Automatic real time detection and localisation of fisher boats in restricted areas with the use of drones

MSc Geomatics Thesis Proposal

Lisa Geers
5351421

1st supervisor: Martijn Meijers
2nd supervisor: Azarakhsh Rafiee
External supervisor: Niels van der Vaart

June 9, 2022

Contents

1 Introduction

Fishing is an important part of the Dutch economy. Despite the ban on pulse fishing, the Dutch fishing industry still yielded 213 million euros in 2020 (H. van Oostenbrugge, 2021). However, fishing can be harmful for several reasons, so it is not permitted to fish in all Dutch waters. In de Dutch inland waters, there are a number of restricted areas where no fish can be caught because of the high dioxin degree in the water, which could have harmful effects when this fish is consumed (Leeuwen et al., 2002). Moreover, both in the Dutch inland waters and in the sea, there are a number of areas where fishing is restricted due to the laws of the Natura 2000 agreements (J. van Oostenbrugge et al., 2010). In these areas, fishing is restricted to conserve habitats and species named in the EU Birds and Habitat directives (Pedersen et al., 2009).

The monitoring of fishers and fisher boats in restricted areas is done by the Netherlands Food and Consumer Product Safety Authority (NVWA). Inspections on boat equipment and fish catch are currently done by boat. Naturally, this is very labour intensive, as this requires at least two inspectors per boat. Additionally, using this method, the checking speed is not very high. Nevertheless, in 2016, 80 cases of illegal fishing in inland waters were reported on eel fishing alone by the NVWA (Bos, 2018).

To increase efficiency and accuracy, a new method needs to be developed that can automatically find fishing boats in restricted areas. By using drones with imaging and GPS equipment, fishing boats can be automatically detected with AI and the exact boat location can be calculated using the drone GPS, followed by an intersection with the restricted areas. In this way, more fishing boats in illegal areas are expected to be found. This method needs to be carried out in real time, because visual inspection by inspectors is still needed to fine or prosecute offending fishers.

Automatic detection of objects with the use of drones has been researched before. For example for the detection of palm trees (Aburasain et al., 2020; Htet & Sein, 2021) and cattle (Shao et al., 2020). Zhang et al. (2019) developed a method to track and localise objects in 3D. Moreover, Prayudi et al. (2020) researched a method to detect and localise fisher boats and derive hull plates from drone images. However, no research has been done to carry detection and localisation out in real time, which is a crucial aspect to catch offending fishers in the act.

The objective of this proposed graduation project is to develop a prototype for a new method that can detect and localise fishing boats in real time with the use of drones. The focus area of this research will be the Dutch inland waters, as data acquisition and testing is more accessible there compared to on sea. Drone image data will be acquired from several inland water areas in the Netherlands. Using this data, a prototype to detect and localise fisher boats in real time will be constructed in Python using a variety of libraries. To evaluate the developed method and prototype, several field tests will be conducted.

The graduation project proposal has the following outline. In the next section, a literature review of the related work is illustrated. In section 3 the research objectives and scope are described by means of a description of the research questions and a MoSCoW analysis. In section 4 the proposed methodology is explained and in section 5 the time planning of the graduation project is given. Finally, in section 6 an overview of the tools and datasets used in the research will be provided.

2 Related work

In this section, the relevant literature for this research project is presented.

2.1 Machine learning

Machine learning is a type of artificial intelligence where machines are able to learn automatically and improve from experience without being explicitly programmed to (Nan, 2022a). Machine learning consists of many different subcategories that are fit for a variety of different purposes. These categories can be split into supervised and unsupervised learning. In unsupervised learning, labelling and pattern finding is performed without training. The main purpose of unsupervised learning is to find patterns in unlabelled data. In supervised learning, input-output pairs are given, which makes the algorithm able to find the correct output for an input through training (Müller & Guido, 2016). The supervised learning algorithms can then be further subdivided into regression and classification algorithms. Regression algorithms are used to predict continuous variables and classification algorithms can predict discrete classes (Müller & Guido, 2016). For the detection of boats, a classification algorithm is needed with a boat category.

