



Graduation thesis
Project for SAXION AMI and Go2Sure

Andrii Roizin
465182

Abstract

The graduation project of Andrii Roizin, a software engineering student at Saxion, was developed as a part of the Sportkeur 2.0 project. This project is led by Go2Sure and the Saxion Ambient Intelligence research group is one of its partners.

The purpose of the project was to discover the possibility of automating the process of inspecting possible hazardous situations of the arenas/tracks used for biking events using computer vision. This checkup is performed by a safety inspector. The idea was to develop such a solution with the help of semantically segmented data to detect hazardous situations like poor road conditions, unclear track directions, steep turns, missing signboards, etc using the videos from the track.

The research and development resulted in a software system that is able to take the video from the biking track as an input, analyze it using computer vision techniques, and output the semantically segmented video. The problem was solved using modern computer vision and software technologies and the solution satisfies the requirements of Go2Sure and the AMI research group. The semantically segmented videos were created using the DeepLabV3+ machine learning architecture. It reached 76.2% of mIOU on the validation data and was able to deal with the traffic situations on Dutch roads. Likewise, the machine learning solution was integrated into the existing SportKeur system using Django API. At the same time, there is room for future improvements and the possibilities for the continuation of the project are described in the document alongside the explanations of the important decisions.

Graduation Supervisors

Danny Plass-Oude Bos	Associate lector at ACT-LC Ambient Intelligence - Saxion graduation supervisor.
Raoul Willemsen MSc	Algemeen directeur (CEO) at Go2Sure SportKeur – Supervisor from Go2Sure.
Iris Heerlien	Researcher/lecturer Applied Data Science at Aml – Supervisor from AMI side.

Table of contents

Glossary	3
Table of abbreviations	3
Chapter 1 – Introduction	4
1.1 Document structure.....	4
1.2 Project SportKeur 2.0.....	4
1.3 AMI research group	4
1.4 Go2Sure	4
1.5 Problem context	5
1.6 Problem statement.....	5
1.7 Assignment goal	5
1.8 Main research questions	5
1.9 Deliverables	6
Chapter 2 – Project management	7
2.1. Methodology	7
2.2 Progress control.....	8
2.3 Quality control.....	8
2.4 Research Methods.....	9
Chapter 3 – Machine learning development cycle.....	10
3.1 Business Understanding	10
3.2 Data Understanding and Preparation	12
3.3 Modeling.....	15
3.4 Model Implementation and Evaluation	19
Chapter 4 – System implementation.....	22
4.1 First version of the system	22
4.2 Integration to SportKeur system	26
Chapter 5 – Evaluation	28

Conclusion	28
Recommendations.....	29
List of figures	30
Version control	30
References	30

Glossary

No	Notation	Definition
1	Machine learning	Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data.
2	Computer vision	Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs
3	Semantic segmentation	Semantic segmentation is the process of labeling every pixel in the image.

Table of abbreviations

No	Abbreviation	Definition
1	Aml	Saxion Ambient Intelligence research group
2	ICT	Information and Communications Technology
3	ML	Machine learning
4	CNN	Convolutional Neural Network
5	mIOU	Mean Intersection Over Union
6	CRISP-DM	CRoss Industry Standard Process for Data Mining

Chapter 1 – Introduction

This document reports the graduation assignment of Andrii Roizin, a Software Engineering student at Saxion University of Applied Sciences. The graduation internship was started in February 2022 at lasted till July 2022, under the guidance of the supervisors from Saxion and company sides. It contains the description of the problem, the solution that was developed to solve it, outcomes of the research, analysis, design, and development that was done for it, and recommendations for the future of the project.

1.1 Document structure

This document consists of the 5 chapters with the following contents:

- Chapter 1 – provides the introduction to the report and describes the project and company background, problem, and assignment.
- Chapter 2 – describes the approach and techniques that were used to manage the project and assure the quality of the solution.
- Chapter 3 – describes the machine learning part of development.
- Chapter 4 – describes the process of developing the system that executes the tasks.
- Chapter 5 – evaluates the project, and describes how each requirement was fulfilled.
- Conclusion and Recommendations – sums up the document and gives recommendations for the future of the project.

1.2 Project SportKeur 2.0

In the Netherlands, around 9.4 million people play sports every week. Sporting events are growing on a large scale and the popularity of more extreme sports is also increasing. These sports also involve risks. However, at the moment, there is no product or approach anywhere in the world that allows us to look at the risks involved in sports. Thus, project Sportkeur provides a risk management tool for sports, sports materials, and sporting events in a systematic way. SportKeur 2.0 focuses on developing and validating in practice a tool to determine the risks of sports, sports facilities, and sporting events and make them transparent and manageable.

This project is led by Go2Sure and the Saxion Ambient Intelligence research group is one of its partners. Both Go2Sure and the research group were involved in the project and supervised me during research and development.

1.3 AMI research group

The Ambient Intelligence Lectorate develops technology that makes 'everything' smart. Enabling IT for a smart world. They focus on IT for a 'smart world', and make this IT suitable for this future 'smart world' by using sensor technology, artificial intelligence, augmented reality, etc. The research group thus develops applications related to safety, sports, smart industry, sustainable energy supply, and animal monitoring; but is in principle application domain independent (AMI, 2022). Likewise, AMI actively involves students in their projects to increase students' research and development skills.

1.4 Go2Sure

Go2Sure is a company that focuses on developing software for risk management. The development of the risk management software takes place with their own team of

programmers, security experts, and risk scientists in the Netherlands. What is unique about the Go2Sure concept is that it involves integral and dynamic risk management, with knowledge of software and content (Go2Sure, 2022).

1.5 Problem context

The Netherlands is very famous for its biking culture. Thus, every year lots of biking competitions are taking place here. Along with the Royal Dutch Cycling Union (KNWU), Go2Sure ensures the safety aspects of such events. To secure the riders from unpredicted situations on the track, it should be reviewed for the possible obstacles that can cause risks for the bikers and those risks should be analyzed. Conventionally, an inspection is performed by a safety inspector who inspects the biking arena/track for possible hazardous situations. This approach causes the risk of some situations being unnoticed because of the human factor (Willemsen, 2022).

1.6 Problem statement

With the expertise of Ambient Intelligence, Go2Sure would like to automate this process where these situations could be detected with the use of machine learning techniques. The idea is to develop such a solution with the help of semantically segmented data to detect hazardous situations like poor road conditions, unclear track directions, steep turns, missing signboards, etc. The goal of the SportKeur project is to develop a robust and deployable real-life scenario system using machine learning. This system should be able to detect various elements in a video from the biking track, determine which of them are the risks for events, and display the semantically segmented video. For example, in Figure 1 – Semantic segmentation example you can see that the driving area is detected, as well as the vehicles and the other objects which could be a potential risk for the bikers if they appear to be on the track.



Figure 1 – Semantic segmentation example

1.7 Assignment goal

The goal of the assignment is to develop a system that is able to take the video from the biking track as an input, analyze it using machine learning techniques, and output the semantically segmented video. The scope of this assignment includes only analysis of the video and distinguishing objects of different classes. Using this system the inspector would have an opportunity to double-check the track using the labeled video. After that, a risk analysis tool can be built on top of this system to fully automate the process.

1.8 Main research questions

The main research question is the question that the whole assignment is aimed to answer. The main question is the core of the project and describes its goal. The question is:

How to provide optimal object detection of road videos through machine learning to support the automation of the safety inspection of biking tracks?

1.9 Deliverables

During the research and development of the project, several professional products were delivered to Go2Sure, AMI, and Saxion. The list of those deliverables is presented below. The deliverables were reviewed and approved by the supervisors.

- List of requirements
- Technical design
- Trained and validated computer vision model
- Deployable system
- Thesis
- Final presentation

Chapter 2 – Project management

This chapter describes approaches, techniques, and tools used to plan and manage the project and ensure the quality of the solution.

2.1. Methodology

Since the main part of the project was related to machine learning, the CRISP-DM methodology was applied. The Cross Industry Standard Process for Data Mining (CRISP-DM) is a process model that serves as the base for a data science process. According to the analytics, the distinctive characteristics have made CRISP-DM to be considered a 'de-facto' standard of data mining methodology and as a reference framework to which other methodologies are benchmarked. One of the important factors of using this method in Data Science is that it is a cross-industry standard for which it can be implemented in any Data Science project regardless of its domains (Why Is CRISP-DM Gaining Grounds, 2022).

