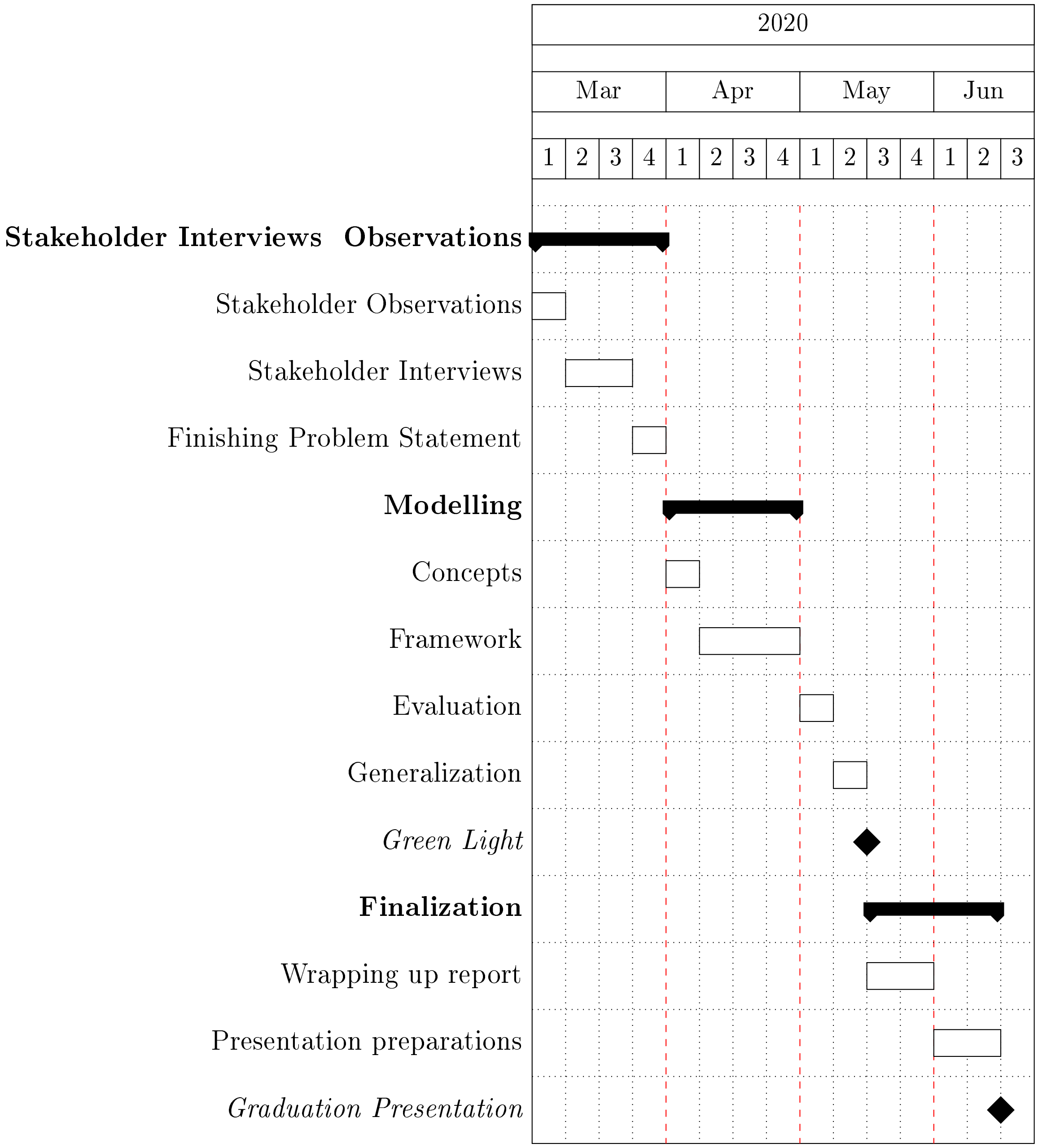


Figure 73.: *Step 8: Close Bridge Allow Traffic (Intergo, 2019)*


2020
Mar
Apr
May
Jun
1 2 3 4 1 2 3 4 1 2 3 4 1 2 3
Stakeholder Interviews Observations
Stakeholder Observations
Stakeholder Interviews
Finishing Problem Statement
Modelling
Concepts
Framework
Evaluation
Generalization
Green Light
Finalization
Wrapping up report
Presentation preparations
Graduation Presentation