A type of supervised classification algorithm is an artificial neural network, which is based on neurons of the biological nerve system. Artificial neural networks algorithms consist of a collection of connected nodes.The basic structure can be seen in Figure 1. It consists of an input, hidden and output layer. The input is a multidimensional vector, for example one containing image pixel values. The hidden layers make decisions based on the previous layer. In the process of learning, the effects of these decisions on the output are evaluated and the decisions are improved consequently. An architecture with multiple stacked hidden layers is called deep learning (O'Shea & Nash, 2015).

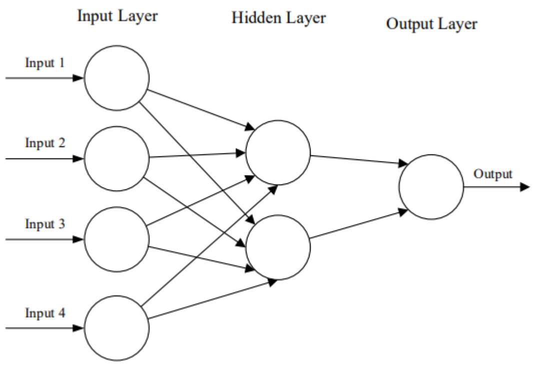

Figure 1: Simple artificial neural network. From O'Shea & Nash (2015)

Convolutional neural networks (CNNs) are artificial neural networks with an architecture created for image processing. CNNs contain a variable set of modules that transform the input at one level into a representation of a higher and more abstract level. With enough of these transformations, complex functions can be learned. The key aspect is that these transformations are not designed by humans, but are learned by the computer from test data (LeCun et al., 2015). The structure can be seen in Figure 2. CNNs consist of three types of hidden layers (O'Shea & Nash, 2015):

- Convolutional layer: Uses an elementwise activation function applied by a rectified linear unit (ReLu) on the output of the previous layer.
- Pooling layer: Downsamples the spatial dimensionality of the input of the previous layer, thereby reducing the number of parameters.
- Fully connected layer: Produces class scores from the activations of the previous layer.

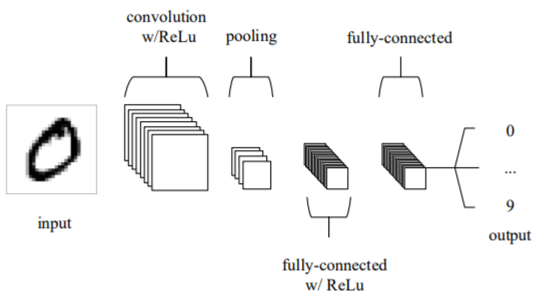

Figure 2: Simple convolutional neural network. From O'Shea & Nash (2015)

Because CNNs are structured to process image data, they can be used well to detect objects from drone data. Classification of objects in drone images has been researched before to find palm trees by Aburasain et al. (2020) and Htet & Sein (2021). Aburasain et al. (2020) used the YOLO-V3 object detector to detect palm trees of different resolution, sizes, spread and degree of overlap. They stated that the benefit of using drone images instead of satellite images is the increased image resolution, which leads to more accurate classification.

Detecting ships with the use of deep learning has been researched several times, for example by Apoorva et al. (2021), Chang et al. (2019) and Voinov (2020). Input images in these researches are imagery data collected by satellites. Apoorva et al. (2021) uses the TensorFlow deep learning library to detect ships on satellite images. Voinov (2020) proposes a method to detect vessels from optical satellite images using images from three different satellites. It is mentioned that the image resolution has a large impact on the detection accuracy. Chang et al. (2019) proposes a method to detect ships on synthetic aperture radar imagery. By using the YOLOv2-reduced deep learning architecture, a significant decrease of the computation time and a detection accuracy of 90% was achieved. Nonetheless, none of these methods used drone data and none implemented a real time or localisation functionality.