It has six sequential phases (What is CRISP DM?, 2022):

1. Business understanding – What does the business need?
2. Data understanding – What data do we have/need? Is it clean?
3. Data preparation – How do we organize the data for modeling?
4. Modeling – What modeling techniques should we apply?
5. Evaluation – Which model best meets the business objectives?
6. Deployment – How do stakeholders access the results?

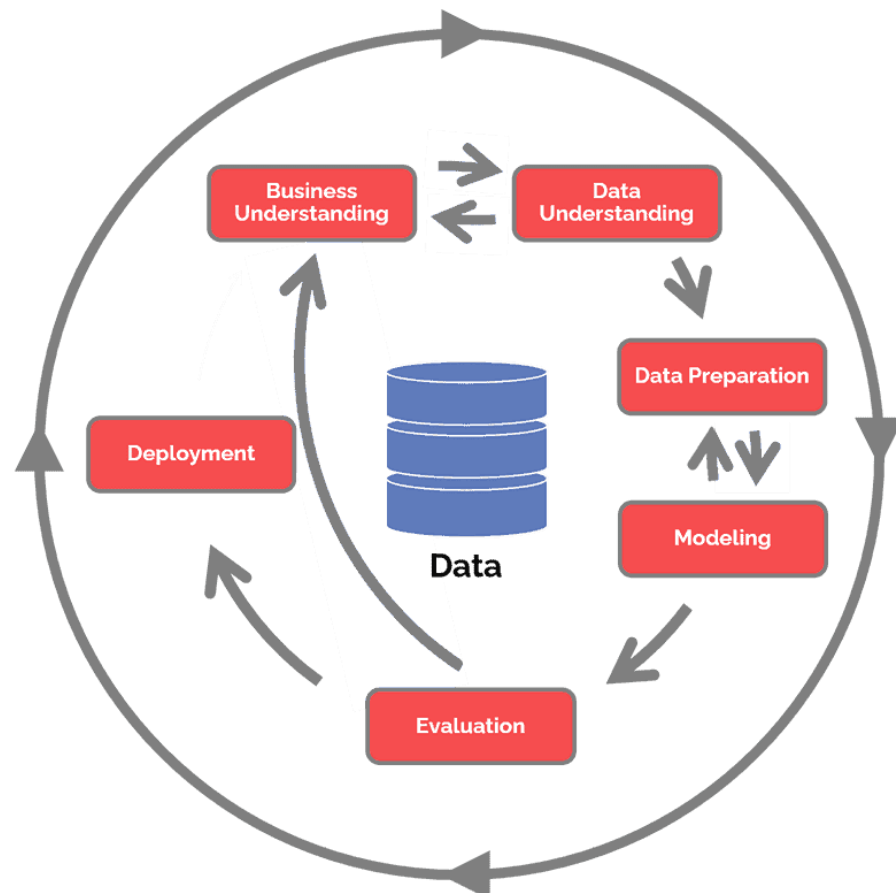


Figure 2 CRIPS-DM

The following process is circular, which means that you can repeat Understanding-Evaluation cycles until you reach the satisfying results and are ready to move to the deployment stage.

2.2 Progress control

The monitoring of the progress was conducted on the weekly basis together with the supervisors from Go2Sure and AMI. The review meetings were usually taking place on Monday mornings and included an overview of the work done the week before, a discussion of successes and obstacles, and plans for the coming week. Likewise, the monthly meetings with the Saxion supervisor were conducted in order to ensure that the project and the thesis are on track regarding the University deadlines.

The task management was conducted by the student himself using the boards in Go2Sure's enterprise Github account. Thus, all the members of the company had access to the boards to monitor the progress if needed. The picture below shows what this board looked like:

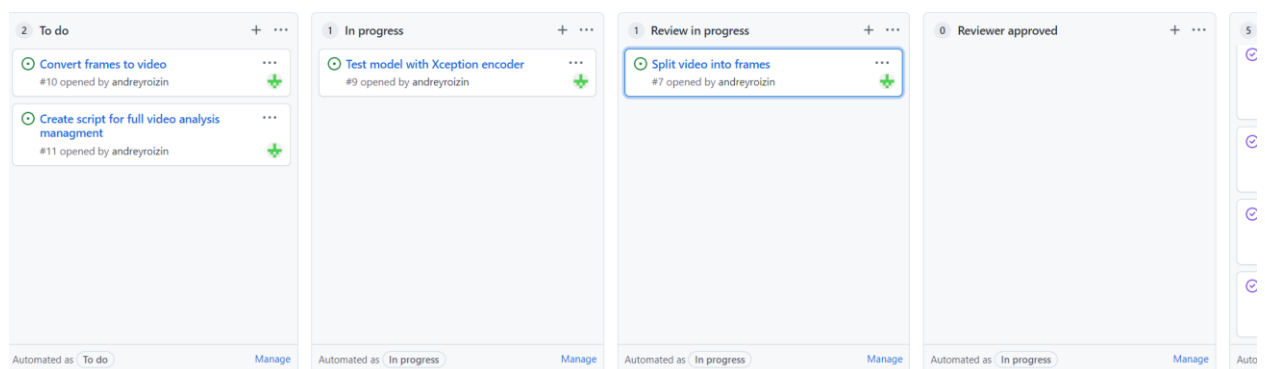


Figure 3 - GitHub board

It included the backlog with the open issues, the ones that are currently in progress, in review, approved, and done. That configuration was handy to perform the code quality control by the company members.

2.3 Quality control

The quality control of the code produced by the student was conducted by the Go2Sure developers Danny Hoojer and Bas van 't Holt. The GitHub repository has consisted 3 kinds of branches – main, development, and separate branch for each issue. This approach is used in Go2Sure to maintain the structure of the project. The main branch was used only at the end of the project to merge the system there. When the work on each issue was done, the merge request to the development branch was created. In order to be approved, the code had to be reviewed and accepted by at least one of the developers. Likewise, the reviewer could leave feedback that needed to be incorporated for approval. When the approval was done, the issue was moved to “Reviewer approved”. After that, the issue was discussed in the weekly session with the supervisors and moved to the “Done” column.

The quality control on documentation was performed by the supervisors. All the documents created were reviewed and approved by them, and the feedback was incorporated.

Also, the work on the final thesis was reviewed on a monthly basis by the Saxion supervisors. Alongside this, the peer review sessions with the fellow students were conducted to get additional feedback on the report, as well as a fresh view.

2.4 Research Methods

ICT Research Methods best practices were used to answer the research question during this project. ICT Research Methods is the toolkit designed for research conducted during ICT projects. It offers a set of possible research methods and a framework to select the appropriate (combination of) methods (ictresearchmethods.nl, 2021). The toolkit consists of the following research methods: Library, Field, Lab, Showroom, and Workshop. ICT research methods also include domain-specific approaches for machine learning, as an extension of the classic research methods.

Chapter 3 – Machine learning development cycle

This chapter describes the CRISP-DM cycle conducted to develop the machine learning part of this project.

3.1 Business Understanding

The first step of the CRISP-DM methodology is to get an understanding of the business needs of the stakeholders. That's why the first research subquestion was introduced.

3.1 Research sub-question 1 – What is the background of the problem that needs to be solved? What are the requirements for the system?

In order to gain a high-level understanding of the problem, the Problem Analysis¹ technique was applied at first. It started with a conversation with Raoul, the supervisor from Go2Sure, to clarify his vision of the project and expectations.

Afterward, the relevant work done by AMI members for this and relevant projects was studied. The purpose of it was to ensure the possibility of developing the solution for this problem using modern technologies. The research "Object detection and recognition in Sportkeur project" (Bagave, 2021), done by the previous AMI member Prachi Bagave who worked for Go2Sure was analyzed. In this work, the writer analyzes the possibility of applying computer vision for object detection for the SportKeur project and comes up with a positive answer. Likewise, this research outlines the possibility of using semantic segmentation for such types of tasks, which is described below.

The outcome of the problem analysis and background understanding has resulted in defining the goal of the assignment which is described in the first chapter of the report.

When the problem understanding was gained, the high-level working principles of the system were discussed with Raoul. The idea was to develop a system that is able to take the video as an input, analyze it using the machine learning algorithm, and output back the semantically segmented video. As the first step, the system should be able to get the manual input with the path to the video that is stored locally and needs to be analyzed to test and prove the possibility of analyzing the video. Afterward, if there is time left for it after completing the machine learning part, it should be connected to the existing SportKeur system and able to get, analyze and output the video by the API call received from it.