References

Adams, A., Lunt, P., & Cairns, P. (2016). A qualitative approach to HCI research. In *Research methods for human–computer interaction.* doi: 10.1017/cbo9780511814570.008

ADP. (2018). *The Workforce View in Europe 2018* (Tech. Rep.).

Allwein, E. L., Schapire, R. E., & Singer, Y. (2001). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research.* doi: 10.1162/15324430152733133

Amidi, A., & Amidi, S. (2018). CS 230-Deep Learning VIP Cheatsheet: Tips and Tricks Data processing. *Stanford University.*

Andrew, A. M. (2001). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* doi: 10.1108/k.2001.30.1.103.6

Antipov, G., Berrani, S. A., Ruchaud, N., & Dugelay, J. L. (2015). Learned vs hand-crafted features for pedestrian gender recognition. In *Mm 2015 - proceedings of the 2015 acm multimedia conference.* doi: 10.1145/2733373.2806332

ANWB. (2020). *Dagelijkse drukke trajecten ochtend en avondspits.* Retrieved 2020-06-12, from `https://www.anwb.nl/verkeer/nederland/verkeersinformatie/dagelijkse-drukke-trajecten`

Bautista, C. M., Dy, C. A., Mañalac, M. I., Orbe, R. A., & Cordel, M. (2016). Convolutional neural network for vehicle detection in low resolution traffic videos. In *Proceedings - 2016 ieee region 10 symposium, tensymp 2016.* doi: 10.1109/TENCONSpring.2016.7519418

Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., ... Venâncio, R. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - Journal of the American Medical Association.* doi: 10.1001/jama.2017.14585

Billings, C. E. (1996). Human-Centered Aviation Automation: Principles and Guidelines. *NASA Technical Memorandum.*

Boureau, Y. L., Ponce, J., & Lecun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Icml 2010 - proceedings, 27th international conference on machine learning.*

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools.*

Brownlee, J. (2018a). Better Deep Learning. Train Faster, Reduce Overfitting, and Make Better Predictions. *Machine Learning Mastery With Python.*

Brownlee, J. (2018b). Better Deep Learning. Train Faster, Reduce Overfitting, and Make Better Predictions. *Machine Learning Mastery With Python.*

Bureau of Air Safety Investigation. (1996). Human Factors in Fatal Aircraft Accidents. *Department of Transport and Regional Development*(Apr). Retrieved from `http://www.atsb.gov.au/publications/1996/sir199604{_}001.aspx`

Campesato, O. (2020). *Artificial Intelligence, Machine Learning, and Deep Learning.* Mercury Learning & Information. Retrieved from `https://books.google.nl/books?id=pqnNDwAAQBAJ`

Chen, Y., Xie, H., & Shin, H. (2018). Multi-layer fusion techniques using a CNN for multispectral pedestrian detection. *IET Computer Vision*. doi: 10.1049/iet-cvi.2018.5315

Chu, Q., Ouyang, W., Li, H., Wang, X., Liu, B., & Yu, N. (2017). Online Multi-object Tracking Using CNN-Based Single Object Tracker with Spatial-Temporal Attention Mechanism. In *Proceedings of the ieee international conference on computer vision.* doi: 10.1109/ICCV.2017.518

Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2016). Fast and accurate deep network learning by exponential linear units (ELUs). In *4th international conference on learning representations, iclr 2016 - conference track proceedings.*