2.2 Object localisation

One of the most fundamental problems in multiple view geometry is triangulation, which is the process of retrieving a 3D coordinate in the world coordinate system from its two local coordinates in two 2D images (Hartley & Zisserman, 2003). To calculate this, the camera intrinsic and extrinsic parameters are used. There exist five intrinsic parameters: two for focal length (α_x, α_y), one parameter for the skew coefficient between the x and the y axis (γ) and two principal point offset parameters (u_0, v_0). Together, they form the intrinsic matrix K. The extrinsic parameters consist of a rotation matrix R and a translation vector T. The extrinsic parameters form the pose of the camera in relation to the world reference system. As can be seen below, the intrinsic and extrinsic parameters together form the camera matrix M. The last equation below displays how a point P in the 3D world coordinate system maps to point p' in 2D on the image plane (Nan, 2021a).

$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$M = K[RT]$$

$$p' = MP$$

The triangulation problem can be seen at Figure 3. Here, R and T are the camera extrinsic parameters and K and K' are the intrinsic parameter matrices of the two images. P is the 3 dimensional point that should be found and p and p' are the corresponding 2D points on the image planes. O_1 and O_2 are the camera origins. In theory, P could be calculated through the intersection of l and l', the two lines of sight. However, because observations are noisy and the camera parameters are not precise, finding this intersection point is problematic Nan (2021b).

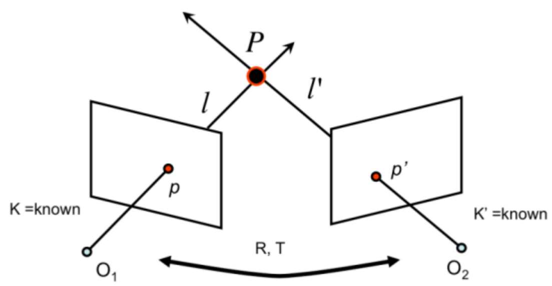

Figure 3: Triangulation problem with two images. From Nan (2021b)

A solution for this problem could be using more than two images. The structure from motion method (SfM) uses multiple views to determine the geometry of the scene and the camera parameters simultaneously. It consists of determining corresponding features in images and subsequenty determining the motion of a feature in each image Nan (2021b)

Bodensteiner et al. (2015) proposes a method to accurately georeference community drone photo and video data to register to 3D LiDAR data. This method relies on the SfM and SLAM techniques, which are combined with appearance and structure matching based on LiDAR data. The method can produce drift free drone image overlays at a speed of 30 frames per second with an average error of 0.4 meters. This method uses existing open drone data for processing and testing. However, it is not runtime optimised and real time execution has not been implemented.

Prayudi et al. (2020) designed a surveillance system framework that can detect and localise ships and derive their hull plates. They used 2450 images for training and stated that ship detection can be challenging due to the difficulty of collecting training images. Nonetheless, the found average matching precision was 96%. The ship coordinates were found by taking the x and y image coordinates from the object bounding box and converting them to latitude and longitude coordinates. It is not clear what algorithm or library was used for this conversion.

3 Research objectives

The objective of this research is to develop a prototype for a new method to catch fishing boats in restricted waters with the use of drones. Boats need to be detected on drone image streams, and the exact location of these boats needs to be derived to be able to determine if they are in restricted waters or not. This prototype also needs to be able to carry this out in real time. The area of interest is the Dutch internal waters, because there are a variety of restricted areas and data can be collected there more easily than on the sea.

3.1 Research questions

To fulfil this objective, the following research question was defined:

To what extent can drones be used to catch fishing boats in restricted areas in real time by means of automatic detection and localisation?

To answer the main question, the following sub-questions were defined:

- How can deep learning be used to detect fisher boats?
- How can detected objects be localised in a geographical coordinate system?
- What hardware and software is needed for this method to be carried out in real time?