After that, the list of objects, which can appear on the track and in its surroundings was brainstormed, determined, and prioritized. Likewise, it was defined that the system should give accurate predictions in at least 75% of the cases, which, according to Raoul, is suitable for this stage of the project.

This high-level understanding resulted in creating the list of requirements. The list has passed through several review and feedback cycles before getting approval from the supervisors. The points of discussion were the relevance of the requirements as well as their prioritization.

¹ Problem Analysis – research technique of the Field ICT research method aimed to collect the needed information about the problem

Requirements were prioritized using the MoSCoW (MoSCoW_method, 2022) prioritization technique. MoSCoW is an acronym that stands for Must, Should, Could, and Would. It defines “Must”, “Should”, “Could”, and “Would” have features to determine the scope of the project.

“Must have” – the features that must be included in the project. Without those features, the project would not make sense.

“Should have” – less critical than “Must” but still crucial for the project. Those features should be developed for the first release of the project.

“Could have” – nice to have but not necessary features. Could be developed if there is time for it before the first release.

“Would have” – features that most probably will not be included in the first version of the project but can be implemented in the future.

Likewise, the requirements were split into functional and non-functional. The functional requirements define the desired behavior of the system, while the non-functional ones describe how it should perform.

The list of requirements and objects to be detected is provided below.

List of requirements

Table 1 - Requirments

Nr	Description	Priority	Functional/Non-functional
1	The system should be able to get the video stored locally by the path provided by the user as a manual input and analyze it.	Must	Functional
2	The system should be able to output the semantically segmented video back to the local storage.	Must	Functional
3	The system should be able to work on the roads in the Netherlands.	Must	Non-functional
4	The system should be designed in a way to ensure the possibility of integration into the SportKeur system.	Must	Non-functional
7	The machine learning model should be able to detect the object classes listed in the objects list below in at least 75% of the cases.	Should	Non-functional
8	The machine learning model should be able to detect the object sub -classes listed in the objects list below.	Could	Non-functional
9	The system should be able to get video by request from the SportKeur system as input and analyze it.	Could	Functional
10	The system should provide and ability to download the semantically segmented video	Could	Functional

	after finishing the analysis by request from the SportKeur system.		
11	The system should be able to output the list of potential risks detected in the video	Could	Functional
12	The system should be able to classify the risk level of the object detected.	Would	Functional

List of objects

List 1 - List of objects

1. Driving area
2. Sidewalk
3. Obstacle
 - a. Cemented block
 - b. Red_white striped indication
 - c. Cemented pole
 - d. Metal poles
 - e. Elevations
 - f. Road island
4. Vehicles
 - a. Cars
 - b. Trucks
 - c. Motor-bikes
 - d. Cycle
5. Non-drivable curbstone
6. Sky
7. Buildings
8. Trees
9. Pedestrians

The objects in the list are indexed by priority from highest to lowest. The final lists were approved by both Go2Sure and AMI supervisors.

According to the methodology, the next step after gathering the business understanding is Data Understanding.

3.2 Data Understanding and Preparation

The goal of this phase is to gather the data needed to train the machine learning model that could satisfy the business needs. For this purpose, the second research question was introduced.

Research sub-question 2 - What data is available? Does it contain enough data on objects/situations that need to be detected? Are there any other datasets that could be helpful?

The investigation of available data started with the research of the datasets from the Netherlands. Unfortunately, no publicly available datasets from the Dutch roads were found by me. Likewise, no dataset from the Netherlands is mentioned by Prachi in the section of her research related to possible datasets. On the other hand, there were several videos taken in the Netherlands by one of the AMI members, as well as the photos of the risky situations on the Dutch roads taken by one of Go2Sure's partners. However, those photos and videos were not labeled to use for the automatic testing of the model. In machine learning, data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it (What is data labeling?, 2022). The manual labeling of every image and video takes a lot of time, however, it's not part of any HBO-ICT competencies and is not related to software engineering. That's why after the consultation with the supervisors, the decision was made to search for the datasets from the neighboring countries that can be used to train and test the model. Likewise, it was decided to use the unlabeled Dutch data that Go2Sure owned to run it through the model and visually check if the developed model is able to work with the Dutch roads. Then the system can be further developed by Go2Sure when they collect and label enough data from NL.

So the next step was to study the available datasets outside the Netherlands. Several open-source datasets for computer vision such as Coco (COCO, 2022), ImageNet (ImageNet, 2022), and Open images (GoogleApis, 2022) were studied but the choice fell to the CityScapes dataset (Cityscapes, 2022). The CityScapes dataset is an open-source large-scale dataset that contains a diverse set of stereo video sequences recorded in street scenes from 50 different cities, with high-quality pixel-level annotations of 5 000 frames. The importance of the pixel-level annotation is described below in the Modeling part(3.3 Modeling). The following dataset contains two major advantages compared to the others. The first one is that it's designed specifically for solving the segmentation problems on the roads. Being very domain-specific, it contains the following list of classes :

Group	Classes
flat	road · sidewalk · parking ⁺ · rail track ⁺
human	person [*] · rider [*]
vehicle	car [*] · truck [*] · bus [*] · on rails [*] · motorcycle [*] · bicycle [*] · caravan ^{**} · trailer ^{**}
construction	building · wall · fence · guard rail ⁺ · bridge ⁺ · tunnel ⁺
object	pole · pole group ⁺ · traffic sign · traffic light
nature	vegetation · terrain
sky	sky
void	ground ⁺ · dynamic ⁺ · static ⁺

Figure 4 - Cityscapes classes

In the list below, the required objects that are present in the CityScapes dataset are marked green.

List of objects

List 2 - List of objects marked

1. Driving area
2. Sidewalk
3. Obstacle
 - a. Cemented block
 - b. Red_white striped indication
 - c. Cemented pole
 - d. Metal poles
 - e. Elevations
 - f. Road island
4. Vehicles
 - a. Cars
 - b. Trucks
 - c. Motor-bikes
 - d. Cycle
5. Non-drivable curbstone
6. Sky
7. Buildings
8. Trees
9. Pedestrians

On the other hand, it contains several other useful objects such as a “Fence” or “Bridge”.

Likewise, all of its data is collected in German cities, and a lot of those cities are very close to the border with the Netherlands. According to the road trip blogger (Differences between Netherlands and Germany road trip, 2019), the Netherlands and Germany are quite similar in the rules, signs, and roads except for a few situations. The picture below shows the cities from which the data was taken.

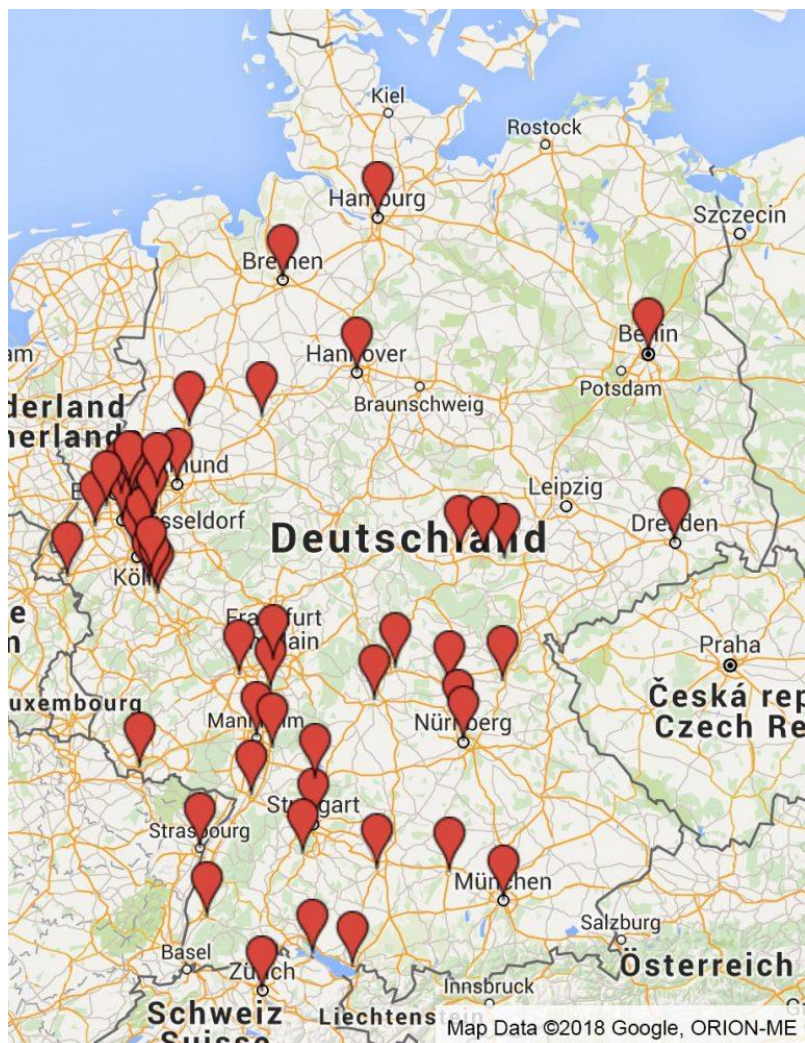


Figure 5 - Cityscapes dataset sources

That's why the assumption was made that the road pictures used in this dataset can be applied to the Netherlands.