Collumeau, J. F., Leconge, R., Emile, B., & Laurent, H. (2011). Hand-gesture recognition: Comparative study of global, semi-local and local approaches. In *Ispa 2011 - 7th international symposium on image and signal processing and analysis.*

Cooper, A., Reimann, R., & Cronin, D. (2007). *About Face 3: The essentials of interaction design.* doi: 10.1057/palgrave.ivs.9500066

Ding, L., Wang, Y., Laganière, R., Huang, D., & Fu, S. (2020). Convolutional neural networks for multispectral pedestrian detection. *Signal Processing: Image Communication.* doi: 10.1016/j.image.2019.115764

Dollár, P., Wojek, C., Schiele, B., & Perona, P. (2009). Pedestrian detection: A benchmark. In *2009 ieee computer society conference on computer vision and pattern recognition workshops, cvpr workshops 2009.* doi: 10.1109/CVPRW.2009.5206631

Draisma, I., & van Heugten, J. (2014). *Centrale post Zaanstad: Visie op werkorganisatie* (Tech. Rep.). vhp human performance.

Durlach, P. J. (2004). Change blindness and its implications for complex monitoring and control systems design and operator training. *Human-Computer Interaction.* doi: 10.1207/s15327051hci1904_10

Dürr, O. (2014). Machine Learning V04 : Support Vector Machines. Zurich University of Applied SciencesandArts.

ECRI Institute. (2020). ECRI Top 10 Health Technology Hazards for 2020. *Journal of Radiology Nursing.* doi: 10.1016/s1546-0843(20)30009-2

Elgammal, A. (2014). Background Subtraction: Theory and Practice. *Synthesis Lectures on Computer Vision.* doi: 10.2200/s00613ed1v01y201411cov006

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting.* doi: 10.1177/154193128803200221

Endsley, M. R. (1995). *Toward a theory of situation awareness in dynamic systems.* doi: 10.1518/001872095779049543

Endsley, M. R., & Robertson, M. M. (2000). Training for situation awareness in individuals and teams. In *Situation awareness analysis and measurement.*

Erlandson, R. F. (2007). *Universal and accessible design for products, services, and processes.* doi: 10.1201/9781420007664

Eysenck, M. W., & Keane, M. T. (2015). *Cognitive Psychology.* doi: 10.4324/9781315778006

Frey, C. B. (2019). *The Technology Trap: Capital, Labor, and Power in the Age of Automation.*

Gemeente Zaanstad. (2017). *Kwaliteitshandboek Havens en Vaarwegen: Centrale Post.*

Gemeente Zaanstad. (2019). *Een brug bedienen: zo gaat dat in Zaanstad.* Retrieved from https://www.youtube.com/watch?v=4wFstRBbKiM

Ghazizadeh, M., Peng, Y., Lee, J. D., & Boyle, L. N. (2012). Augmenting the Technology Acceptance Model with trust: Commercial drivers' attitudes towards monitoring and feedback. In *Proceedings of the human factors and ergonomics society.* doi: 10.1177/1071181312561481

Goldstein, E. B. (2009). *Sensation and Perception.* Cengage Learning. Retrieved from `https://books.google.nl/books?id=2tW91BWeNq4C`

Gregor, S. (2006). The nature of theory in Information Systems. *MIS Quarterly: Management Information Systems.* doi: 10.2307/25148742

Hahnioser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex- inspired silicon circuit. *Nature.* doi: 10.1038/35016072

Hartel, C., Smith, K., & Prince, C. (1991). Defining aircrew coordination. In *Sixth international symposium on aviation psychology.*

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee computer society conference on computer vision and pattern recognition.* doi: 10.1109/CVPR.2016.90

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., ... Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. Retrieved from `http://arxiv.org/abs/1712.00409`

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks.* doi: 10.1109/72.991427

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... Murphy, K. (2017). Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. In (pp. 3296–3297). doi: 10.1109/CVPR.2017.351

human performance, V. (2017). *Incident Bosrandbrug* (Tech. Rep.).