3.2 Scope

The scope of this research is displayed in Table 1 by means of a MoSCoW analysis. In this analysis, tasks are divided into four categories. The tasks in the Must have category are critical to the research outcome and should be carried out. The Should have category is also necessary for the research, but is not as critical because results can still be obtained without these items. The Could have category contains tasks that are interesting and research on them could improve the results, but they are not necessary to obtain a good outcome. The Will not have category details tasks that are interesting to research, but lay outside the scope of the research.

The aim is to carry out the tasks in the Must and Should have category. Therefore, the tasks in the Must and Should have categories are both included in the methodology of this research. The tasks in the Could have category will be carried out if there is still time left after completing the necessary tasks. The items in the Will not have category will not be carried out and are outside the scope of this research.

The challenges of this research lie in the localisation of the fisher boats and the real time aspect of the method. For this reason, different machine learning methods, like decision trees and support vector machine, to detect of boats will not be researched. Also, besides testing the performance of different deep learning models, no time will be spent on improvement of the detection accuracy, as the focus of this research lies on the object localisation and real time aspects. Additionally, determination if detected boats are fishing or not will be out of the scope of this research.

The detection of boats with deep learning, localisation of objects, the real time aspect of these methods and the result evaluation is critical for the outcome of the results, hence they are in the Must have category. Other tasks that are non-critical but should be carried out are the creation of a dashboard with a map on which the restricted areas and boats are visualised in real time. Additionally, building on the pre-trained deep learning model that will be used in the Must have task, research and tests will be done on different deep learning models using the acquired data.

Other tasks that would benefit the project are research on the performance of the prototype in another area, for example on sea. Research and tests on the performance of the prototype with other types of drones and other drone settings would also be interesting. Additionally, it would be beneficial if the boat type and ID's could also be automatically detected by the prototype.

Must have	**Should have**
• Deep learning model that can detect boats • Localisation algorithm that can calculate the geographical coordinates of detected objects • Real time object detection, localisation and intersection with restricted areas • Evaluation of the results	• Real time dashboard with map • Research on different deep learning detection models
Could have	**Will not have**
• Researching the prototype performance in other areas • Researching prototype performance with other drone types • Researching prototype performance with other drone settings • Detecting boat type and ID	• Research on usefulness of different machine learning methods • Extensive improvement of detection model

Table 1: MoSCoW analysis of the research project

4 Methodology

This section describes the proposed methodology that will be used to answer the research questions. Figure 4 depicts an overview of the methodology. It consists of two main phases that will be explained in more detail in the following subsections.

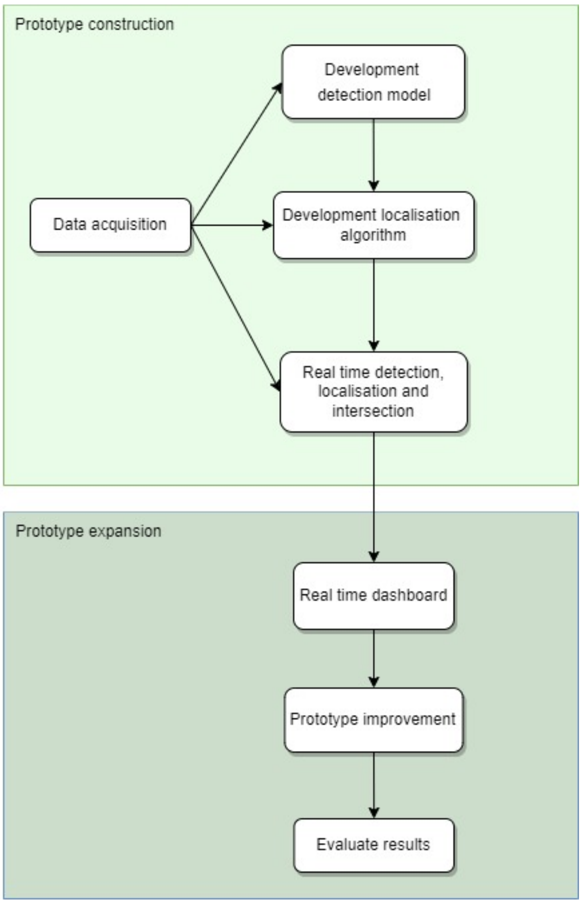

Figure 4: Methodology overview

4.1 Prototype construction

In this phase, the initial prototype is constructed. The phase consists of four main tasks: the acquisition of data, the development of the detection model, the construction of the localisation algorithm and the construction of the prototype where intersection of the object with the restricted area is possible in real time. The output of this phase is the initial prototype with the crucial functionalities on which expansion will be carried out in the next phase.