For data preparation, the training-validation split provided by the CityScapes creators was used (The Cityscapes Dataset, 2020). The data is split into a proportion of 85% for training and 15% for validation of the publicly available data. According to the CityScapes team, the split ensures each split is representative of the variability of different street scene scenarios. So it was decided to rely on their opinion since they might know their data the best.

The research resulted in the dataset chosen for training and validation of the first version of the model. After that, the project moved to the next step of the CRISP-DM lifecycle – Modeling.

3.3 Modeling

The first step of the modeling phase was to define what type of machine learning algorithm is applicable for this case.

Research sub-question 3 - What type of machine learning algorithm needs to be implemented? What object detection and classification techniques need to be applied?

For answering this question Literature study technique was applied to research the theoretical background and usage of the applicable machine learning algorithms.

The very first way how machine learning can be applied to the image is the Image Classification. It's the process of assigning the label for the whole image. This approach is typically used for images that contain only one object or have one biggest object that dominates the picture. The ML model trained for the Image Classification would probably recognize the following picture as a dog:



Figure 6 Image classification example

Obviously, this method is not applicable for this project since the image retrieved from the biking track contains lots of objects and each of them needs to be detected.

The more complex way of processing the image is called Object Detection. This technique is able to detect the multiple objects within the image as well as identify their location. When Object Detection is applied to the image it can make it look as follows:

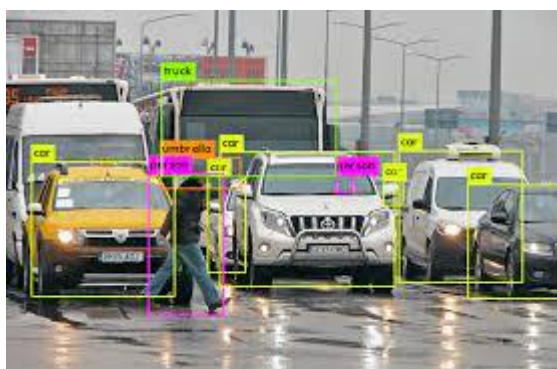


Figure 7 Object detection example

While this approach might be good enough for a single image, for a video, that changes multiple frames within the second, it might not be the best. The objects in the video are changing their locations, so the squares that identify the objects are constantly moving, which would make it difficult for the safety inspector to pay attention to everything that is going on on the track.

That's why the semantic segmentation technique was applied. Semantic Segmentation or image segmentation is the process of assigning a label to every pixel in the image. (Beyondminds.ai Research team, 2022). Label, in this case, is a class to which the object most probably belongs (eg. Vehicle, road, sky). The example of a semantically segmented image is presented above in the first chapter (Figure 1 – Semantic segmentation example). This technique is used for the autonomous driving solution, which has a very similar problem of

detecting the objects on the road, defining their shape, and determining whether these objects can produce a risk to the driver.

The next step after defining what type of algorithm will be used was to search for the existing technologies that are able to perform the semantic segmentation of the image.

Research sub-question 4 - Are there any open-source model architectures available that could be used as a starting point? Are there any relevant pre-trained ones?

To answer this question Library research techniques, such as Literature study and Best good and bad practices were applied. It helped to study the existing solution and go deeply into the theoretical background of the semantic segmentation models architecture.

The main principle of architecture for image segmentation is that it consists of an encoder and decoder. In semantic segmentation tasks, the goal is to extract features before using them to separate the image into multiple segments. An encoder is the convolutional neural network that is used to extract the features of the image. A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data (TechTarget, 2022). According to (Pooling Layers Explained, 2022), a convolutional layer slides a filter over the image and extracts features resulting in a feature map that can be fed to the next convolutional layer to extract higher-level features. Thus, stacking multiple convolutional layers allows CNNs to recognize increasingly complex structures and objects in an image. The main problem of this approach is that the produced feature map is location dependent, which means that as it is it would be able to recognize the object only in the specific location. That's why pooling layers are added after the convolutional ones. Pooling layers operate with the feature maps produced by convolution by applying the smaller filters to them. With every additional layer, the map shrinks, preserving only the important information about the presence of the features of interest. As the map becomes small, it becomes increasingly independent of the location of the feature (Pooling Layers Explained, 2022). The diagram below visualizes the Convolution-Pooling working principle(Figure 8 Convolution-Pooling working principle).

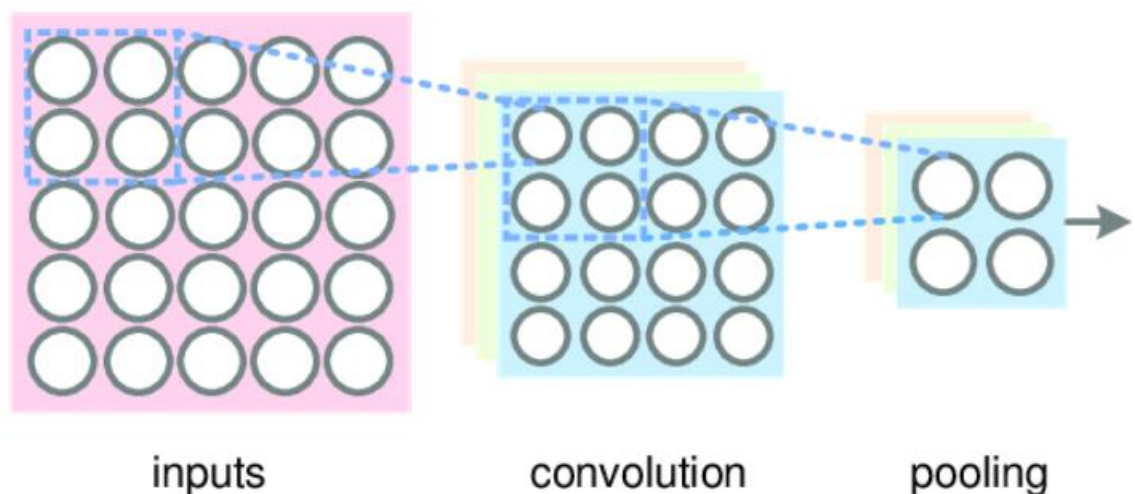


Figure 8 Convolution-Pooling working principle

This approach allows extracting the detailed, location-independent features from the image. However, the image is reduced in size while passing through convolutional-pooling layers. After that, in order to efficiently separate the image into multiple segments, it needs to be upsampled back. That's achieved by the usage of the deconvolutional layers, which perform the reverse operation with the feature maps created by downsampling. The network used for upsampling is called a decoder.

There are several architectures exist that solve semantically segmented tasks by applying the approach above. According to the research conducted by Cristian Anghel (Anghel, 2021) - a previous AMI and Go2Sure intern from the University of Twente, a state-of-art deep learning model for semantic image segmentation is the DeepLab. The main competitors of the Deeplab architecture are FCNs(Fully Convolutional Networks). According to the research conducted at Stanford University (Peggy (Yuchun) Wang, 2021), the most recent version of DeepLab – DeepLabv3+ slightly outperforms the compared FCNs. The other popular type of architecture for semantic segmentation is called U-Net. This network was initially developed for segmenting biomedical images. Despite the fact that no research on the comparison between the DeepLab and U-Net was found, the Deeplab architecture was chosen because it's more widely used in road image segmentation while the U-Net appears to be more domain specific for medical science. The ability of DeepLab to deal with road image segmentation and autonomous driving was confirmed by the research of Cristian Anghel (Anghel, 2021) for Go2Sure, researchers from Stony Brook University (Karim, Li, Qin, & Yin, 2021), and the research created in collaboration between researchers from Chinese Universities (Li, et al., 2021). The figure below represents the architecture of DeepLabv3+. This specific example includes ResNet101 as an encoder, however, any other supported encoder can be used with the same principle.

