Automatic analysis of digital aerial surveys

A roadmap for the short, medium and long term

Author(s): Mardik Leopold, Freek Daniëls, Hans Verdaat, Afra Asjes, Paul Goethals en Lydia Meesters

Wageningen University & Research report C082/21

WAGENINGEN UNIVERSITY & RESEARCH

Automatic analysis of digital aerial surveys

A roadmap for the short, medium and long term

Author(s): Mardik Leopold, Freek Daniëls, Hans Verdaat, Afra Asjes, Paul Goethals en Lydia Meesters

Wageningen Marine Research

Wageningen Food & Biobased Research

Wageningen Marine Research
Den Helder, November 2021

CONFIDENTIAL no

Wageningen Marine Research report C082/21

WAGENINGEN
UNIVERSITY & RESEARCH

Keywords: automatic, image recognition, high definition images, deep learning, seabirds, marine mammals

Client: Rijkswaterstaat Water, Verkeer en Leefomgeving
Attn.: Dagmar van Nieuwpoort
Postbus 2232
3500 GE Utrecht

This report can be downloaded for free from https://doi.org/10.18174/556349
Wageningen Marine Research provides no printed copies of reports

Wageningen Marine Research is ISO 9001:2015 certified.

Contents

Preface

Wageningen University & Research has been commissioned by Rijkswaterstaat WVL (RWS) to investigate how aerial surveys of seabirds and marine mammals can be automated using artificial intelligence-based image analysis techniques such as deep learning. The wishes and requirements of RWS and the approach were discussed during two workshops with experts from diverse backgrounds. The purpose and use of an automatic tool have been discussed with experts who perform the surveys and analyses to map the populations of marine animals and report them to the Dutch authorities. The approach to develop an automatic tool has been discussed with data and AI scientists and application developers. In a third workshop, the same experts were asked to provide feedback on the draft version of this report and to discuss outstanding questions. This report is the result of these workshops and individual discussions with experts.

1 Introduction

1.1 Objective

The aim of this assignment is to develop an approach and process by which automatic image recognition software can be developed based on high-definition (hidef) images collected during aerial surveys at sea. This document will provide a perspective on the short, medium, and long term (1 year (concrete), 3 years (global) and 5 years or longer (dot on the horizon)). The ultimate aim is to develop an open-source platform for automatic image recognition and species identification for all types of footage collected by the most common types of aerial vehicles.

1.2 Scope

The scope of this project includes the following topics:

- Organizing three workshops with (inter)national image recognition experts to discuss, based on the quality and annotations of the already collected images, the possible approach for the development of automatic image recognition software.
- Writing a roadmap that describes the process to develop an automatic image recognition tool on the short, medium, and long term:
    - Description of the expected quality of image recognition and the various testable intermediate stages of the end product for process control. Short, medium, and long term for future technical development are distinguished.
    - Brief exploration of the Dutch and international standards used in the data structure, thereby ensuring that the output matches those international standards.
    - A focus on the species that are important to Wozep (seabirds and marine mammals) but takes into account possible broadening of the application (for instance other marine animals, schools of fish, plastic pollution, algal blooms) of the developed software during development.

What is not part of this roadmap:

- Detailed plan of deployment
- How to collect representative datasets
- Applying machine learning algorithms to explore the feasibility of an automatic tool
- Operational phase of the developed system including aspects of ownership and governance
- Extensive literature study
- Digital survey design

1.3 Benefits of an automated tool for digital surveys

We realise that there will be pros and cons in switching from human visual observers to digital surveys and to automated evaluation of the survey results. These have been discussed by Žydelis *et al.* (2019) and in the workshops preceding the writing of this document. Some of the more obvious advantages of moving towards a fully automated approach are:

- Observer independent
- Quality assurance and transparency; images can be saved and viewed again later.
- More detections / more identifications
- Ultimately more cost efficient

2 Marine animal surveying

2.1 Current approach to at-sea seabird surveys

Several different platforms and methods have been and are still being used to survey seabirds at sea. Methods that yield seabird densities (n/km^2) are: aerial, ship-based and land-based surveys (Tasker *et al*. 1984; Baptist & Wolf 1991; Komdeur *et al*. 1992; Durinck *et al*. 1994; Camphuysen *et al*. 2004). All three survey methods have in common that seabirds are counted within pre-determined survey areas. These might be counts within strips of a given length and width or counts of birds within a larger area of sea, visible from vantage points on land, such as secluded bays or estuaries. Land-based counts are, in principle, total counts within a rather large area, while both aerial and ship-based surveys take multiple, small samples. The basic technique in the latter two is a strip-transect method, in which all birds seen within a given surface area of sea are counted. It was soon realised, however, that birds at larger distances from the central (transect) line had a lower probability of detection than birds at closer range. To correct for birds being missed by the observer at larger perpendicular distances, sub-bands within the counting band, or strip were introduced (Tasker *et al*. 1984), allowing comparisons of apparent bird densities in sub-bands at various distances, so that densities in more distant sub-bands could be corrected, using Distance analysis techniques from line transect theory (Buckland *et al*. 1993; Skov *et al*. 1995; Poot *et al.* 2011). The end result of all survey techniques mentioned should be estimates of seabird densities (n/km^2), and results from different surveys might be combined (e.g. Skov *et al*. 2007; Bradbury *et al*. 2014; Leopold *et al*. 2015). These survey techniques may be used for any species that is sufficiently abundant and detectable at the surface, be it seabirds or cetaceans (e.g. Fijn *et al*. 2020; Waggitt *et al.* 2020), pinnipeds (Leopold *et al*. 1997), or turtles and large fish (Ridoux *et al*. 2010), or, for that matter, non-biota (Scheidat & Feint-Herr 2012).

2.2 Digital aerial surveys – domain knowledge

Today, there is a clear movement towards digital surveys of seabirds and cetaceans at sea, using high-definition cameras applied from (mostly at this time) fixed-wing aircrafts. Other carriers might be deployed as well, such as helicopters or unmanned vehicles (drones), e.g., from ships in remote waters (e.g., Herr *et al.* 2019). One clear advantage of digital surveys is increased safety: less personnel is needed, while the cameras are flown at much greater altitudes than human observers, who need to fly at rather risky low altitudes to be able to identify some of the more difficult targets (van Roomen *et al*. 2013). Surveying at higher altitudes has the added advantage that offshore wind farms are no longer obstacles in the survey design. Other advantages of digital surveys are accountability: digital images can be retained for second opinion analyses, and that between-observer differences in detection and identification skills, that can be substantial (van der Meer & Camphuysen 1996), can be ruled out. Whether or not digital surveys can outcompete human observers in all situations, and can thus be more cost-effective, is not yet clear (Collier *et al.* 2021). However, there are also several probable disadvantages of digital surveying: flying at greater altitudes might decrease the number of weather windows for surveys; "context" (items attracting seabirds, such as platforms, fishing vessels, front lines etc) are mostly not captured on the images; initial costs are higher.

Žydelis *et al*. (2019) describe the aerial, high-definition video survey method that is used in Borssele and that will also likely be used in the future Gemini project. They report a better performance of digital surveys over classic surveys with human observers, in terms of coverage, numbers of birds detected and identified, reduced disturbance of targets and reduced observer bias. However, specific identification of some species (grebes, auks) was problematic, and the collected images were evaluated by human observers (“trained ornithological specialists”), who might agree or disagree on the identification of certain targets. Target identification by human specialists is very time-consuming and, like classic observer surveys, subjective to a certain extent. The next step would therefore be to build an automatic tool to detect and identify targets of interest on the digital images collected.

2.3 International standards

Ship-based and aerial surveys, carried out by human observers, have been used alongside each other in most North Sea countries for decades. The step towards using digital, aerial surveys, particularly in relation to offshore wind farm studies, has now firmly been made in the UK and in Germany. All UK Round 3 and 4 (see e.g., www.thecrownestate.co.uk/round-4/) offshore wind farm studies rely principally on aerial digital survey methods while Germany now requires that aerial digital survey methods are used for offshore wind farm studies. Likewise, digital survey methods are now used in offshore wind farm impact studies in the states of Maryland and Texas (USA). Digital methods are being evaluated in other regions. For instance, for the Baltic, the Baltic Marine Environment Protection Commission (2020) notes that: "Though currently too expensive for the purpose of large-scale monitoring, digital imaging from aircrafts may be an(other) option in future."

3 Automated analysis of digital aerial surveys

3.1 Introduction

Since many years, digital aerial images are taken from specific water plots, e.g., in the North and Baltic Seas, for analysis of the presence of marine mammals and seabirds. Rijkswaterstaat is interested in estimating how many individuals of which species occur in which location. This analysis is currently performed in a non-automated way by trained humans. It is a highly specialised and time-consuming task.

Until recently, the specialised task was too complex to consider automating it. However, in recent years, artificial intelligence (AI) has developed rapidly and has opened up possibilities to automate tasks that were previously performed manually. Therefore, Rijkswaterstaat has asked for this investigation into the possibility of performing an automated analysis of digital aerial surveys.

Several studies reported in literature (see section 3.2) on automatic interpretation of wildlife photography have shown the potential of AI, and in particular of deep learning, for large scale, accurate analysis of animals (and plants) in digital images.

The proposed approach to apply deep learning and build an automatic analysis tool for localisation and identification of seabirds and marine mammals in digital aerial surveys is outlined in section 3.3.

3.2 Related work from scientific literature

Automatic and semi-automatic methods for the surveying of animals can be partitioned into three classes of methods: (1) pixel-based methods (2) object-based methods and (conventional) machine learning, and (3) deep learning (Wang *et al.* 2019).

Pixel-based methods are typically applied to VHR images captured from space where single animals may take up only a few pixels or even a single pixel. Because we will be considering the case where animals occupy more of the image, these methods fall beyond the scope of this overview.

The next two classes of methods (conventional machine learning and deep learning) rely on imagery where animals occupy enough of the image that features such as shape and texture can be employed in making decisions. The difference between these methods lies in how these features are chosen for decision making support. Conventional machine learning methods often rely on humans to choose sensible features to operate on. While human expertise can be helpful in this process, such features are not necessarily optimal or robust. The power of deep learning-based approaches is that they can automatically learn to use features of interest within sets of images. These features can be more robust in decision making than features a human might choose. On ImageNet, a benchmark database with over 14 million annotated images, the best conventional machine learning approach, reported an error rate of around 25% prior to deep learning (Langlotz *et al.* 2019). After its introduction in 2012, deep learning models reported error rates of 15% in 2012 to less than 3% in 2017, outperforming humans whose error rate is reported at 5% (Langlotz *et al.* 2019). Systems based on deep learning have achieved state of the art performance in a large and growing number of problem domains, and so we choose to focus on these systems in this overview.