4.1.1 Data acquisition

The acquisition of data is the first main step in the research. The data that will be collected is image data from drones. The drone specification can be found at section 6.2. Data will be collected during the whole prototype construction phase. Data will be collected in different areas in the inland waters, for example data will be collected at the Biesbosch area and the IJsselmeer lake in the next month. Collected drone data will be stored and used during several steps of the research. Additionally, the real time functionality will be investigated during data acquisition. The drone will fly in fixed patterns over the inland waters. This will be done in cooperation with inspectors of the NVWA, as a drone certificate is needed to operate drones. During flight, the goal is to collect data of as many fisher boats as possible. If there are no fisher boats present on site, the NVWA will make a boat available to gather drone images of.

Data acquisition has already been carried out once at a harbour called Geersdijk in de Dutch Province Zeeland. Several images of a fish cutter were collected that can be used in the research. A few example images are displayed at Figure 5.

Figure 5: Drone images of a fish cutter

4.1.2 Development detection model

The detection of boats on drone images will be done using a deep learning model. The detection model that is used in the initial prototype will be a pre-trained model from the Keras API, which is part of the TensorFlow library. The benefit of using a pre-trained model is that less time and memory needs to be spent on training the model and fewer data will be necessary. This model is trained on images of the ImageNet database (see 6.3). A brief literature study will be conducted about which pre-trained Keras model will be used. The model will be used to detect boats, so no distinction will be made between fisher boats and non-fisher boats. The pre-trained model and drone images are invoked by means of a python script. When the model is used on collected drone images, the output are labelled images with the coordinates of the box around the detected object, which can be visualised with the Matplotlib library.

4.1.3 Development localisation algorithm

For this task, the geographic coordinates will be derived from a detected object in an image. The algorithm will be constructed according to the camera geometry principles. The intrinsic and extrinsic camera parameters will be derived from the drone and image metadata and will be used as input in the algorithm. The local upper left and lower right coordinates of the box around the detected object will be converted to 3-dimensional geographic coordinates. This set of geographic coordinates is the output of this algorithm.

4.1.4 Real time detection, localisation and intersection

In this step, the previous tasks will be combined into one real time prototype. First, an algorithm needs to be constructed that can compare the localised coordinates of the detected objects to the restricted areas. To get easy access to the layer with restricted areas, the constructed python script will be loaded and ran in ArcGIS Pro. By running the script from there, a connection can be made to the layer using the ArcGIS Python library.

Secondly, the prototype needs to be able to handle real time drone data. The real time drone data will be streamed through the Larix Broadcaster app, so the script needs to be able to make a connection to it. During this task, several tests will be conducted in the field to test the real time functionality of the prototype. The output of the prototype after this task will be the geographic coordinates of the ship and a boolean value whether this location is in a restricted area or not.

4.2 Prototype expansion

During this phase, the prototype constructed in the previous phase will be enhanced with additional features and functionalities.

4.2.1 Real time dashboard

The first task in improving the prototype is to create an insightful real time dashboard. The goal is to visualise a map of the restricted areas and update it with boat locations when one is found. The real time flight track of the drone will also be visualised. The construction of this dashboard app will be done with ArcGIS Dashboards. The input for this dashboard will be spatial data. The map with restricted areas will be shown, as well as the found boats. Additionally, other useful open spatial like weather data could be added to the dashboard. With this dashboard, the prototype is made more insightful for the inspectors.