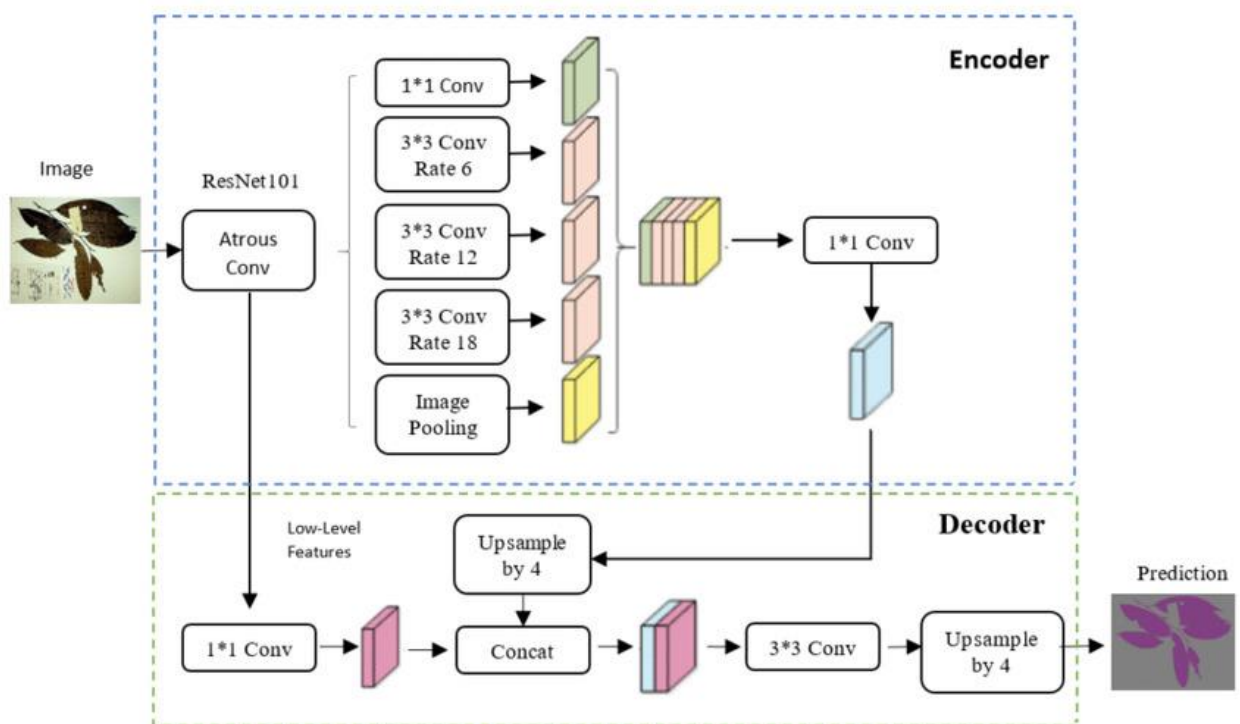


Figure 9 - DeepLabV3+ architecture (ResearchGate, 2022)

The outcome of the research was the decision on using the DeepLabv3+ as a model architecture for the first CRISP-DM cycle of this project. After that, the work on the implementation and evaluation of this model started.

3.4 Model Implementation and Evaluation

At this stage, the actual development is started. The first step was to implement the DeepLab architecture. Since DeepLab is a widely used architecture for semantic segmentation, the choice was to first try to find and run the ready-made open-source implementation of it. During this research, the Pytorch (Pytorch, 2022) implementation of it was found, which was developed by Gongfan Fang (Fang, 2022). The choice was made for this implementation because it already had 2 different encoders pre-trained on the Cityscapes dataset. The first one was the MobileNet, the second is the ResNet101 model. So the question of how to choose between those two encoder options appeared.

Research sub-question 5 - What are the relevant metrics to assess the quality of the model for this case?

To answer this question the machine learning and semantic segmentation metrics were researched. The accuracy metric, which is quite common for machine learning projects, was not taken into account in this case. In semantic segmentation, the accuracy rate is the percentage of pixels identified correctly. However, the main goal of the project is to identify the risks for the bikers, which usually take the least part of the picture. For example, the model can correctly identify the road and the sky but not recognize the obstacle on the road. In this case, the accuracy will be high but the job won't be done.

That's why two other metrics were found which provide the balance between the labels presented in the picture.

Intersection-Over-Union (IoU, Jaccard Index)

IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. (TowardsDataScience, n.d.). To validate the performance of the dataset the mIOU (mean-intersection-over-union) is used, which is the average IOU of the segmented objects over the entire validation split.

Dice Coefficient (F1 Score)

The dice Coefficient is $2 * \text{the Area of Overlap}$ divided by the total number of pixels in both initial and segmented images.

Those two metrics use similar approaches and are positively correlated with each other. If one of those metrics shows that classifier A is better than classifier B, the other will show the same. The difference comes when identifying how much classifier A is better than classifier B. The Dice Coefficient tends to measure something closer to average performance, while the IoU score measures something closer to the worst-case performance.

Thus, the mIOU metric was chosen to assess the quality of the model and determine whether it satisfies the business requirement. That's because if mIOU satisfies the minimum need, then the Dice Coefficient would satisfy it for sure.

Using the outcomes of this question, the encoders mentioned above could be compared. The MobileNet one had an mIOU of 72.1% over the validation split of the CityScapes dataset, while the ResNet101 model had 76.2%. The advantage of the MobileNet compared to the other networks is that it is a smaller and faster network designed for mobile devices. The table below shows the advantages of those two types of encoders.

Table 2 MobileNet – ResNet101 comparison

Model	Speed	mIOU
MobileNet	✓	
ResNet101		✓

Since this system was planned to be run on the server-side, thus the choice was made for the more complex but more accurate ResNet one. ResNet is a convolutional neural network, where the number in the name stands for the number of layers in it, which is 101 in this case.

The chosen implementation was downloaded and set up according to the instructions.

After that, this implementation was tested manually on the available data from the Netherlands. It appeared to deal well with the Dutch road, which confirmed an assumption that the chosen German dataset was applicable for the case in the Netherlands. The two photos below illustrate the Enschede road situation and the result of semantic segmentation of it.



Figure 10 - Enschede road situation

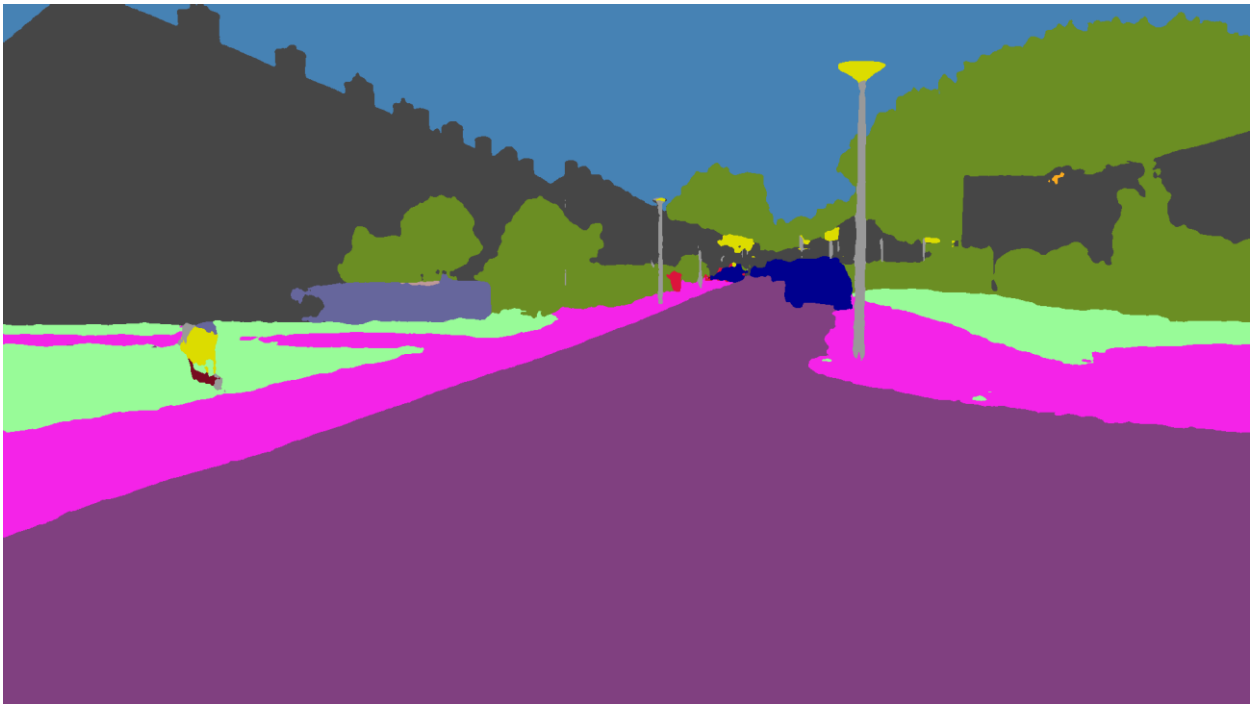


Figure 11 - Enschede road situation segmented

That conducted the first cycle of the CRISP-DM. The model above was already satisfying the requirements but still had room for improvements. However, the model itself could only work with single photos. That's why it was decided to first implement the system that would be able to semantically segment the video.