Existing systems based on deep learning can be divided by the strategy they employ to localize animals within a larger image. Perhaps the most straightforward way of doing so is to partition the image into smaller sub images on a regular grid of tiles. The system then tests for the presence of an animal in each of the sub images. This approach was taken by Gray *et al.* (2019) to predict the presence of sea turtles using a convolutional neural network (CNN) acting on each sub image. Similarly, Borowicz *et al.* (2019) detect cetaceans in satellite imagery by applying the CNNs ResNet and DenseNet to tiles. In detecting dugongs, Maire *et al.* (2015) take a different approach in choosing the sub images by centring them on super pixels, allowing them to extract animal positions by equating these with the super pixel centres.

Alternatively, one can attempt to process the entire image using a deep network and try to predict the density of animals or probability of finding animals within subregions of the image. This approach was taken by Kellenberger *et al.* (2021) where they predict probability maps for seabirds within input images using a variant of the ResNet architecture. These maps are then processed to localize individual seabirds within the image. Padubidri *et al.* (2020) instead attempts to predict density maps of sea lion population by using the UNet architecture. Images are mapped to density maps, which can then be processed to obtain an estimated count of individuals within the input image.

Another possible approach is to try to predict bounding boxes for all animals in the image directly. Guirado *et al.* (2019) take a hybrid approach to count whales where they first eliminate sub images that do not contain any whales. The remaining sub images are then processed using a network based on the Faster RCNN architecture, which produces bounding boxes for candidate whales within each sub image. Hayes *et al.* (2021) use a similar network based on a ResNet-50 backbone to localize seabirds in images of colonies.

Concluding: we state that the use of deep learning methods has proven its worth in automated animal detection in digital images. We suggest using this approach as the starting point for the proposed automated image recognition approach. The goals and approach are elaborated upon in the next sections.

3.3 Objective of the project

The overall objective of the project is to provide a roadmap to develop a tool that automatically detects and identifies seabirds and marine mammals in digital aerial images. The tool needs to work with sufficient accuracy and have an option for humans to correct false identifications. It must learn from such corrections. The tool needs to be tested, verified, and made available through an open-source online platform for various animal detection tasks.

This roadmap distinguishes three phases: the short term (1 year), the mid-term (3 years) and the longer term (5 years). In the short term, the proof of principle should be given, in the mid-term the development of a more complete software suite for data-driven biodiversity purposes should be put central and models for other animal classes might be developed, in the long term the software suite should become available for world-wide use. The goals of these three phases are elaborated upon below. In addition, chapter 4 will elaborate more on starting up this process.

3.3.1 Goals for Phase 1: the short term

In the short term (one year) the goal is **to demonstrate the feasibility of automated detection, identification and counting of birds and marine mammals on digital video surveys using deep learning (AI).** More specifically, the starting point of phase 1 is the existing set of Borssele wind farm images (frames from the video, each provided with GPS coordinates), that have been annotated by experts. This set of images and annotations are the ground truth for method development. In addition, it might be advisable to put human observers on board of service vessels in wind farms under high def surveillance, to get some ground truthing from independently assessed ratios of e.g. guillemot/razorbills on site. For now, the subgoals of phase 1 are to automatically measure the following parameters in this dataset:

- In each image, the location of the seabirds and mammals that are present is determined
- In each image, the identity of the present seabirds and mammals is determined

- From each present seabird and mammal, other characteristics annotated by experts are determined as much as possible. One can think of the activity of the animal (flying, swimming), of the gender or plumage of the species, flying directions etc.
- For each video an overall count of detected and identified seabirds and mammals is determined (along with beforementioned characteristics if included).

At the start of phase 1 the desired list of species and parameters to be detected and identified needs to be specified as well as targets for the performance/accuracy. Possibly the list of species from the annotations can be used, but this can also be further refined. As for the targets, an indication of human performance can be found in Žydelis *et al. (*2019) where a validation of similar annotation work took place and a detection rate between 91.51–99.71% and an identification rate of 93.20–96.76% was reported.

Phase 1 will focus on model development and evaluation to demonstrate the feasibility and report on the intermediate and achieved performances. At the end of this phase a 'go/no go' decision will be made (based on the SMART criteria set to be described at the start of phase 1) and the developed models will be deployed to an environment where it can be run by trained operators (see Figure 1). Only in the next phases time will be spent on (pure) software development that will realize the end-goal, an interactive platform for automated analysis of digital video surveys.

In order to improve the image recognition model other annotated datasets can be added, if made available by project partners or other contributors. For instance, should the PPS-project Gemini be granted (see Annex 1), which will be announced beginning of November 2021, this will generate a dataset similar to that of the Borssele surveys. A partner in this PPS also has datasets from similar surveys, but collected in other regions, with species relevant to the surveys preformed in the Southern North Sea.

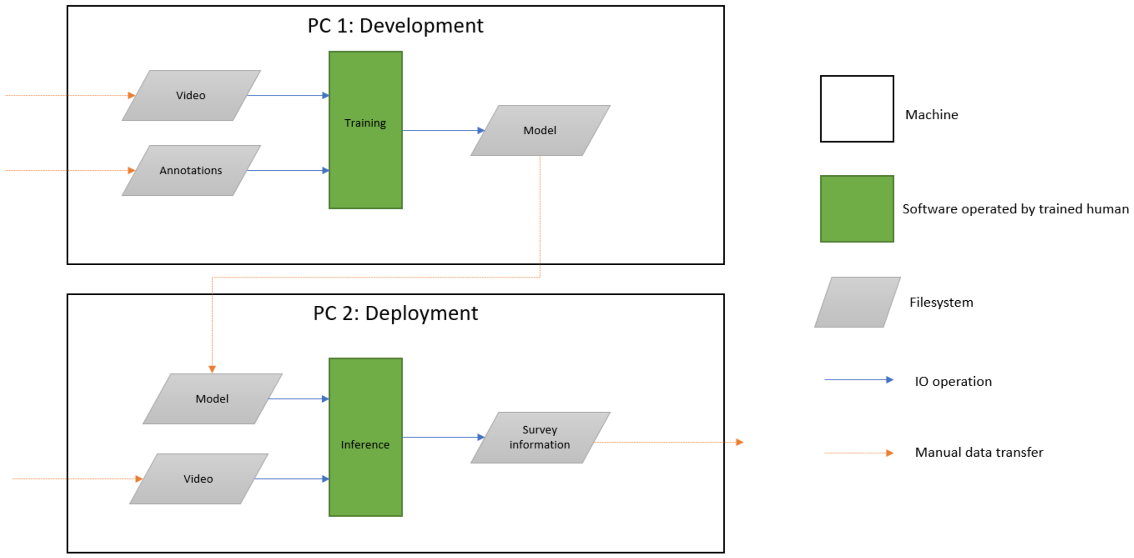

Figure 1. *Slightly simplified workflow at the end of phase 1. Models are developed using the available videos and annotations. Developed models are then deployed on a separate pc on which new video material can be analysed. "Model" here is assumed to both identify and count seabirds and mammals on videos where in practice both an AI model and additional algorithms are needed for these tasks.*

3.3.2 Goals for Phase 2: the mid-term

For the second phase of the project, to be concluded in approximately three years, the goal is to design and build a platform standardizing models and data, integrating data sets, evaluating existing models on new data and extending the applications to different domains and imaging technologies (UAV, satellite).

The existing models will be tested and evaluated with new survey data when available to assess their performance. It is not unimaginable that the models require retraining when presented with data that differ to some degree from the original training data. This could happen for instance when data is obtained from a different area where the background is much more complex than seen during the training phase. This could lead to an increase of false positive detections and some retraining will be required to familiarize the model with the new background. Typically, such retraining requires much less annotations. An evaluation should take place to see if retraining is needed and how many annotations it requires. In practice a domain expert could be asked to annotate several images after which the model is retrained. Retraining can be either in an online fashion, i.e. as annotations or corrections are provided, or offline. Model retraining strategies should be explored and their performances evaluated depending on the desired functionalities of the platform.

Next to the model developments additional focus is on software development to build the platform that will facilitate the development of models and the automated analysis of video surveys. A major development will be the design of databases and ontologies (knowledge models describing the data, their properties and the relations between data) (Gruber 1993). Different data sets will have different annotators, and different annotators can use different names (or rather pieces of text) to identify the same species. To standardize this, ontologies should be used (such as WoRMS for marine mammals). Ontologies not only make sure that unique and universal identifiers for species (or objects in general) are used, but also capture the relation between objects. For example, all annotations for a particular family of seabirds can be found using the established relationship between family, genus and species in an ontology. Similarly, models will also be described using ontologies. This allows automated matching between type of models and annotations. Models can for example require each image to have one label (object present yes/no) or require each image to have a list of labelled bounding boxes, giving the position and identification of species. Such ontologies support model development in efficiency as relevant annotations can be automatically retrieved and utilise the full potential of available data by having it integrated and connected. Furthermore, databases will be designed to store data according to the FAIR principles (findable, accessible, interoperable, reusable). When we use a model on a data set, we want to log on what data the model was trained, what the performance obtained, who provided the annotations etc.

A second development would be to build a user interface. The view of the project team during the workshops has been that in this phase the data and annotations are (still) entered to the database through software developers/maintainers and that the models are run also by trained operators rather than end users (domain experts). This because of the massive amount of video data, associated storage costs and the software developments needed to support such a functionality. The user interface would give relevant users the possibility to request analysis on data sets and retrieve the results as well as inspect any relevant information related to how the results were achieved.

At the end of phase 2 the platform, depicted in Figure 2, should serve as an infrastructure where all data are standardized and stored. Model development and inference will be connected to the data (through an intermediate backend) to fetch relevant annotations, store models and their performances as well as completed analyses. This platform will likely run on a server with cloud storage for the data and high-performance clusters to train and run the models.