Inoue, H. (2018). *Data Augmentation by Pairing Samples for Images Classification* (Tech. Rep.).

Intergo. (2019). *Uniform bedienconcept: Gemeente Zaanstad* (Tech. Rep.).

International Ergonomics Association. (2000). *Human Factors/Ergonomics (HF/E).* Retrieved 2020-06-27, from `https://iea.cc/what-is-ergonomics/`

Jensen, M. S., Yao, R., Street, W. N., & Simons, D. J. (2011). Change blindness and inattentional blindness. *Wiley Interdisciplinary Reviews: Cognitive Science.* doi: 10.1002/wcs.130

Jiang, Y., & Wang, X. (2012). A Background Model Combining Adapted Local Binary Pattern with Gauss Mixture Model. *Advances in Intelligent and Soft Computing*, *159*, 7–12. doi: 10.1007/978-3-642-29387-0_2

Karg, M., & Scharfenberger, C. (2020). Deep Learning-Based Pedestrian Detection for Automated Driving: Achievements and Future Challenges. In *Studies in computational intelligence.* doi: 10.1007/978-3-030-31764-5_5

Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). A Guide to Convolutional Neural Networks for Computer Vision. *Synthesis Lectures on Computer Vision.* doi: 10.2200/s00822ed1v01y201712cov015

Kim, J., & Lee, M. (2014). Robust lane detection based on convolutional neural network and random sample consensus. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* doi: 10.1007/978-3-319-12637-1_57

KNMI. (2020). *Tijden van zonopkomst en –ondergang 2020.* Retrieved 2020-06-12, from `https://cdn.knmi.nl/system/ckeditor{_}assets/attachments/102/`

`tijden{_}van{_}zonopkomst{_}en{_}-ondergang{_}2020.pdf`

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, *25*. doi: 10.1145/3065386

Lee, J. D., & See, K. A. (2004). *Trust in automation: Designing for appropriate reliance.* doi: 10.1518/hfes.46.1.50_30392

Lee, S. Y., Tama, B. A., Moon, S. J., & Lee, S. (2019). Steel surface defect diagnostics using deep convolutional neural network and class activation map. *Applied Sciences (Switzerland)*, *9*(24). doi: 10.3390/app9245449

Lemley, J., Abdul-Wahid, S., Banik, D., & Andonie, R. (2016). Comparison of recent machine learning techniques for gender recognition from facial images. In *Ceur workshop proceedings.*

Li, C., Song, D., Tong, R., & Tang, M. (2019). Illumination-aware faster R-CNN for robust multispectral pedestrian detection. *Pattern Recognition*. doi: 10.1016/j.patcog.2018.08.005

Li, H., Wang, P., You, M., & Shen, C. (2018). Reading car license plates using deep neural networks. *Image and Vision Computing*. doi: 10.1016/j.imavis.2018.02.002

Lidwell, W., Holden, K., & Butler, J. (2010). Universal Principles of Design. *Universal principles of design: 125 ways to enhance usability, influence perception, increase appeal, make beter design decisions, and teach through design.*. doi: 10.1007/s11423-007-9036-7

Light, K. E., Reilly, M. A., Behrman, A. L., & Spirduso, W. W. (1996). Reaction times and movement times: Benefits of practice to younger and older adults. *Journal of Aging and Physical Activity*. doi: 10.1123/japa.4.1.27

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*. doi: 10.1007/s11263-019-01247-4

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*. doi: 10.2307/258792

Merket, D., Bergondy, M., & Cuevas-Mesa, H. (1997). Making sense out of teamwork errors in complex environments. In *18th annual industrial/organizational behavior conference.*

Michelucci, U. (2019). *Advanced applied deep learning: Convolutional neural networks and object detection.* doi: 10.1007/978-1-4842-4976-5