4.2.2 Prototype improvement

In this task, the prototype will be improved with additional functionalities. Firstly, the detection model will be updated. Research will be done on different available deep learning detection models. The found most optimal model will be used and trained. For this training, the drone image data that is collected during the previous phase is used. This data will be manually labelled.

If the time is available, more functionalities will be added. These are described in the Could have section of the MoSCoW analysis at Table 1.

4.2.3 Evaluate results

The final prototype will be evaluated on three main criteria: detection accuracy, localisation accuracy and prototype speed.

The detection accuracy can be measured as the accuracy of the detection model correctly predicting ships in images. Performance metrics that will be used are the confusion matrix and the overall accuracy (Nan, 2022b).

To evaluate the localisation accuracy, the found coordinates will be compared to the real coordinates. To do this, a test will be conducted in the field with the prototype. The localisation results of the prototype are compared to a GPS tracker on the boat in real time.

To test the speed of the prototype, the time from image capture by the drone to the ship showing up on the dashboard is measured. This will also be tested in the field.

To determine the usefulness of the prototype as a new method for the NVWA inspectors, KPI's will be constructed on the three main criteria in consultation with the NVWA.

5 Time planning

The graduation project will be carried out in the academic years 2021/2022 and 2022/2023. In the next section, a Gantt chart is displayed and explained, showing the time planning of the research phases and their underlying steps. In section 5.2 the project milestones are given. Finally, in section 5.3 the meeting planning is discussed.

5.1 Gantt chart

The Gantt chart of the time planning can be found at Figure 6. The time planning is divided into four main phases: topic exploration (P1) in blue, graduation plan development (P2) in red, thesis draft (P4) in green and the final thesis (P5) in purple. In the writing period, one to two days per week will be dedicated to writing the thesis.