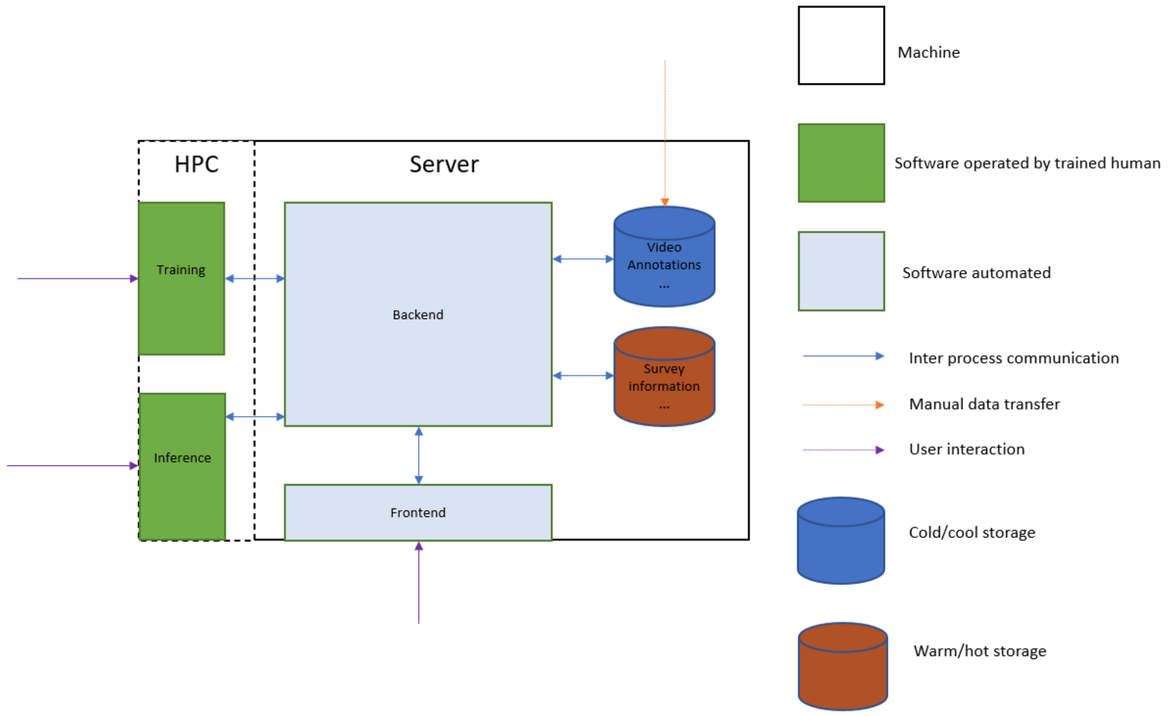

Figure 2*. Schematic overview of the platform at the end of phase 2. Data are stored in a cloud environment and connected through a backend to training and inference procedures which run on a high performance cluster (HPC). A frontend is made for user interactions to review model performances, survey outcomes and to request analyses.*

3.3.3 Goals for Phase 3: the long term

For the long term, the platform will be further automated and developed open source to make it available to (international) parties outside the project group. Developed models will be made available for users through the platform. Users will be given the possibility to upload data and annotations to the platform, either through the user interface or through software APIs (application programming interface). Additionally, it can be considered to incorporate functionalities to support users and domain experts to make annotations and corrections within the user interface. Users can search for suitable models and run these (without the interferences of model developers) on their data should they exist. The available annotations can be used to evaluate the performance and, if needed, for (re)training. Selected retraining strategies from phase 2 could be integrating allowing retraining to be done online in an automated fashion if suitable models exist. In case of an entirely new application, offline training with the help of a model developer might be needed. In both cases, uploaded annotations are only used for models of that particular user unless verified by a domain expert. This prevents the introduction of noise in the database. In Figure 3 the components of the platform at the end of phase 3 are shown.

At this point we assumed the deep learning models work on individual frames and an additional algorithm counts the detected and identified individuals over all frames in the videos (removing duplicate detections). The algorithm for counting is not necessarily a trained algorithm but rather a designed one which works under certain conditions such as minimal amount of overlap between images. When external users add their videos for analysis, such conditions might not hold. The user could then use the models for detection and identification on individual frames and export them for further analysis elsewhere or the algorithm could be adjusted and made available through the platform. The assumption is made with regards to the current state-of-the-art, but in five years' time perhaps the best performing models are trained directly on videos in which case this is less relevant.

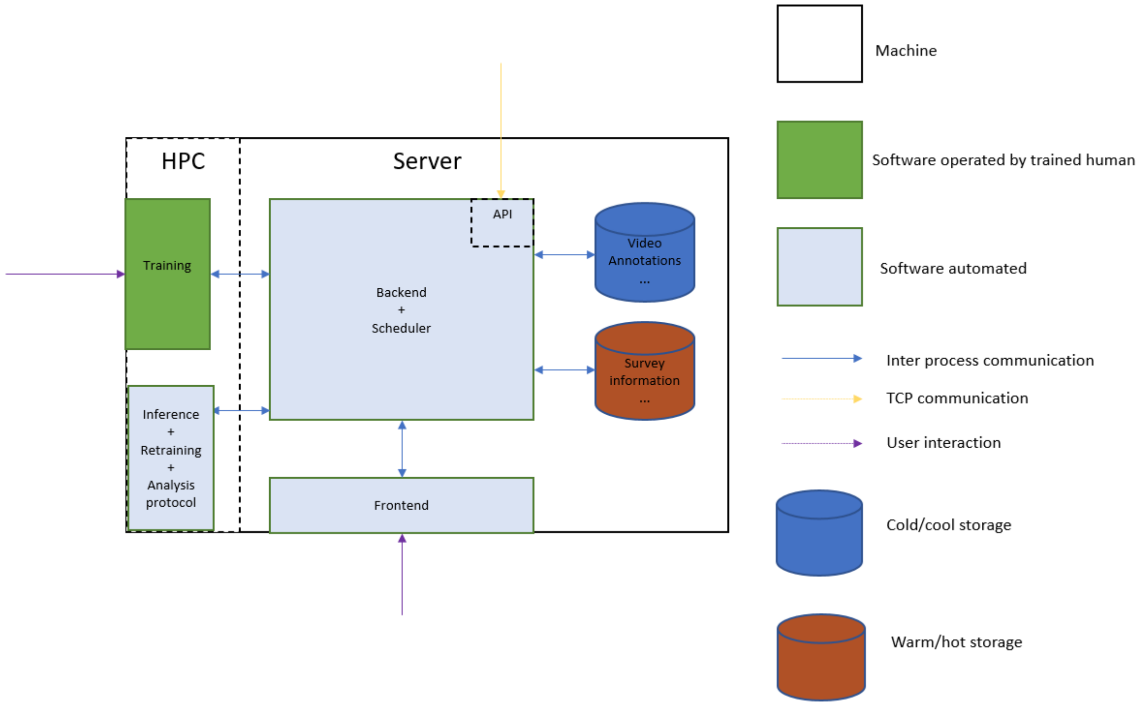

Figure 3*. In phase 3 we allow external users to upload data and annotations and automate the inference of models which they can search for, select and start through the user interface (frontend).*

3.4 Current and future projects

In the two years leading up to the writing of this document, project partners had been communicating with the commissioner about developing a publicly available aerial image recognition tool developed by a broad developing team of experts within The Netherlands. It gradually became clear that data of possibly two projects can be made available for use in the first phases of development. From February 2021 to February 2022 (possibly to be extended to February 2023), digital video footage is collected in, and adjacent to Borssele offshore wind farm by a collaboration of Bureau Waardenburg and HiDef/BioConsult (total 24 to 48 surveys). Digital images will also be collected (2022-23) in and adjacent to Gemini offshore wind farm if the PPS-subsidy that has been requested is awarded (see annex 1). When using the data collected in both projects, it might be possible to develop software for automated detection, identification and counting of the seabirds and marine mammals captured on the footage. By using the data collected in both the Borssele and the Gemini project for training the software, these two projects benefit from each other.

Once the software is up and running as required, there is the potential for other monitoring projects to benefit from this methodology, like:

- MWTL seabird and cetacean surveys
- seaduck and eider surveys in the Wadden Sea and North Sea coastal waters
- specific surveys of certain “difficult species groups” such as grebes and divers in North Sea coastal waters
- wind farm-related waterbird surveys in the North Sea and in IJsselmeer
- future seabird surveys in the North Sea (both in Dutch domestic waters, and in other parts)
- future SCANS surveys.
- foreseeable developments in other regions might be surveys of certain charismatic species (cetaceans, sharks, flying fish, marine turtles, seabirds) in e.g., the Caribbean Sea.
- other uses, e.g., identifying / quantifying floating litter, algal blooms, fishing vessels’ activities, etc.

4 Starting the development with phase 1

4.1 Analysis of first images

4.1.1 Survey area: Borssele wind farm

Sample data have been made available and have been analysed to estimate the feasibility of automating detection, identification and counting of birds and mammals present and the methods needed to do so. The data consist of images, and annotations of species present on those images, from an area in and next to the Borssele wind farm (Figure 4). From this first data set, already some first insights into the difficulty for the objective at hand have been obtained. Example images are shown in Figures 5 and 6. In each a single individual (Great Blacked-backed Gull and Harbour Porpoise, respectively) can be seen within a yellow bounding box. From an altitude of around 550 meters the pixel resolution is roughly 2.5 cm per pixel.