Mileti, D., & Sorensen, J. (1990). Communication of emergency public warnings : A social perspective and State-of-the-art assessment. *Landslides*. doi: 10.2172/6137387

MLMath.io. (2019). *Math behind SVM.* Retrieved 2020-03-02, from `https://mc.ai/math-behind-svmsupport-vector-machine/`

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve Restricted Boltzmann machines. In *Icml 2010 - proceedings, 27th international conference on machine learning.*

NOS Nieuws. (2019, sep). Zaanstad pakt onveilige bruggen aan en burgemeester maakt excuses. Retrieved from `https://nos.nl/artikel/2303248-zaanstad-pakt-onveilige-bruggen-aan-en-burgemeester-maakt-excuses.html`

Onderzoeksraad voor de Veiligheid. (2016, jan). *Ongeval Den Uylbrug, Zaandam* (Tech. Rep.). Den Haag: Onderzoeksraad voor de Veiligheid.

Onderzoeksraad voor de Veiligheid. (2019, sep). *Veiligheid van op afstand bediende bruggen* (Tech. Rep.). Den Haag: Onderzoeksraad voor de Veiligheid.

Patterson, J., & Gibson, A. (2017). *Deep Learning. A Practitioner's Approach.* doi: 10.13209/j.0479-8023.2014.011

Peemen, M., Mesman, B., & Corporaal, C. (2011). Speed sign detection and recognition by convolutional neural networks. *Proceedings of the 8th ....* Retrieved from http://parse.ele.tue.nl/system/attachments/11/original/paperspeedsigncnn.pdf

Proctor, R. W., & Van Zandt, T. (2008). *Human factors in simple and complex systems.* doi: 10.1080/00140139.2019.1638068

Ribeiro, D., Nascimento, J. C., Bernardino, A., & Carneiro, G. (2017). Improving the performance of pedestrian detectors using convolutional learning. *Pattern Recognition.* doi: 10.1016/j.patcog.2016.05.027

Rifkin, R., & Klautau, A. (2004). In defense of one-vs-all classification. *Journal of Machine Learning Research.*

Rijkswaterstaat. (2019). *Snelweg A7.* Author. Retrieved from https://www.rijkswaterstaat.nl/wegen/wegenoverzicht/a7/index.aspx

Ruskin, K. J., & Hueske-Kraus, D. (2015). *Alarm fatigue: Impacts on patient safety.* doi: 10.1097/ACO.0000000000000260

Salvendy, G. (2012). *Handbook of Human Factors and Ergonomics: Fourth Edition.* doi: 10.1002/9781118131350

Satzinger, J. W., & Olfman, L. (1998). User interface consistency across end-user applications: The effects on mental models. *Journal of Management Information Systems.* doi: 10.1080/07421222.1998.11518190

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In *Automation and human performance: Theory and applications.* doi: 10.1201/9781315137957

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly: Management Information Systems*, *35*(1), 37–56. doi: 10.2307/23043488

Sendelbach, S., & Funk, M. (2013). Alarm fatigue: A patient safety concern. *AACN Advanced Critical Care.* doi: 10.1097/NCI.0b013e3182a903f9

Shaikh, S. H., Saeed, K., & Chaki, N. (2014). Moving object detection using background subtraction. In *Springerbriefs in computer science.* doi: 10.1007/978-3-319-07386-6_3

Sharma, S. (2019, dec). *Activation Functions in Neural Networks.* Retrieved from https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception.* doi: 10.1068/p281059

Stanford University. (2018, feb). *Introduction to {Convolutional} {Neural} {Networks}* [Lecture]. Retrieved from https://web.stanford.edu/class/cs231a/lectures/intro{_}cnn.pdf

Statnikov, A., Aliferis, C. F., Hardin, D. P., & Guyon, I. (2013). *A Gentle Introduction to Support Vector Machines in Biomedicine.* doi: 10.1142/7923

Stauffer, C., & Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* doi: 10.1109/34.868677