Chapter 4 – System implementation

This chapter describes the process of implementation of the system that is able to handle the video and analyze it using the developed machine learning solution according to the requirements.

4.1 First version of the system

At first, it was needed to identify the high-level architecture of the system.

Research sub-question 6 - What is the optimal architecture/communication flow within the system?

The research technique applied for this question was IT architecture sketching.² During this session, the ideas regarding the system architecture were defined and discussed with supervisors.

The high-level system architecture was defined. This system takes a video input, analyzes it using machine learning, and outputs the semantically segmented video with the risks predicted. At the moment of writing this report, there is no machine learning algorithm that could be applied directly to the video. That's why the video should be split into frames(images). Then each photo is evaluated and transformed individually. After that, the system compiles the semantically segmented frames back to the video. The video is taken from the local storage and the transformed one is saved locally. Likewise, it could be taken from the SportKeur system when it's integrated there. The diagram below shows the architecture of the software system.

²Part of the workshop research method, aimed to define the high-level IT architecture.

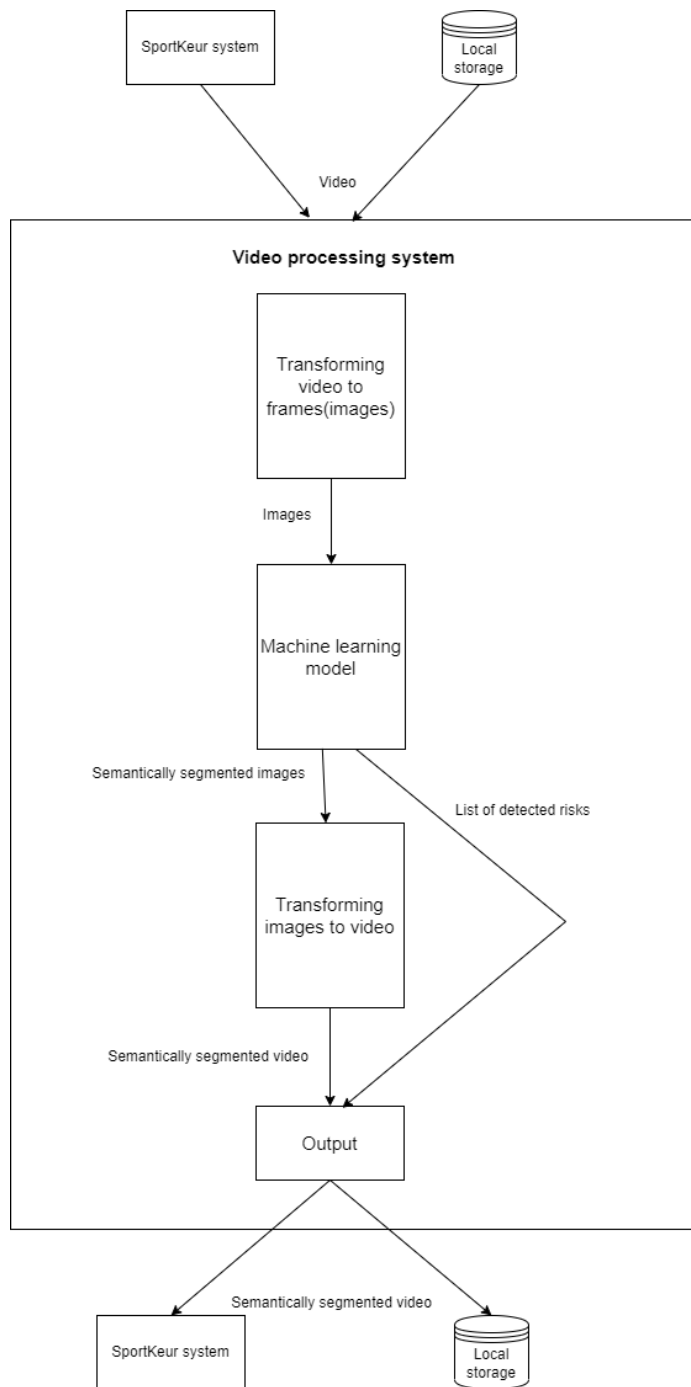


Figure 12 - System architecture

The next step after defining the architecture was choosing the technologies that are able to satisfy the requirements and be integrated and maintained within the SportKeur system.

Research sub-question 7 - What technology stack is optimized to fit the requirements? What technologies are optimal to process and output the video in a needed way and be able to integrate it into the existing SportKeur stack?

To answer this question two Library research techniques were applied. The first one was the Expert interview³ with one of the Go2Sure developers. During this interview, the current and planned Go2Sure technology stack was discussed and preferred technologies for this project were defined. After that Best good and bad practices⁴ research was conducted to determine whether the system could be implemented using suggested technologies.

During the interview with the Go2Sure developer (Hooijer, 2022) it was found out that the company uses Python programming language for such types of projects and is planning to move towards a new backend using Django. The usage of those technologies was not strictly required by the company but it was preferable for the developers if there is a possibility for that. That's why the best practices for such projects were researched keeping in mind the preferences of the company.

After that, the research on best practices for ML projects was conducted. According to various researches, Python is considered the best option for machine learning. The main reason for it is that the community around the available tools and libraries makes Python particularly attractive for workloads in data science, machine learning, and scientific computing. According to a recent KDnuggets poll that surveyed more than 1,800 participants for preferences in analytics, data science, and machine learning, Python maintained its position at the top of the most widely used language in 2019 (Sebastian Raschka, 31 Mar 2020). Various experts also claim Python is the most advantageous programming language for ML and Data Science. Josh Tobin states that Python allows you to iterate faster (in ML) than any other language (Tobin, 2022). According to Tareek Amr, Python is indeed ML's lingua franca. It is flexible, easy to read, and as a non-compiled language, it is suitable for quick iterations. It also became entrenched deeper into the field due to the presence of different ML toolings such as Scikit-Learn, TensorFlow, and Pytorch (Amr). That confirms Python to be the best, as well as the most widely used technology for ML nowadays.

After that, the research on the optimal Python framework was conducted. The diagram below shows that the most widely used Python frameworks are Flask (Flask, 2022) and Django (Django, 2022).

³ Research approach used to gain the knowledge from the expert in the field.

⁴ Method applied to research the best practices and approaches used in a similar high-quality projects.

Which One Is The Best Python Web Framework?

What web frameworks / libraries do you use in addition to Python?

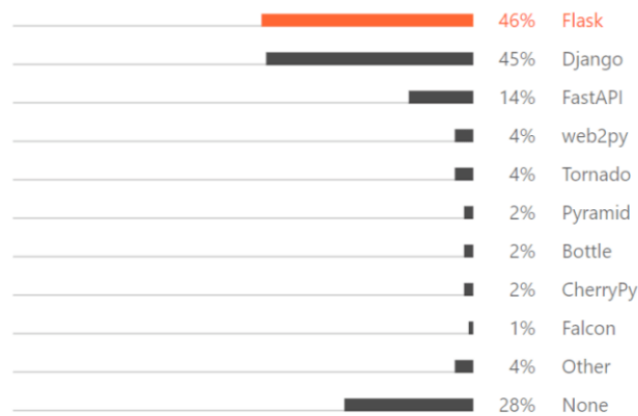


Figure 13 Most popular Python frameworks

So the comparison between those two options was conducted to determine which one suits better for the following project. Four criteria were outlined to compare the frameworks:

- Usability for such types of projects
- Performace
- Size of community, number of learning materials, etc.
- Scalability

From the usability point of view, both Django and Flask appeared to be used for machine learning-related projects. According to the research conducted by Gaurav Sharma (Sharma, 2022), Django is better for bigger and more complex machine learning projects while Flask is more suitable for smaller ones. That makes Flask a better option for the current project by this criteria, however, it could cause problems if the Risk Management tool would be built by Go2Sure on top of this project.