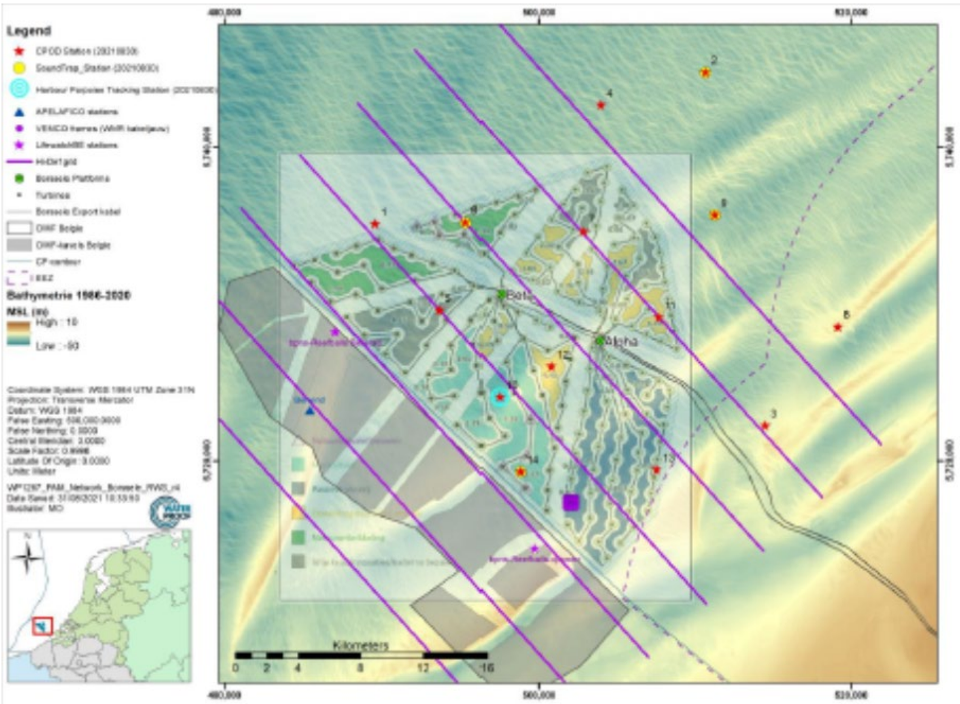

***Figure 4**. Survey area with transect lines covering both Dutch and Belgian waters. The map also includes CPOD / SoundTrap positions that are part of the RWS PAM network (source: WaterProof 2021).*

4.1.2 Foreseeable problems with detections, identifications and statistical treatment of the data collected

Digital aerial surveys supposedly have (near) 100% registrations of the objects of interest. However, image quality may not always be optimal (Figures 5 & 6), e.g., sun glare in parts of the survey area and white caps on the water surface may hinder detection, over-exposure of images may affect white and dark parts of seabird plumage differently, while some species look very similar to each other from above. Other species might be too small for detection (depending on image resolution) or are too rare on the collected images for robust statistical analyses, while there are also the known problems of non-random occurrence and zero-inflation and effects of presumed impact factors that are unclear or dwarfed by other confounding factors on animal distribution (Leopold 2018; Zuur 2018; Zuur *et al.* 2014). In the early days of (visual) aerial surveys (Baptist & Wolf 1993) the rule of thumb was that 25 detections of a given species per survey were needed for analysis of densities or trends in density. How many hits are necessary for full training of a deep learning system needs to be explored. Table 1, taken from Fijn *et al.* (2020), with several rarer species of seabirds added, gives an overview of these potential problems.

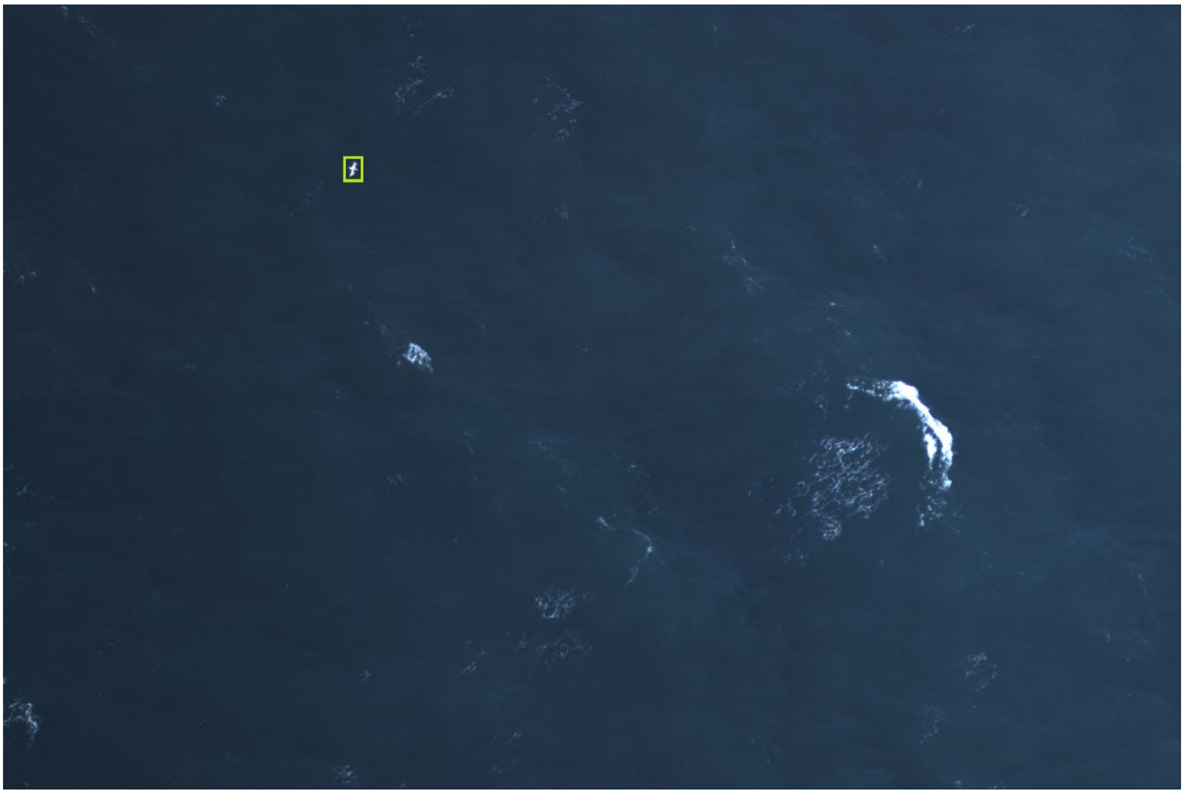

Figure 5. *Example image from Borssele data set, an immature Great Back-blacked Gull can be seen in the yellow bounding box.*

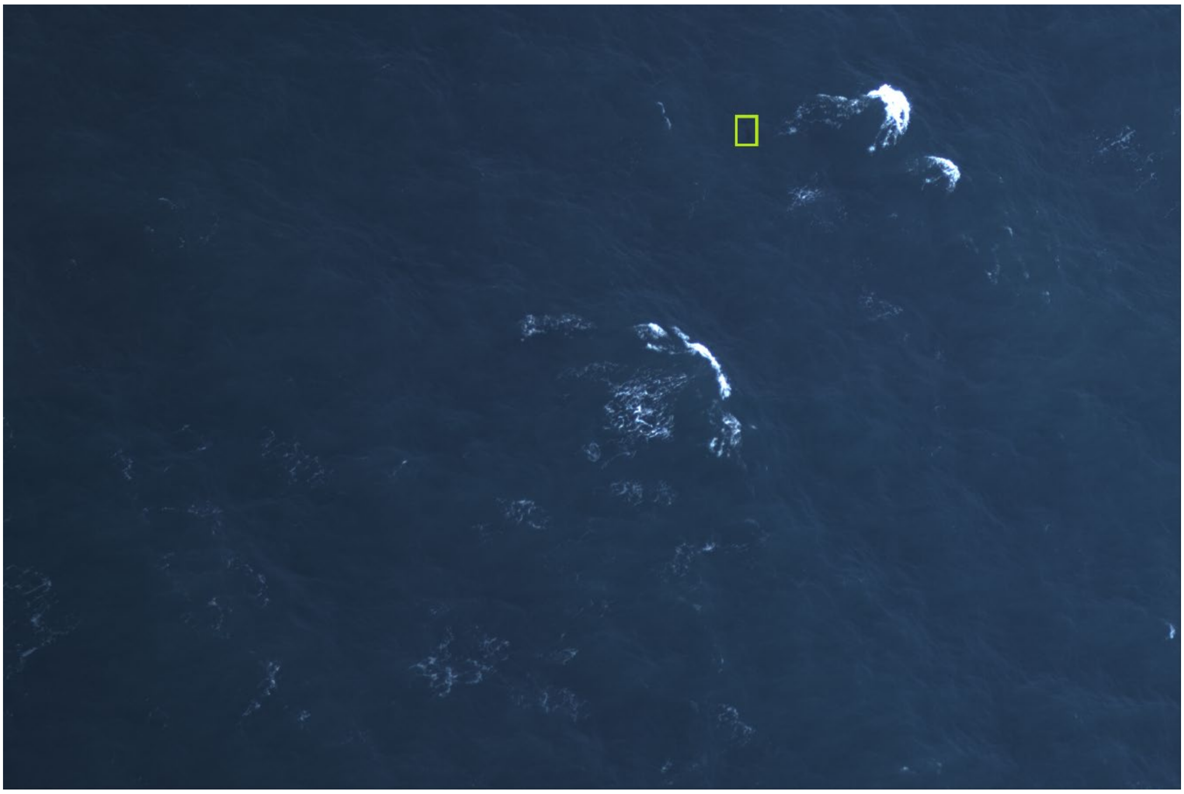

Figure 6. *Example image from Borssele data set, a porpoise can be seen in the yellow bounding box.*

Table 1. *List of seabird species that are of potential interest in digital aerial surveys in Dutch marine waters. The "number of observations" gives an indication of relative densities in the target area: these were the total numbers of observations (of 1 or more birds) seen during 6 MWTL monitoring flights over the Dutch sector of the North Sea, in 2019-2020 (*Fijn *et al.* 2020)*. Note that unidentified gulls, terns or auks, or rare species of gulls, are not included here.*

Euring code	Species	Number of Observations	Too rare?	Too small?	Cryptic	confusing species
9	Great Creasted Grebe	19	1		1	Smaller grebes, particularly Red-necked Grebe
20	Red-throated Diver	128			1	Black-throated Diver, Great Northern and White-billed Divers
30	Black-throated Diver	0	1		1	Red-throated Diver, Great Northern and White-billed Divers
110	Slavonian grebe	2	1	1	1	Black-necked Grebe
220	Northern Fulmar	303				
460	Manx Shearwater	1	1		1	Sooty and Balearctic Shearwaters, Razorbill
520	European Storm Petrel	0	1	1	1	Leach's Storm Petrel
550	Leach's Storm Petrel	0	1	1	1	European Storm Petrel
710	Northern Gannet	635				
720	Great Cormorant	193			1	European Shag
2060	Common Eider	20	1			
2130	Common Scoter	111			1	Velvet Scoter (swimming)
5660	Pomarine Skua	0	1		1	Arctic Skua
5670	Arctic Skua	6	1		1	Pomarine, Long-tailed Skuas
5690	Great skua	8	1			
5780	Little Gull	47	1		1	Kittiwake (juv)
5820	Black-headed Gull	61	1			Little Gull
5900	Common Gull	192			1	Kittiwake, Herring Gull, Lesser Black-backed Gull (juv)
5910	Lesser Black-backed Gull	1771			1	Great Black-backed Gull, juvenile gulls in general
5920	Herring Gull	526			1	Common, Lesser & Great Black-backed Gulls
6000	Great Black-backed Gull	226			1	Lesser Black-backed Gull, juvenile gulls in general
6020	Kittiwake	1618			1	Common Gull, Little Gull (juv)
6110	Sandwich Tern	987			1	other terns
6150	Common Tern	599			1	Arctic Tern
6160	Arctic Tern	11	1		1	Common Tern
6340	Common Guillemot	2683			1	Razorbill
6360	Razorbill	957			1	Common Guillemot
6470	Little Auk	0	1	1	1	Atlantic Puffin
6540	Atlantic Puffin	64	1	1	1	Little Auk

In digital aerial surveys that collect more than one image simultaneously (Figure 7), numbers of detections should be similar on all images, in other words, images collected closer to the track line of the plane should have the same animal density as images taken at larger perpendicular distances (after correction for the relative surface areas on the images). Whether or not this is the case needs to be tested.