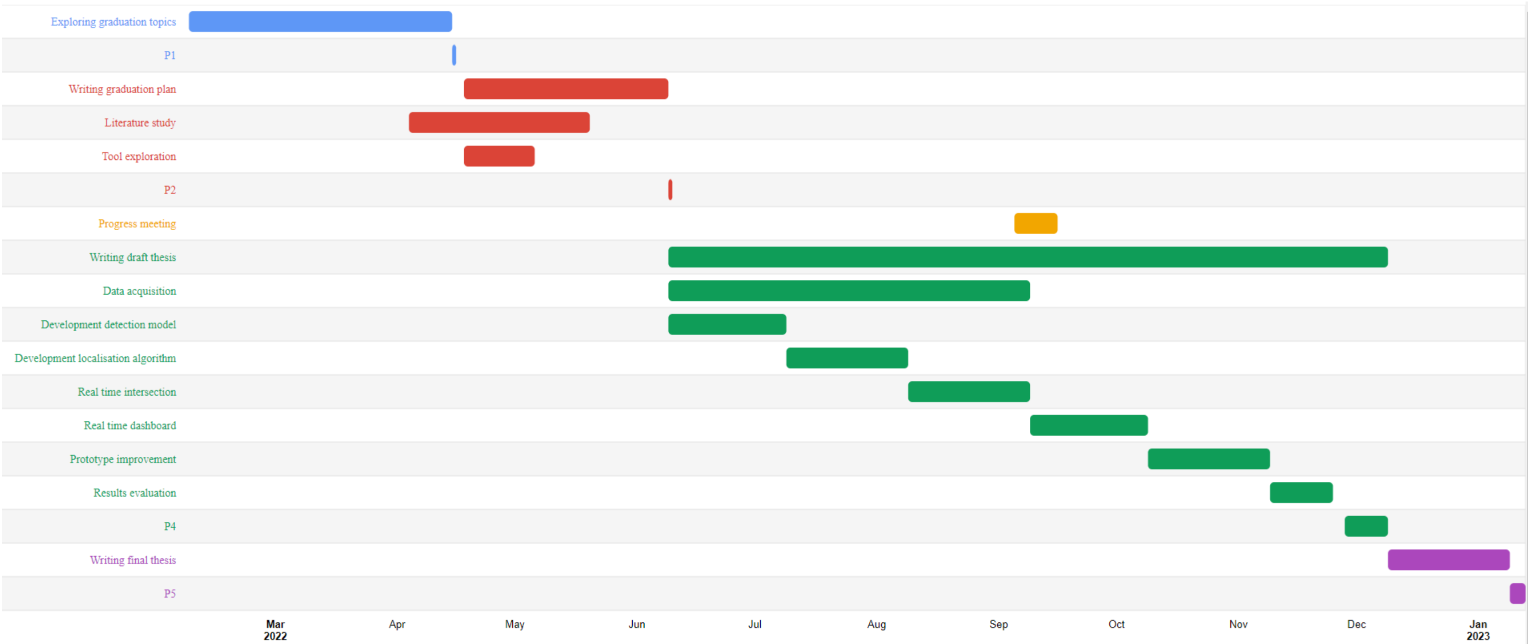

Figure 6: Gantt chart

5.2 Milestones

Table 2 shows the milestones of the graduation project. The dates are based on the official graduation calendar. The exact dates of P3, P4 and P5 are still to be determined by the graduation committee.

Milestone	Date
P1	April 15, 2022
P2	June 9, 2022
P3	September 5 - September 16, 2022
P4	November 28 - December 9, 2022
P5	January 9 - January 13, 2023

Table 2: Milestones during the graduation project

5.3 Meetings

Bi-weekly meetings are held on Wednesday afternoon with the TU Delft supervisors and the external Esri supervisor. These meetings will be held online through Microsoft Teams. A separate weekly meeting is held Wednesday morning with the Esri supervisor. In both meetings, the progress is discussed and guidance is provided when needed.

6 Tools and datasets

Several software and hardware tools that will be used during the research are described in the following subsections. Additionally, the used datasets are illustrated.

6.1 Software tools

The mainly used software tool for this project is ArcGIS. Additionally, open software libraries will be used. ArcGIS has many useful tools for spatial analysis and AI. Moreover, integration of other open spatial software libraries with ArcGIS is very efficient and a great variety of data types are supported. At Table 3 an overview of the used software is given.

Tool	Version	Description
ArcGIS Pro	2.9.2	Desktop Geographic Information Software used to visualise, analyse and process geographic information.
ArcGIS Python API	2.0.0	Python library that is used for mapping, spatial analysis, data science, geospatial AI and automation.
ArcGIS Dashboards	10.9.1	Tool to create real time dashboards that can present location-based analytics using intuitive and interactive data visualizations.
TensorFlow	2.8	Python end-to-end open source platform for machine learning.
OpenCV	4.5.5	Python and C++ library used for real time computer vision.
Larix Broadcaster	1.2.9	App that allows real time streaming of drone data over WiFi, EDGE, 3G, LTE and 5G.

Table 3: Software tools

6.2 Hardware tools

The drone used for this research is the DJI Matrice 30 [1]. This UAV has a maximum flight time of 41 minutes, a maximum speed of 23 meters per second and is resistant to water, wind and high temperatures. Moreover, it has a video and image camera, a thermal imaging camera and a laser rangefinder where the range from the drone to an object in meters can be found. The intrinsic parameters like focal length, the yaw pitch and roll and the 3 dimensional GPS location are recorded in the drone image metadata. The laptop used to develop and train the models is a Dell with a 4 GB memory GPU, which should just be sufficient to train detection models. If it is found this GPU is not sufficient, several virtual machines can be made available to train models on.

6.3 Datasets

The data used for training and testing of the detection model will be collected in the field using the drone previously mentioned. Collected data will be in the form of image data. Data collection will be done at least twice in at least two different locations, making sure the collected data is diverse enough.

During the research, the initial detection model that will be used is pre-trained by TensorFlow. Training of this model is done on ImageNet [2], which is a visual database that contains 14 million images intended for image classification.