Performance-wise, Flask has a slight advantage over Django. Flask framework is lighter than Django and performs better with negligible differences (Django Vs Flask, 2022). Both frameworks are considered to have a high performance.

From the community and learning materials perspective, Django has a bit wider range of options than Flask. That's because it was invented five years earlier, thus had more time to develop that community. However, community support for both frameworks is wide enough to get the needed information.

Despite this project being relatively small, it can become a basis for the SportKeur Risk Management tool, thus scalability is important. Django offers better scalability and is designed for more high-traffic solutions, while Flask is considered to be a lighter framework for smaller projects.

Table 3 Django vs Flask

	Django	Flask
Usability	✓	✓
Performance		✓
Community	✓	✓
Scalability	✓	

The research above shows that Django and Flask are the approximately equal choice for this project. Flask might be a bit better solution for this project itself but Django offers more possibilities for it to be extended in the future. Likewise, as mentioned above, Django is preferred more by the company. That's why Django was chosen as a framework for it.

Afterward, the Django app was created, the following DeepLabV3+ implementation was added to it, and the scripts to deal with the videos were created. Those scripts work as described in the design above – taking video from the local storage, splitting it into the frames, analyzing each frame individually by the machine learning implementation, and converting the frames back to the video. The example of one of the videos from Dutch road, that was semantically segmented by the system is attached to the Graduation Portfolio. To execute the system, the user should provide at least the location of the video to be analyzed. Optionally the location to save video, as well as the location to save the frames can be provided, otherwise, the system uses the default location.

At this stage of the project, all the Must and Should requirements were satisfied, so the discussion on what to do next was conducted with Raoul. There were two options on the table:

1. Try to improve the machine learning part by conducting more CRISP-DM cycles.
2. Develop the API that would allow the integration of the system into the SportKeur project.

Raoul mentioned that the second option was more prior for him at that stage, so the work on the integration started.

4.2 Integration to SportKeur system

In order to start the integration, the SportKeur system workflow was discussed with the Go2Sure developers. During this session, the idea of how my system should be connected to the main one was defined. The defined communication flow is provided below (Figure 14 - Communication with SportKeur system). In the SportKeur system, the driving inspector is able to upload the video from the track using the provided UI. After the video is uploaded to the system it's used in several cases, one of which would be the semantic segmentation. So then it's sent to my system which processes it and starts the segmentation. Whenever the user wants to watch the segmented video the SportKeur system tries to retrieve the analyzed video from my part. If the semantic segmentation is finished by this time, it responds to the system with the video, otherwise, it asks to wait until the process is finished and the SportKeur UI displays the message to the user.

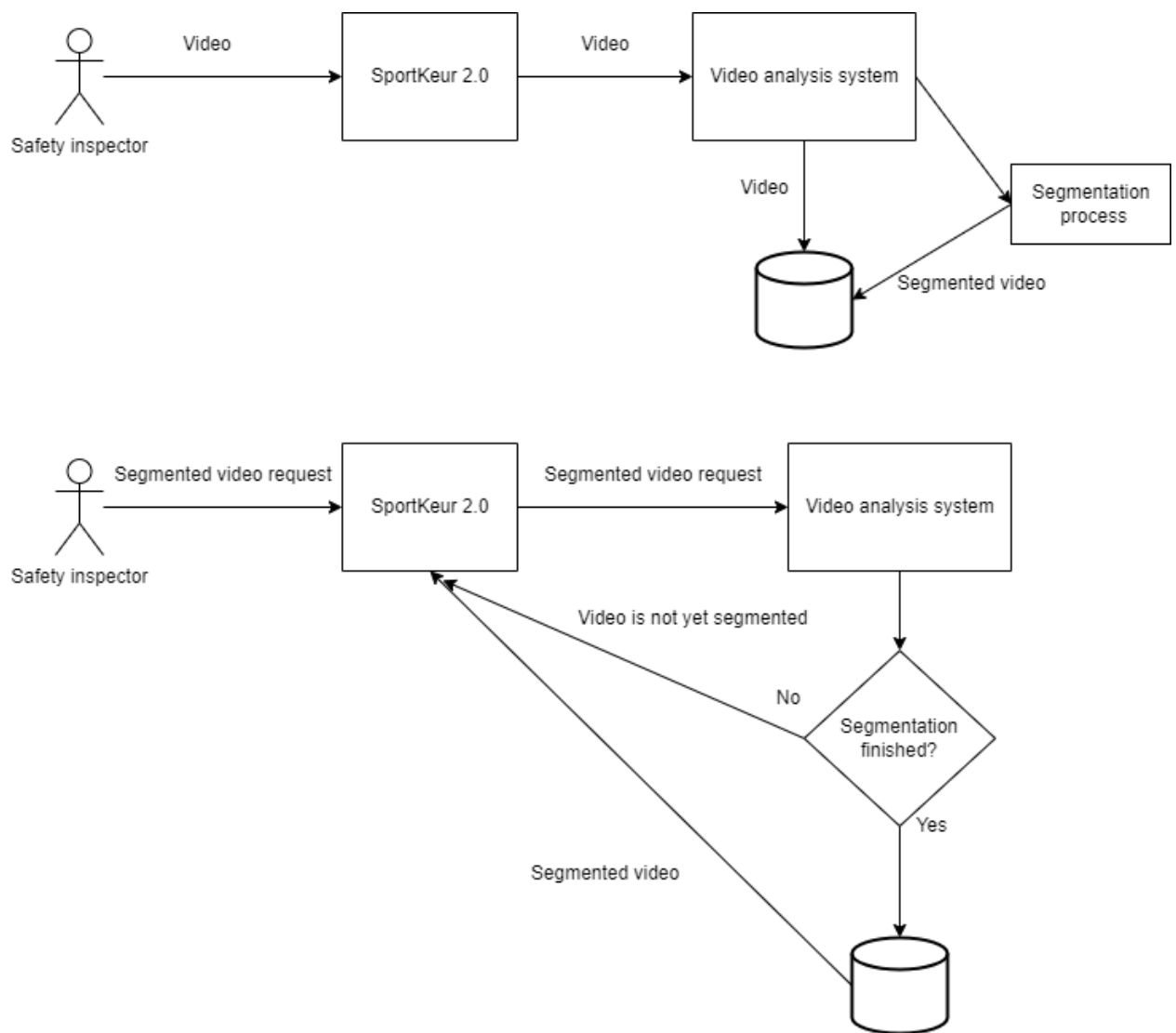


Figure 14 - Communication with SportKeur system

The system described above was implemented using the Django REST (Django REST, 2022) API calls.

Chapter 5 – Evaluation

This chapter is aimed to go through the requirements that were defined at the beginning of the project and show how they were fulfilled in the final delivery.

To start with, the machine learning model that is able to perform semantic segmentation was set up. As it was mentioned above, it shows the 76.2% mIOU on the CityScapes dataset which satisfies the initial requirement of being able to give accurate prediction in 75% of the cases on average. Likewise, the model was tested with various photos and videos from the roads in the Netherlands. Unfortunately, the mIOU percentage can not be taken from this test because, as was mentioned, the data is not labeled. However, visually it seems to deal well with the Dutch roads and this part can be tested automatically by Go2Sure as soon as they have enough labeled data available. The model is able to detect almost all of the objects from the list defined at the beginning, as well as the additional ones that the dataset contained, which also adds value to the segmentation. The photos and videos segmented by the model were presented to Go2Sure and the AMI supervisor. The Acceptance Test on the quality of prediction on the Dutch road was conducted with Raoul to ensure the model is able to deliver the expected result. Raoul has confirmed that the quality of semantically segmented videos is suitable for this stage of the SportKeur project.

From the system part, the software that is able to get the video input, split it into images, analyze it using the model, convert it back to the video, and output the semantically segmented video back to the user was developed. Likewise, the API for integration of that system to the SportKeur 2.0 core was developed, which was initially the “Could” requirement. The system and the API request were covered with the Unit Tests to automate the process of testing the system parts.

The “Would” requirement of classification of the risk level was not fulfilled because of the lack of time. However, it was initially planned as outside of the scope of this assignment. Go2Sure developers can extend my system with this feature while developing the Risk Management Tool in the future.