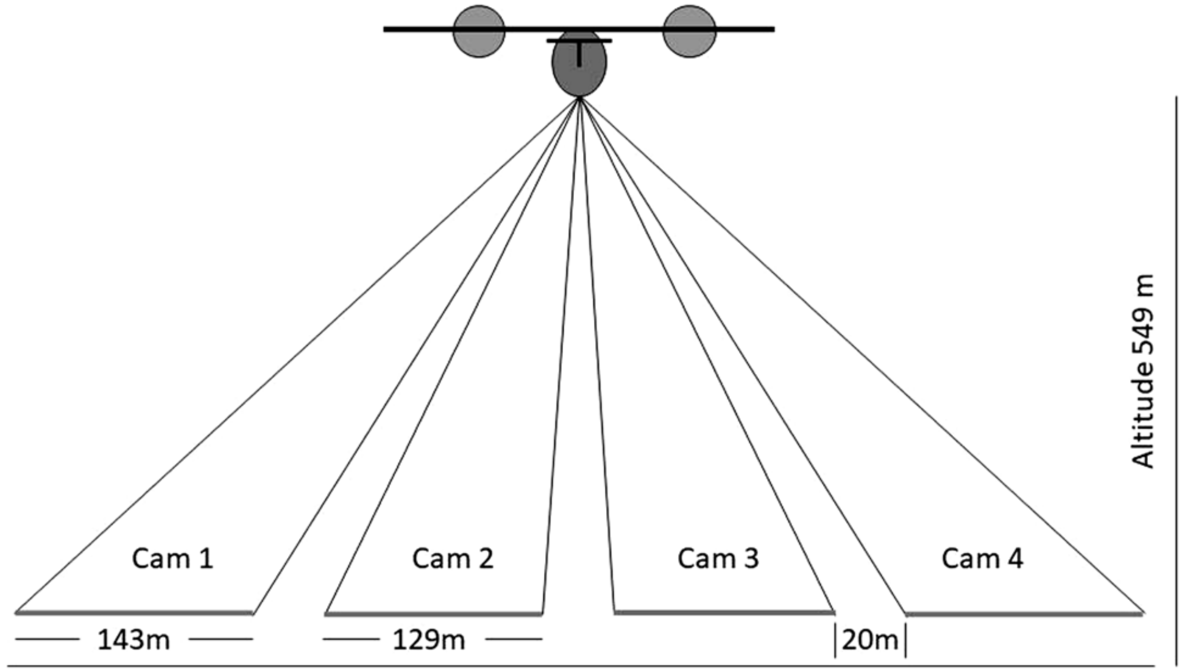

Figure 7. *(taken from Žydelis et al. 2019). Surveying with four digital cameras from a fixed-winged aircraft. Statistically, the same target densities should be found left and right of the plane and in the inner (129 m wide) and outer (143 m wide) strips.*

The effects of weather conditions, particularly glare and white caps can seriously hamper visual observations. How this affects digital surveys and the numbers of validated images for each observation condition, needs to be evaluated.
Many seabird species might look very similar from above and this might hamper species identification. Moreover, some species have different plumages over time, that at least require a relatively large data set of validated images, for each plumage that occurs over time. The same applies to species that have different "colour morphs" (Northern Fulmar, skuas), or subspecies of different coloration (Lesser Black-backed Gull: from Herring Gull-like grey to Great Black-backed Gull-like black). There may also be a lower size limit, given the quality of the images available, for species that can be detected or identified. Animals that spend time under water, like cetaceans or pinnipeds will have varying probabilities of detection with the depth at which they occur as the plane passes overhead, and the same applies to diving seabirds.

4.1.3 Description of the available data

Table 2 lists the species present in the Borssele sample dataset. In the second column the number of distinct individuals is indicated. Each individual is typically captured around four to seven times on succeeding frames. The total amount of annotations is given in the third and last column. The overlap on succeeding frames gives more opportunity for an algorithm to detect an individual. On the other hand, when counting it must be deduced if detected individuals were already seen. Having the individuals annotated multiple times rather than once is very valuable as it increases the number of examples, although not all annotations of the same individual will be as informative as those from different individuals. When birds are swimming for instance, multiple frames do not show much change in appearance. Therefore, this will only add a little bit of new information. For birds that are flying however, the change of wing position can cause significant variation in appearance and additional annotations of the same individual are highly informative.

From Table 2 we can furthermore deduce that species identification is not always possible. Human experts are sometimes only able the infer the family or reduce the possible species to up to two species. It remains to be specified if these categories are relevant or if they should be refined. The other information we can clearly deduce is that the numbers per species are highly skewed. A few species dominate and for those a lot of annotations are available. Many species on the other hand have only a few annotations. Though this table lists only those species from the sample data and (much) more data are expected, this skewedness is likely to remain to some extent. Collecting data from multiple seasons and combining these data with data from different areas will over time allow for more species to be automatically identified. This phenomenon is exactly what calls for a platform, standardizing and centralizing such datasets. Simultaneously ongoing developments in AI will lead to models requiring less examples.

Table 2. *Number of species and individuals present in the sample data received.*

Species (activity)	Number of individuals	Number of annotations
Razorbill (sitting)	727	4788
Common guillemot (sitting)	279	1642
Common guillemot or razorbill (sitting)	225	1319
Black-legged kittiwake (sitting)	121	657
Lesser black-backed gull (sitting)	110	564
Northern gannet (sitting)	65	337
Harbour porpoise (moving)	58	302
Black-legged kittiwake (flying)	54	304
Lesser black-backed gull (flying)	52	272
Northern gannet (flying)	49	296
Unidentified songbird (flying)	30	143
Razorbill (flying)	26	130
Unidentified large gull (sitting)	26	143
Great black-backed gull (sitting)	19	102
Unidentified pinniped (grey or harbour seal) (stationary)	15	81
Herring gull (sitting)	14	66
Unidentified small gull (sitting)	11	57
Common guillemot or razorbill (flying)	11	61
Common guillemot (flying)	7	51
Common gull (flying)	6	31
Unidentified bird (sitting)	6	28
Unidentified gull (sitting)	6	34
Herring gull (flying)	6	29
Great black-backed gull (flying)	5	23
Common gull (sitting)	5	22
Great or lesser black-backed gull (flying)	3	17
Little gull (flying)	2	9
Unidentified pinniped (grey or harbour seal) (moving)	2	10
Black-headed gull (flying)	2	10
Great cormorant (flying)	2	8
Unidentified larus gull (sitting)	2	9
Unidentified large gull (flying)	2	11
Seal or small cetacean (moving)	2	12
Great or lesser black-backed gull (sitting)	1	5
Unidentified dove or pigeon (flying)	1	5
Harbour seal (stationary)	1	5
Unidentified bird (flying)	1	4
Unidentified diver (flying)	1	6
Common or herring gull (sitting)	1	6
Black-throated diver (sitting)	1	5

Figures 8-11 show example annotations (cropped out from the larger images) of flying birds, swimming (floating) birds, songbirds, and cetaceans respectively. Each individual is shown on five consecutive frames. When estimating the feasibility for automating the detection and identification, these annotations are viewed using criteria such as the information present, intra-species variation, inter-species similarities and background variation. The (amount of) information present (number of pixels, colour, in focus or not) determines the complexity of features that can be extracted or learned. On small blurry blobs no abstract robust features can be learned. With more intra-species variation, more data are generally needed to get a good representation of the variation. Inter-species similarity gives an indication on how difficult identification will be, high similarity simply means the visual differences between species are small and subtle. Finally, background variation is especially important in the detection part. If this variation is high, it is expected that there is either a difficulty in finding rare species or large numbers of false positives might occur.