Suleiman, A., & Sze, V. (2016). An Energy-Efficient Hardware Implementation of HOG-Based Object Detection at 1080HD 60 fps with Multi-Scale Support. *Journal of Signal Processing Systems.* doi: 10.1007/s11265-015-1080-7

Sun, Y., Gao, H., Guo, L., Hong, X., Song, H., Zhang, J., & Li, L. (2020). A New Intelligent Fault Diagnosis Method and Its Application on Bearings. In (pp. 618–628). doi: 10.1007/978-981-13-8331-1_46

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*. doi: 10.1037/h0030300

Tensorflow. (2019, dec). *Tensorflow detection model zoo.* Retrieved from `https://github.com/tensorflow/models/blob/master/research/object{_}detection/g3doc/detection{_}model{_}zoo.md`

The Learning Machine. (2020). *Classification: Convolutional Neural Network (CNN).* Retrieved 2020-06-02, from `https://www.thelearningmachine.ai/cnn`

Todd, H. (2016, aug). *people-walking-across-the-bridge.* Retrieved from `http://www.public-domain-image.com/public-domain-images-pictures-free-stock-photos/people-public-domain-images-pictures/people-walking-across-the-bridge.jpg`

Tomè, D., Monti, F., Baroffio, L., Bondi, L., Tagliasacchi, M., & Tubaro, S. (2016). Deep Convolutional Neural Networks for pedestrian detection. *Signal Processing: Image Communication*. doi: 10.1016/j.image.2016.05.007

van Nieuwenhuizen Wijbenga, C. (2019, sep). *{OVV} rapport '{Veiligheid} van op afstand bediende bruggen'* [Kamerbrief]. Retrieved from `https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/09/04/ovv-rapport-veiligheid-van-op-afstand-bediende-bruggen/ovv-rapport-veiligheid-van-op-afstand-bediende-bruggen.pdf`

van Veelen, H. (2018). *Uitkomsten Nationale Enquête Brug- en Sluiswachters 2017* (Tech. Rep.). 's-Gravenhage: VHP Human Performance BV.

Venkatesan, R., Li, B., Venkatesan, R., & Li, B. (2018). Convolutional Neural Networks. In *Convolutional neural networks in visual computing.* doi: 10.4324/9781315154282-4

Wang, A. (2019, dec). *Convolutional Neural Networks(CNN) #6 Pooling in Backward pass.* Retrieved from `https://www.brilliantcode.net/1781/convolutional-neural-networks-6-backpropagation-in-pooling-layers-of-cnns/`

Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. doi: 10.1109/TPAMI.2003.1251144

Weber, E. (1978). *E. H. Weber: the sense of touch.*

Wickens, C. D., Lee, J., Gordon-Becker, S., & Liu, Y. (2014). *An Introduction to Human Factors Engineering.* Pearson. Retrieved from `https://books.google.nl/books?id=930ingEACAAJ`

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a "cry wolf" effect? *Human Factors*. doi: 10.1177/0018720809344720

Wieringa, P. A., & Wawoe, D. P. (1998). Operator support system dilemma: balancing a reduction in task complexity vs. an increase in system complexity. In *Proceedings of the ieee international conference on systems, man and cybernetics.* doi: 10.1109/icsmc.1998.725546

Xu, Z., Li, S., & Deng, W. (2016). Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In *Proceedings - 3rd iapr asian conference on pattern recognition, acpr 2015.* doi: 10.1109/ACPR.2015.7486482

Yoo, Y., Tang, L. W., Brosch, T., Li, D. K., Metz, L., Traboulsee, A., & Tam, R.

(2016). Deep learning of brain lesion patterns for predicting future disease activity in patients with early symptoms of multiple sclerosis. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics).* doi: 10.1007/978-3-319-46976-8_10

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). *Dive into Deep Learning.*

Zuboff, S. (1988). Dilemmas of transformation in the age of the smart machine. In *In the age of the smart machine: The future of work and power.*