Conclusion

To conclude, the project can be considered successful because it solves the main problem and reaches the goal of the assignment. During the project, the problem and the goal were defined and scoped based on the problem research done by previous AMI and Go2Sure employees. After that, the requirements for the project were gathered, the available data were analyzed and a suitable dataset was chosen. Following, the applicable computer vision techniques were researched as well as a suitable machine learning architecture was chosen. The machine learning model was set up and evaluated using the validation set as well as the data from the Netherlands. Afterward, the video handling system was built on top of it that is able to take the video input, analyze it using the ML model, and output the semantically segmented video. Furthermore, the tool is integrated into the existing SportKeur system using API. It creates a possibility for Go2Sure to create the Risk Management tool on top of it to automate the process of safety inspections of biking tracks. Meanwhile, the system can be used to give the inspector an option to perform an additional check of the track using the semantically segmented video. Likewise, the project has room for improvements that Go2Sure

developers or future interns could implement. The recommendations for it are described in the next chapter.

Recommendations

The very first recommendation would be to collect and label the data from the Netherlands. The German dataset has proven the concept but I assume that with the Dutch data the results could be much better. Despite the German roads being similar to Dutch ones, they have some differences, especially in biking infrastructure. For example, Dutch roundabouts involve bike paths, pedestrians, and cars all at once, which is not the case in Germany (Differences between Netherlands and Germany road trip, 2019). It would be good to train the model in such situations as well. As far as I know, Go2Sure has partners that can be involved in this process and supply the needed photos or videos. Afterward, the collected data should be manually labeled. Ideally would be to get enough data with all the needed classes and risky situations to completely retrain the model or improve it with more data. Alternatively, if the company or the research group would not be able to get that much data, they could collect and label the amount that is enough for testing. Using this data, the involved data scientists can compare different models and architectures trained on the other datasets, such as the CityScapes.

The next step for this project should be to build the actual risk analysis tool to automate the process. The current system is able to detect different objects on the track, so the risk-management system can be built on top of it. This tool can analyze if the objects on the road are a risk for the bikers and alarm the inspector in case it has some level of confidence in it. Likewise, it can classify risks by the level of danger for the rider.

List of figures

Figure 1 – Semantic segmentation example	5
Figure 2 CRIPS-DM	7
Figure 3 - GitHub board	8
Figure 4 - Cityscapes classes	13
Figure 5 - Cityscapes dataset sources.....	15
Figure 6 Image classification example	16
Figure 7 Object detection example	16
Figure 8 Convolution-Pooling working principle	17
Figure 9 - DeepLabV3+ architecture (ResearchGate, 2022)	18
Figure 10 - Enschede road situation	20
Figure 11 - Enschede road situation segmented	21
Figure 12 - System architecture.....	23
Figure 13 Most popular Python frameworks	25
Figure 14 - Communication with SportKeur system	27

Version control

Date	Version	Description	Author
08.04.2022	0.1	The draft of the document for the first feedback session.	Andrii Roizin
05.04.2022	0.2	Feedback from the first feedback session was applied. Draft version for the interim session.	Andrii Roizin
27.05.2022	0.3	Feedback from the interim session was incorporated. The concept version of the report.	Andrii Roizin
10.06.2022	1.0	Feedback from the concept version was incorporated. The final version of the report.	Andrii Roizin

References

- (n.d.). Retrieved from TowardsDataScience: <https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2>
- AMI. (2022). *Ambient Intelligence*. Retrieved from Saxion.nl: <https://www.saxion.nl/onderzoek/smart-industry/ambient-intelligence>
- Amr, T. (n.d.). Is Python the best language for machine learning?
- Anghel, C. (2021). *Effect of Image Quality in Computer Vision for Semantic Segmentation of Road Images*. Enschede: University of Twente.
- Bagave, P. (2021). *Object detection and recognition in Sportkeur project*. Enschede: AMI Saxion research group. Retrieved from

https://saxion.data.surfsara.nl/index.php/apps/files/?dir=/video_analysis&fileid=39522546

Beyondminds.ai Research team. (2022). *A simple guide to semantic segmentation*. Retrieved from beyondminds.ai/: <https://beyondminds.ai/blog/a-simple-guide-to-semantic-segmentation/>

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (n.d.). *Rethinking Atrous Convolution for Semantic Image Segmentation*.

Cityscapes. (2022). Retrieved from <https://www.cityscapes-dataset.com/dataset-overview/#labeling-policy>

COCO. (2022). Retrieved from <https://cocodataset.org/#home>

DeepLab: Deep Labelling for Semantic Image Segmentation. (n.d.). Retrieved from github.com: <https://github.com/tensorflow/models/tree/master/research/deeplab>

Differences between Netherlands and Germany road trip. (2019).

Django. (2022). Retrieved from <https://www.djangoproject.com/>

Django REST. (2022). Retrieved from <https://www.django-rest-framework.org/>

Django Vs Flask. (2022). Retrieved from <https://www.softwaretestinghelp.com/django-vs-flask-vs-node/>

Fang, G. (2022). *DeepLabV3Plus-Pytorch*. Retrieved from github.com: <https://github.com/VainF/DeepLabV3Plus-Pytorch>

Flask. (2022). Retrieved from <https://flask.palletsprojects.com/en/2.1.x/>

Go2Sure. (2022, 03 30). *go2sure.nl*. Retrieved from WIE ZIJN WIJ: <https://go2sure.nl/>

GoogleApis. (2022). Retrieved from <https://storage.googleapis.com/openimages/web/index.html>

Hooijer, D. (2022). (A. Roizin, Interviewer)

ictresearchmethods.nl. (2021). Retrieved from https://ictresearchmethods.nl/Main_Page

ImageNet. (2022). Retrieved from <https://www.image-net.org/>

Karim, M. M., Li, Y., Qin, R., & Yin, Z. (2021). *A SYSTEM OF VISION SENSOR BASED DEEP NEURAL NETWORKS*. Stony Brook: Stony Brook University.

Li, J., Jiang, F., Yang, J., Kong, B., Gogate, M., Dashtipour, K., & Hussain, A. (2021). *Lane-DeepLab: Lane semantic segmentation in automatic driving*.

MoSCoW_method. (2022). Retrieved from Wikipedia.org: https://en.wikipedia.org/wiki/MoSCoW_method

- Peggy (Yuchun) Wang, K. A. (2021). *Semantic Image Segmentation for Autonomous Driving Scenarios Combining*. Stanford: Stanford University.
- Pooling Layers Explained*. (2022). Retrieved from <https://programmatically.com/what-is-pooling-in-a-convolutional-neural-network-cnn-pooling-layers-explained/>
- Pytorch. (2022). Retrieved from <https://pytorch.org/>
- ResearchGate. (2022). *DeepLabv3-architecture-Conv-Convolutional-layer*. Retrieved from [www.researchgate.net: https://www.researchgate.net/figure/DeepLabv3-architecture-Conv-Convolutional-layer_fig1_335516870](https://www.researchgate.net/figure/DeepLabv3-architecture-Conv-Convolutional-layer_fig1_335516870)
- Sebastian Raschka, J. P. (31 Mar 2020). *Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence*. Madison: University Ave.
- Sharma, G. (2022). *Which is Better for Machine Learning: Flask vs Django?* Retrieved from <https://www.analyticsvidhya.com/blog/2022/01/which-is-better-for-machine-learning-flask-vs-django/>
- TechTarget. (2022). *Convolutional neural network*. Retrieved from <https://www.techtarget.com/>: <https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network>
- The Cityscapes Dataset*. (2020). Retrieved from <https://github.com/mcordts/cityscapesScripts/blob/master/README.md>
- Tobin, J. (2022). Is Python the best language for machine learning?
- What is CRISP DM?* (2022). Retrieved from <https://www.datascience-pm.com/>: <https://www.datascience-pm.com/crisp-dm-2/>
- What is data labeling?* (2022). Retrieved from [aws.amazon.com: https://aws.amazon.com/ru/sagemaker/data-labeling/what-is-data-labeling/](https://aws.amazon.com/ru/sagemaker/data-labeling/what-is-data-labeling/)
- Why Is CRISP-DM Gaining Grounds*. (2022). Retrieved from <https://analyticsindiamag.com/why-is-crisp-dm-gaining-grounds/#:~:text=CRISP%2DDDM%20is%20designed%20to,which%20other%20methodologies%20are%20benchmarked.>
- Willemsen, R. (2022). (A. Roizin, Interviewer)
- Zhang, R., Du, L., Xiao, Q., & Liu, a. J. (n.d.). *Comparison of Backbones for Semantic Segmentation*.