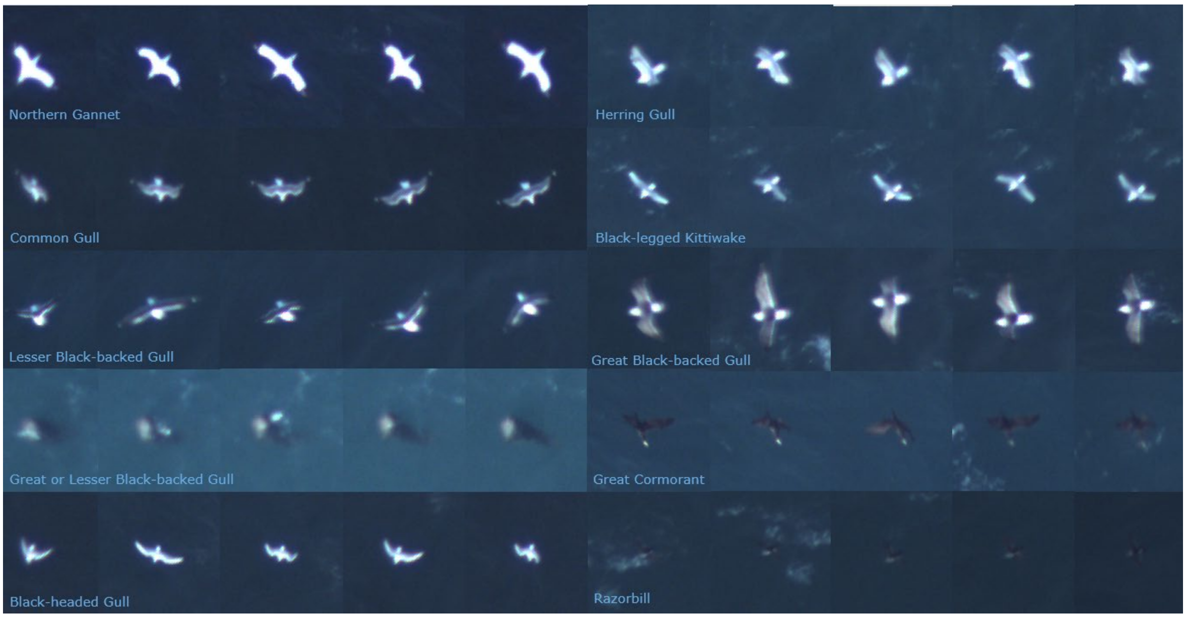

Figure 8. *Example annotations of seabirds flying.*

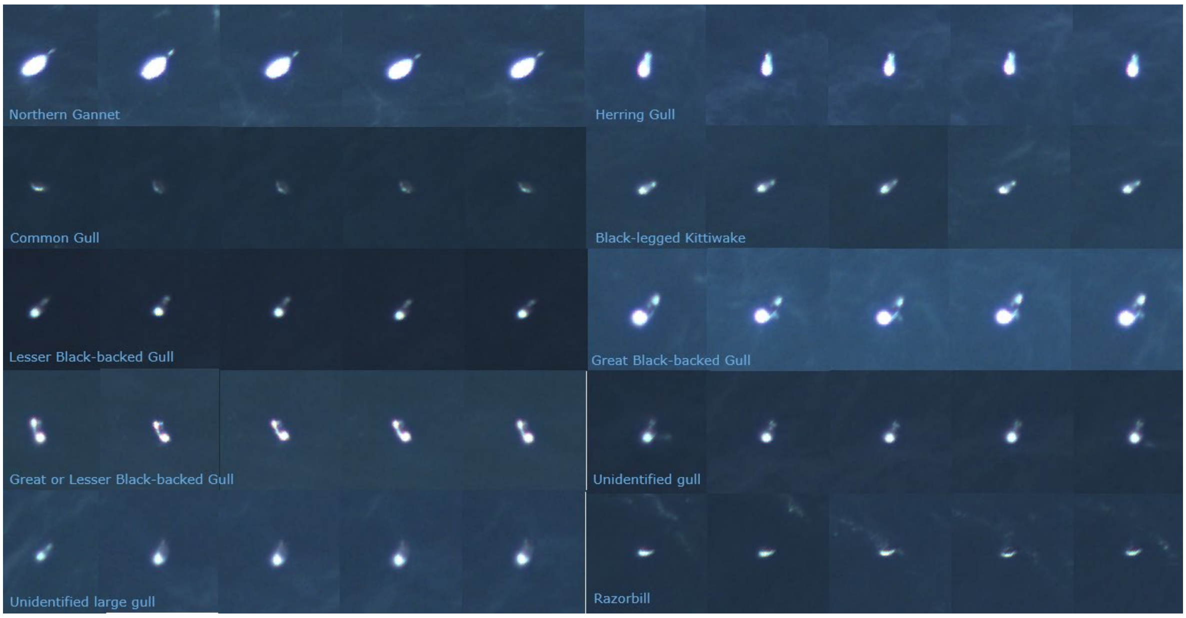

Figure 9. *Example annotations of seabirds swimming.*

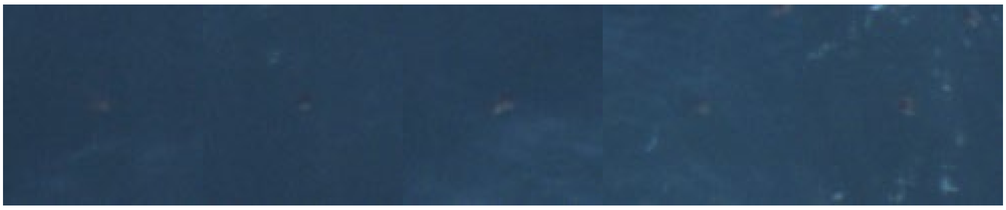

Figure 10. *Example annotations of an unidentified songbird (flying).*

Figure 11. *Example annotations of cetaceans*

When we inspect the annotations, we see that the information present can be limited. Smaller species cover fewer pixels which leaves little to extract robust features from. It is worthwhile to discuss this in the future with the human experts on how they identify smaller species such as songbirds (and possibility if full resolution data were received or a compressed version). Larger species are more clearly visible, and shape and some colour can be seen. Although especially white birds are often overexposed losing most of the texture information, making inferring parameters such as plumage unlikely. Some inter-species similarities can be seen especially between sitting birds but from the annotations we can see that humans can identify them, nevertheless. Also, here it is good to discuss with experts what it is that distinguishes similarly looking species. With modern tools we can inspect what deep learning models pay attention to in images and compare this to what experts use to make distinctions and so can increase trust in the models. However, ground-truthing for abundant, yet similar species such as auks ("razormots") may prove useful. With regards to the intra-species variation, we see for example significant differences due to motion of flying birds. It is very valuable that the annotations cover multiple frames of individuals so that the variation is captured as much as is possible. Finally, what seems to make automating the detection using state of the art technology highly feasible is the fact that the background variation is low. Background here means everything not considered an object of interest. Because of this we expect that the detection of rare species (without identification) should even be possible. The exact feasibility and number of annotations cannot be calculated beforehand. However, an estimate for automating identification of larger species when flying would probably be around 200 annotations of individuals (captures in multiple frames), for smaller species or species sitting more annotations might be needed.

Based on paragraphs 1.2 and 1.3, a list of issues one might run into when developing the deep learning model can be made:

- Depending on the survey method, when more than one image is collected simultaneously, correct for relative surface of the image opposite to the capture angle from the plain
- Weather conditions, like glare of the sun or white caps on the sea surface, species might look identical from above
- Some species show plumage changes during for instance their lifetime
- Some species have colour morphs
- Marine mammals might be less visible when deeper under water due to water turbulency.
- For rare species there might not be enough images available for training

4.1.3.1 Feedback on image capture

From the images seen so far, we observe two areas that could be considered for improvements. The resolution (pixel per cm) is quite low, especially when dealing with smaller birds. This leaves little information for models to extract robust abstract features from. No discussion has taken place with the company responsible for the image acquisition on the used hardware nor has an inventory taken place on available and suitable cameras. As such no comment can be given at this stage on the possibility or expected cost for upgrading cameras to improve the resolution. Secondly, there is a lot of overexposure such that texture information of birds is lost. If plumage is needed to identify species or gender this could be a limitation. The use of polarizing filters could be considered to remove or reduce the effects of glare and overexposure if not already in place. The need for improvements is related to the demand for performance. From the given images human experts can identify most birds to a certain degree. We expect with enough annotations to reach similar or better performances with current state-of-the-art deep learning models. Improvements in image acquisition could lower the number of needed annotations or improve performance by providing more discriminatory information.

4.2 Roadmap to model and algorithm development and evaluation

It is advised to use the Borssele dataset as the starting point for building a platform for automated digital video or image surveys. In this first phase, of approximately one year, this dataset can be used to demonstrate the performance of state-of-the-art deep learning models and algorithms for:

1. Detecting objects of interest in videos surveys: perhaps the most crucial part of this development. If one can detect all birds and mammals in images (frames of videos), experts are no longer needed to scan through lengthy videos in search of objects of interest but rather the machine can select the images and regions within those images where animals are located. This promises to be a major time and therefore cost saver.
2. Identify the detected species. For species with sufficient numbers of annotations it will be possible to identify them automatically. These species therefore no longer need to be analysed by experts at all.
3. Count the detected and identified species. Species are detected, and if possible, identified, on extracted frames from videos. Because it is video footage, individuals are captured on multiple frames. This requires an additional algorithm to prevent the same individuals from being counted more than once throughout the videos.

The task of localising and identifying multiple objects in images is commonly known as an *object detection* problem. The goal of which is to predict objects by bounding boxes and provide for each bounding box a label or class. In our case these classes would encode the different species. The best performing deep learning models for object detection at this moment include Faster-RCNN (Ren *et al.* 2017) and YoLo (Redmon *et al.* 2016). These could be considered as a starting point to try and solve goal one and two in the beforementioned list. To track objects over frames and obtain a count (goal three) a combination of Kalman filters (Kalman 1960) and the Hungarian algorithm (Kuhn 1955) could for example be considered.

Another possible approach would be to use methods like DeepSort which integrates both object detection and tracking (Wojke *et al*. 2017). DeepSort also uses Kalman filters and the Hungarian algorithm, but additionally trains a deep learning model on the appearances of objects between frames. Using this model, it is attempted to reduce the number of misassignments of objects between frames. Such errors typically occur when paths of objects cross or occlude each other. Methods like DeepSort place a much heavier burden on the annotation process as they require that objects are annotated and linked throughout the video. However as mentioned earlier in this chapter for the Borselle dataset this is the case.

The performance of these three goals needs to be evaluated in two parts. Firstly, a subset of the videos needs to be kept aside during algorithm development and training as a test set. On this test set, the animals that were annotated by (human) experts are compared with where animals were predicted (detected (and possibly identified)) by the algorithm. This comparison could be quantified using two metrices, 'precision' and 'recall'. Recall gives the percentage predicted animals that were annotated. Precision gives the percentage of predictions that are matching with the annotations. In an extreme example, one can obtain a high recall by predicting animals everywhere, this will lead to a very low precision though (often wrong). As an opposite, one can obtain a high precision by predicting animals nowhere (never wrong), but that will result in a very low, or zero, recall. Typically, a balance between both is sought, though perhaps in this project a higher recall is desired, and a slightly lower precision might be tolerated as long as human experts are still in the loop for making corrections. This would mean the system gives more false positives (animals detected which are not there) but fewer false negatives (animals not detected). Such a performance is expected to still lead to significant cost reductions as domain experts are only needed to review detections rather than analyse entire video surveys.

This first part of the evaluation leads to a performance relative to the annotations. However, it is not unlikely that errors exist in the annotations. Species can be missed or misidentified. Therefore, the second part of the evaluation has to be that independent experts review those detections and identifications where the predictions and the annotations do not agree. Without knowing what was predicted and what was annotated the experts will decide who was right. This final evaluation should lead to a 'go'/'no go' to complete phase 1 and start phase 2. At the start of phase 1 the criteria for a 'go' should be formulated by the involved parties using the SMART principles. If a 'go' is given, phase 1 will be completed by deploying the solution (i.e., developed algorithms) to an environment where new surveys can be analysed upon request. This environment will be controlled by the developers while no interface for domain experts is yet developed.