[1] https://www.dji.com/nl/matrice-30
[2] https://www.image-net.org

References

Aburasain, R., Edirisinghe, E., & Albatay, A. (2020). Palm tree detection in drone images using deep convolutional neural networks: Investigating the effective use of yolo v3. In *Conference on multimedia, interaction, design and innovation* (pp. 21–36).

Apoorva, A., Mishra, G. K., Sahoo, R. R., Bhoi, S. K., & Mallick, C. (2021). Deep learning-based ship detection in remote sensing imagery using tensorflow. In *Advances in machine learning and computational intelligence* (pp. 165–177). Springer.

Bodensteiner, C., Bullinger, S., Lemaire, S., & Arens, M. (2015). Single frame based video geo-localisation using structure projection. In *Proceedings of the ieee international conference on computer vision workshops* (pp. 1–8).

Bos, O. (2018). *Report on the eel stock and fisheries in the netherlands 2016/2017* (Tech. Rep.). Stichting Wageningen Research, Centre for Fisheries Research (CVO).

Chang, Y.-L., Anagaw, A., Chang, L., Wang, Y. C., Hsiao, C.-Y., & Lee, W.-H. (2019). Ship detection based on yolov2 for sar imagery. *Remote Sensing*, *11*(7), 786.

Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

Htet, K. S., & Sein, M. M. (2021). Toddy palm trees classification and counting using drone video: Retuning hyperparameter mask-rcnn. In *2021 7th international conference on control, automation and robotics (iccar)* (pp. 196–200).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Leeuwen, S., Traag, W., Hoogenboom, L., & De Boer, J. (2002). Dioxins, furans and dioxin-like pcbs in wild, farmed, imported and smoked eel from the netherlands. *Organohalogen compounds*, *57*, 217–220.

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with python: a guide for data scientists*. " O'Reilly Media, Inc.".

Nan, L. (2021a). *Camera models.* `https://3d.bk.tudelft.nl/courses/geo1016backup/handouts/01-camera_models.pdf`.

Nan, L. (2021b). *Reconstruct 3d geometry.* `https://3d.bk.tudelft.nl/courses/geo1016backup/handouts/04-reconstruct_3D_geometry.pdf`.

Nan, L. (2022a). *Introduction to machine learning.* `https://3d.bk.tudelft.nl/courses/geo5017/handouts/01-Introduction.pdf`.

Nan, L. (2022b). *Performance metrics for classification.* `https://3d.bk.tudelft.nl/courses/geo5017/handouts/09-ClassificationMetrics.pdf`.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pedersen, S. A., Fock, H., Krause, J., Pusch, C., Sell, A. L., Böttcher, U., ... others (2009). Natura 2000 sites and fisheries in german offshore waters. *ICES journal of Marine Science*, *66*(1), 155–169.

Prayudi, A., Sulistijono, I. A., Risnumawan, A., & Darojah, Z. (2020). Surveillance system for illegal fishing prevention on uav imagery using computer vision. In *2020 international electronics symposium (ies)* (pp. 385–391).

Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., & Naemura, T. (2020). Cattle detection and counting in uav images based on convolutional neural networks. *International Journal of Remote Sensing*, *41*(1), 31–52.

van Oostenbrugge, H. (2021). *Nederlandse visserijsector in zwaar weer.* `https://www.wur.nl/nl/onderzoek-resultaten/onderzoeksinstituten/economic-research/show-wecr/nederlandse-visserijsector-in-zwaar-weer.htm`.

van Oostenbrugge, J., Bartelings, H., & Buisman, F. (2010). *Distribution maps for the north sea fisheries; methods and application in natura 2000 areas*. LEI, part of Wageningen UR.

Voinov, S. (2020). *Deep learning-based vessel detection from very high and medium resolution optical satellite images as component of maritime surveillance systems* (Unpublished doctoral dissertation). Universität Rostock.

Zhang, H., Wang, G., Lei, Z., & Hwang, J.-N. (2019). Eye in the sky: Drone-based object tracking and 3d localization. In *Proceedings of the 27th acm international conference on multimedia* (pp. 899–907).