4.2.1 Infrastructure

In phase 1 all developments can in principle be done locally if the size of the data permits this. The environment in which the models will be deployed can be a standalone pc. If it so happens to be that the data of the Borssele dataset is of such size that local storage is inefficient or awkward in use, it is advised that all operations start in a cloud environment. This will require budget for the renting of space and computation time. Cloud storage is available in different modes: from cold, cool to warm and hot. Thought should be given on which mode is needed for what type of data. For example, cold and cool modes are cheaper, but these data are not readily available and need to be requested (e.g., for analysis) in advance. Given the expected size of the video survey data (>20 TB per flight) and because most of it will be needed sparsely (only in training) one of these modes would seem to be the best fit.

4.2.2 Agile project approach

We recommend approaching this project using an iterative or agile approach where frequent communication takes place between the end user(s) of the software and the software development team. The advantage of such an approach is that not all wishes and requirements need to be fully mapped out beforehand (as is the case in the traditional 'waterfall method') but can be gathered incrementally. Incremental prototypes make it easier for users to understand the capabilities and limitations of the automated tool and to adjust, refine and prioritise the tool's requirements during the project. The iterative process is depicted in Figure 12.

Figure 12. *Schematic depiction of an agile development process, each cycle is called a 'sprint' and each sprint results in an implemented 'minimum viable product'.*

The initial planning of the project should take place together with all stakeholders. A product vision document is created, in which a detailed description of the software tool is given and in which user stories are defined. User stories are descriptions in the form "as <user> I want <functionality> because of <reason>". The user stories are prioritised and one or a few are selected as basis for the first minimum viable product. This list of selected user stories is referred to as the 'back log'. In the sprint, the user stories are elaborated, the corresponding algorithms are designed and implemented, the created software is tested, and the resulting minimum viable product is evaluated with the stakeholders. Before the second cycle starts, new user stories are added to the back log according to the stakeholders' priorities and the process is repeated.

In order to achieve a good agile software development process, it is important to identify and involve relevant stakeholders on a regular basis. Apart from the project commissioners, these are:

- Data providers, who provide the image sets, meta data and manual annotations.
- Users, who will be using the output generated by the automated tool, e.g., to generate reports for policy makers.
- Users, who will use the tool to automatically detect and identify the animals in the tool.
- Policy makers, who will need to use the reports to determine and adjust the relevant policies.
- An application provider, who will eventually deploy and maintain the automated tool.

Not all stakeholders will have the same role, but it is important to keep track of every stakeholder's interests.

4.2.3 Deliverables

The expected deliverables of a project following Phase 1 would be:

- Describe a set of SMART criteria as a basis for the 'go/no-go' decision to move from phase 1 to phase 2 by the parties involved in the development of the model and algorithm. These criteria reflect the wish list of species that the system needs to be able to detect and identify and the minimum performance level considered acceptable (for example in terms of precision and recall as mentioned in 4.2).
- A report describing the architecture of the software tool with a summary of the results of the automated annotation on the Borssele data set
- A hands-on session for the users of the software to learn to work with the software tool
- Demo / video of the created software tool
- Roadmap for the next phase

5 Quality Assurance

Wageningen Marine Research utilises an ISO 9001:2015 certified quality management system. This certificate is valid until 15 December 2021. The organisation has been certified since 27 February 2001. The certification was issued by DNV GL.

References

Baltic Marine Environment Protection Commission 2020. Draft HELCOM guidelines for monitoring seabirds at sea. Online: https://portal.helcom.fi/meetings/STATE%20-%20CONSERVATION%2013-2020-779/MeetingDocuments/3N-8%20Draft%20HELCOM%20Guidelines%20for%20monitoring%20seabirds%20at%20sea.pdf

Baptist H.J.M. & Wolf P.A. 1991. Vogels monitoren per vliegtuig. Sula 5 16-23.

Baptist H.J.M. & Wolf P.A. 1993. Atlas van de vogels van het Nederlands Continentaal Plat. Rapport DGW-93.013, Dienst Getijdewateren, Rijkswaterstaat, Middelburg, 168p.

Borowicz A., Le H., Humphries G., Nehls G., Höschle C., Kosarev V. & Lynch H.J. 2019. Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. PLoS ONE 14(10): e0212532.

Bradbury G., Trinder M., Furness B., Banks A.N., Caldow R.W.G. & Hume D. 2014. Mapping Seabird Sensitivity to Offshore Wind Farms. PLoS ONE 9(9): e106366.

Buckland S.T., Anderson D.R., Burnham K.P. & Laake J.L. 1993. DISTANCE Sampling: Estimating abundance of biological populations. Chapman & Hall, London, reprinted 1999 by RUWPA, University of St. Andrews, Scotland.

Camphuysen C.J., Fox T.J., Leopold M.F. & Petersen I.K. 2004. Towards standardised seabirds at sea census techniques in connection with environmental impact assessments for offshore wind farms in the U.K. - A comparison of ship and aerial sampling methods for marine birds, and their applicability to offshore wind farm assessments. NIOZ Report to Cowrie / The Crown Estate, 38p.

Collier, M. P., F. Weiß, C. G. Irwin and R. C. Fijn (2021). High-definition bird and marine mammal aerial survey image collection in Borssele: Implementation plan., Bureau Waardenburg.

Durinck J., Skov H., Jensen F.P. & Pihl S. 1994. Important marine areas for wintering birds in the Baltic Sea. EU DG XI research contract no. 224/90-09-01, Ornis Consult Report 1994: 1-110, Copenhagen.

Fijn R.C., van Bemmelen R.S.A., de Jong J.W., Arts F.A., Beuker D., Bravo Rebolledo E.L., Engels B.W.R., Hoekstein M., Jonkvorst R.-J., Lilipaly S., Sluijter M., van Straalen K.D. & Wolf P.A. 2020. Verspreiding, abundantie en trends van zeevogels en zeezoogdieren op het Nederlands Continentaal Plat in 2019-2020. RWS-Centrale Informatievoorziening BM 20.22. Bureau Waardenburg Rapportnr. 20-324. Bureau Waardenburg & Deltamilieu Projecten, Culemborg.

Gray P.C., Fleishman A.B., Klein D.J., McKown M.W., Bézy V.S., Lohmann K.J. & Johnston D.W. 2019. A convolutional neural network for detecting sea turtles in drone imagery. Methods in Ecology and Evolution 10: 345–55.

Gruber T. 1993. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43: 907–928. doi:10.1006/ijhc.1995.1081.

Guirado E., Tabik S., Rivas M.L., Alcaraz-Segura D. & Herrera F. 2019. Whale counting in satellite and aerial images with deep learning. Scientific Reports 9: 14259.

Hayes M.C. Gray P.C., Harris G., Sedgwick W.C., Crawford V.D., Chazal N., Crofts S. & Johnston D.W. 2021. Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. Ornithological Applications 123: 1-16. doi:10.1093/ornithapp/duab022

Herr H., Kelly N., Dorschel B., Huntemann M., Kock K.-H., Lehnert L.S., Siebert U., Viquerat S., Williams R. & Scheidat M. 2019. Aerial surveys for Antarctic minke whales (*Balaenoptera bonaerensis*) reveal sea ice dependent distribution patterns. Ecology and Evolution 2019; 1-19. doi:10.1002/ece3.5149

Kalman, R. E. (March 1, 1960). "A New Approach to Linear Filtering and Prediction Problems." ASME. *J. Basic Eng*. March 1960; 82(1): 35–45. https://doi.org/10.1115/1.3662552

Kellenberger B., Veen T., Folmer E. & Tuia D. 2021. 21 000 Birds in 4.5 h: Efficient large-scale seabird detection with machine learning. Remote Sensing in Ecology and Conservation, doi:10.1002/RSE2.200

Komdeur J., Bertelsen J. & Cracknell G. (eds) 1992. Manual for aeroplane and ship surveys of waterfowl and seabirds. IWRB Special Publication 19, National Environmental Research Institute Kalø, 37p.

Leopold M.F., van der Werf B., Ries E.H. & Reijnders P.J.H. 1997. The importance of the North Sea for winter dispersal of harbour seals Phoca vitulina from the Wadden Sea. Biological Conservation. 81: 97-102.

Leopold M.F., Booman M., Collier M.P., Davaasuren N., Fijn R.C., Gyimesi A., de Jong J., Jongbloed R.H., Jonge Poerink B., Kleyheeg-Hartman J., Krijgsveld K.L., Lagerveld S., Lensink R., Poot M.J.M. van der Wal J.T. & Scholl M. 2015. A first approach to deal with cumulative effects on birds and bats of offshore wind farms and other human activities in the Southern North Sea. IMARES Report C166/14, 188p.

Leopold M.F. 2018. Common guillemots and offshore wind farms: an ecological discussion of statistical analyses conducted by Alain F. Zuur. Wageningen University & Research report C093/18 (project WOZEP Birds-1). doi:10.18174/466106

Langlotz, Curtis & Allen, Bibb & Erickson, Bradley & Kalpathy-Cramer, Jayashree & Bigelow, Keith & Cook, Tessa & Flanders, Adam & Lungren, Matthew & Mendelson, David & Rudie, Jeffrey & Wang, Ge & Kandarpa, Krishna. (2019). A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology. 291. 190613. 10.1148/radiol.2019190613.

Maire F., Alvarez L.M. & Hodgson A. 2015. Automating marine mammal detection in aerial images captured during wildlife surveys: A deep learning approach. Advances in Artificial Intelligence, Lecture Notes in Computer Science, eds. Bernhard Pfahringer and Jochen Renz. Cham: Springer International Publishing, 379–85.

Meer J. van der & Camphuysen C.J.1996. Effect of observer differences on abundance estimates of seabirds from ship-based strip transect surveys. Ibis 138: 433-437

Padubidri C., Kamilaris A., Karatsiolis S. & Kamminga J. 2021. Counting sea lions and elephants from aerial photography using deep learning with density maps. Animal Biotelemetry 9: 27.

Poot M.J.M., Fijn R.C., Jonkvorst R.J., Heunks C., de Jong J. & van Horssen P.W. 2011. Aerial surveys of seabirds in the Dutch North Sea May 2010 - April 2011. Seabird distribution in relation to future offshore wind farms. Report 10-235 Bureau Waardenburg, Culemborg, 233p.

Ridoux V., Certain G., Dorémus G., Laran S., Van Canneyt O. & Watremez P. 2010. Mapping diversity and relative density of cetaceans and other pelagic megafauna across the tropics: general design and progress of the REMMOA aerial surveys conducted in the French EEZ and adjacent waters. IWC SC/62/E14.

Redmon, J., Divvala, S., Girshick, R. (2016) "You Only Look Once: Unified, Real-Time Object Detection". IEEE Conference on Computer Vision and Pattern Recognition CVPR. http://dx.doi.org/10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., Sun, J. (2017) "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6):1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

van Roomen M., Stahl J., Schekkerman H., van Turnhout C. & Vogel R. 2013. Advies monitoringplan vogels in het Nederlandse Noordzeegebied. Sovon-rapport 2013/22. Sovon Vogelonderzoek Nederland, Nijmegen. https://www.sovon.nl/sites/default/files/doc/Rap_2013-22_monitoringplan_noordzee.pdf

Scheidat M. & Feint-Herr H. 2012. Collecting data on marine debris during cetacean aerial surveys. IWC SC/64/E.

Skov H., Durinck J., Leopold M.F. & Tasker M.L. 1995. Important bird areas for seabirds in the North Sea, including the Channel and the Kattegat. Birdlife International, Cambridge, 156p.

Skov H., Durinck J., Leopold M.F. & Tasker M.L. 2007. A quantitative method for evaluating the importance of marine areas for conservation of birds. Biological Conservation 136: 362-371

Tasker M.L., Jones P.H., Dixon T.J. & Blake B.F. 1984. Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardized approach. Auk 101: 567-577.

Waggitt J.J., Evans P.G.H., Andrade J., Banks A.N., Boisseau O., Bolton M., Bradbury G., Brereton T., Camphuysen C.J., Durinck J., Felce T., Fijn R.C., Garcia-Baron I., Garthe S., Geelhoed S.C.V., Gilles A., Goodall M., Haelters J., Hamilton S., Hartny-Mills L., Hodgins N., James K., Jessopp M., Kavanagh A.S., Leopold M., Lohrengel K., Louzao M., Markones N., Martinez-Cediera J., 'O Cadhla O., Perry S.L., Pierce G.J., Ridoux V., Robinson K.P., Santos M.B., Saavedra C., Skov H., Stienen E.W.M., Sveegaard S., Thompson P., Vanermen N., Wall D., Webb A., Wilson J., Wanless S. & Hiddink J.G. 2020. Distribution maps of cetacean and seabird populations in the North-East Atlantic. Journal of Applied Ecology 57: 253-269.

Wang D., Shao Q. & Yue H. 2019. Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): A review. Remote Sensing 11: 1308.

WaterProof 2021. Implementation- and PAM operations Plan. Risk Assessment Method Statement Bruinvis Netwerk Borssele. Report number: 31167838.

N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.

Zuur A.F. 2018. Effects of wind farms on the spatial distribution of guillemots. Highland Statistics Ltd. Report for Wageningen Marine Research and Rijkswaterstaat

Zuur A.F., van Bemmelen R.S.A. & Leopold M.F. 2014. Zero-inflated sea bird data sampled at offshore wind farms. Ch 6 (pp 145-186) in: A.F. Zuur, A.A. Saveliev & E.N. Ieno. A Beginner's Guide to Generalised Additive Mixed Models with R. Published by Highland Statistics Ltd., Newburgh, United Kingdom.

Žydelis R., Dorsch M., Heinänen S., Nehls G. & Weiss F. 2019. Comparison of digital video surveys with visual aerial surveys for bird monitoring at sea. Journal of Ornithology 160: 567-580.

Justification

Report C082/21
Project Number: 4312100133

The scientific quality of this report has been peer reviewed by a colleague scientist and a member of the Management Team of Wageningen Marine Research

Approved: ir. S.C.V. Geelhoed
Researcher

Signature:

Date: 2nd of November 2021

Approved: drs. J. Asjes
Manager Integration

Signature:

Date: 2nd of November 2021

Annex 1 Adjacent PPS-project

1.1 Contribution

Part of the contribution to the development described in this roadmap is a requested subsidy PPS-project, a collaboration between public and private parties, that aims to run parallel to and provide input to the larger Image Recognition project described in this roadmap. If granted, the "Gemini / PPS project" will contribute to the development by collecting additional field data, attributed images and (software & AI) development. It is stressed that this PPS is not part of this roadmap but will apply the steps described in this document in the development.

1.2 Deliverables

Table 1. *Project deliverables of PPS. Abbreviations: WMR = Wageningen Marine Research, BuWa = Bureau Waardenburg, AFR = Agro Food Robotics*

WP	Deliverable	Title	Who	Type	Delivery date
1	1.1	Annotated images Gemini wind farm	Bioconsult-HiDef	Data set	31-06-2023
1	1.2	Validation of annotated images Gemini & Borssele wind farms	WMR/BuWa	Data set	30-09-2023
1	1.3	Article annotation validation	WMR/AFR/BioConsult/ BuWa	Scientific paper	31-12-2024
2	2.1	Deep learning model: iterative updates (three per year)	AFR/WMR	Model	31-12-2025 (Final report)
2	2.2	Article on deep learning models	AFR/WMR	Scientific paper	31-12-2025
2	2.3	Method description collected field data.	All project partners	Scientific paper	31-12-2025
WP	**Deliverable**	**Title**	**Who**	**Type**	**Delivery date**
2	*2.4*	*(Software) platform automated digital video surveys*	*AFR*	*Software*	*31-12-2025*
3	3.1	Infographics on social media	WMR	Outreach	31-12-2025

1.3 Project team PPS

- Gemini is enabling this project within their wind farm and making the budget for the data collection and analysis of auks and guillemots habitat loss available for the broader analysis for all species and for use for the general purpose of developing image recognition software (AI).
- WMR is project leader, responsible for project management and guarantees cooperation within the project, the connection of this project with the development of the National Infrastructure Image Recognition Aerial Surveys.
- WUR AgroFoodRobotics is a group within Wageningen Research with more than 10 years of experience in research and development of AI and image recognition software, mainly in the Agriculture and Greenhouse Horticulture sectors, but recently also more in the marine

environment in collaboration with WMR. They will develop the AI together with development partners.

- Observation International has already gained a lot of experience with automatic image recognition. Dylan Verheul is co-founder of the Waarneming.nl/Observation.org platform and director of the Observation International foundation. He will contribute best practices/lessons learned from his role as client for building AI/Deep Learning model for image recognition species and as person ultimately responsible for the implementation of image recognition as part of their platform (incl. mobile apps; ObsIdentify).
- BioConsult is a German-based company of which the English company HiDef Aerial Surveying Ltd is a part. Since 2014, BioConsult SH has been conducting research on seabirds and marine mammals in the North Sea and the Baltic Sea using the high-resolution digital camera system developed by HiDef. The HiDef method has set a new standard internationally for marine ecological research and provides high quality data for impact assessment studies and conservation projects. Within this project, BioConsult will perform the digital aerial surveys above the Gemini wind farm and annotate the images. In addition, they will contribute their extensive experience and expertise with image recognition.

An advisory committee will be set up that will ideally consist of, but not be limited to, Gemini wind farm, Rijkswaterstaat (WVL and CIV Datalab), Royal Belgian Institute of Natural Sciences (RBINS) and Bureau Waardenburg. All these parties have experience with or are involved in digital aerial survey projects and their analysis.

Annex 2 Involved parties

This roadmap (and the associated PPS) has been developed by a wide range of contributors, representing an equally wide range of (international) organisations. To continue the work of developing an AI for aerial survey image recognition a commitment to participate in a future project team or take seat in a guidance committee was given by most of the involved organisations.

Parties involved in the workshops and drafting of this roadmap:

- **WUR AgroFoodRobotics** is a part of Wageningen Research with over ten years' experience in research and development in AI and image recognition software, originally active in agri- and horticulture, since recent years collaborating with WMR on topics related to automated biodiversity monitoring. Responsible for the AI insights in this project.
- **WUR Wageningen Marine Research (WMR)** has a broad range of seabird and marine mammal expertise. WMR is project leader of the PPS project, responsible for project management and for organizing the cooperation between PPS project partners. WMR has been conducting (international) seabird and marine mammal aerial surveys for many years, including surveys for the SCANS projects.
- **Observation International,** partner in the NatureAI consortium (that also includes Naturalis Biodiversity Centre) that has developed a species identification model for identification of all species on images that is made available on their platform Observation.org and the mobile app ObsIdentify. This party has offered to serve as an advisor for AI and platform developments with the potential to also participate in the development team.
- **Bureau Waardenburg (BuWa)** is an ecological consultancy that has been conducting seabird and marine mammal aerial surveys for Rijkswaterstaat for many years, including surveys for the MWTL program.
- **Rijkswaterstaat CIV Datalab** is a team part of Rijkswaterstaat working on data storage, software and AI development.

Contributors for the proposed development could include:

- **Rijkswaterstaat (RWS); CIV Datalab** collects data at sea in a wide range of projects potentially other survey data could constitute a useful contribution to the proposed AI development
- **BioConsult** is a commercial company in Germany, in which the UK-based company **HiDef Aerial Surveying Ltd** is embedded. BioConsult SH has conducted seabirds and cetacean studies since 2014 in North and Baltic Seas, using the digital high-definition camera systems and software developed by HiDef. The HiDef-method has set an international standard for marine research and provides high quality data for impact and conservation studies. BioConsult will conduct the digital aerial surveys for both Borssele and Gemini within this project and will annotate the collected footage. BioConsult and HiDef can share their extensive experience in image recognition with the software developers.

It might be good to also set up a **Guidance committee** to receive direct feedback from parties involved in such a development. Such a committee could include:

- **Gemini offshore wind farm** or any other windfarm operators for insight into what the wind farm owners consider relevant
- **Rijkswaterstaat** (WVL) for insight into what the policy makers need for information
- **Royal Belgian Institute for Natural Sciences** (RBINS) or other international institutes for insight into the developments abroad
- **Bureau Waardenburg** bringing in their experience with bird counts both from ships as from planes

These bodies are all highly experienced in offshore wind farm impact studies and/or digital aerial surveying and data analysis.

Wageningen Marine Research
T +31 (0)317 48 7000
E: marine-research@wur.nl
www.wur.eu/marine-research

Visitors' address
- Ankerpark 27 1781 AG Den Helder
- Korringaweg 7, 4401 NT Yerseke
- Haringkade 1, 1976 CP IJmuiden

With knowledge, independent scientific research and advice, **Wageningen Marine Research** substantially contributes to more sustainable and more careful management, use and protection of natural riches in marine, coastal and freshwater areas.

Wageningen Marine Research is part of Wageningen University & Research. Wageningen University & Research is the collaboration between Wageningen University and the Wageningen Research Foundation and its mission is: 'To explore the potential for improving the quality of life'