UiT The Arctic University of Norway

The Faculty of Science and Technology, Department of Computer Science.

Combination of satellite imagery and wind data in deep learning approach to detect oil spills

Hans Berg Borhaug

INF-3990 Master's Thesis in Computer Science – September 2021

Abstract

The ocean is vulnerable to oil related activities such as oil production and transport that can harm the environment. Environmental damages from oil spills can be large if not dealt with. Satellite images from radar are useful to detect oil spills because they cover both day and night and penetrates clouds. However, detecting oil spills in ocean areas from satellite images are not a trivial task due to abundance of lookalikes from other natural sources, like river inputs or geological seepage. Auxiliary data such as wind speed in the monitored area, are used to separate oil spills from natural occurring slicks in the manual oil detection process. One solution to detect oil spills is applying artificial intelligence techniques like convolutional neural networks. These convolutional neural networks have usually been a candidate to create an automatic oil detection process. However, the convolutional neural networks have problems with distinguishing between spilled oil spills and look-alikes.

This project is about exploring the possibility of detecting oil spills from satellite images and distinguish between spilled oil spills and natural occurring ones by using wind speed data of the area. The convolutional neural network takes in both satellite images and auxiliary wind speed data of the area monitored.

Two convolutional neural networks are designed and setup, where one includes auxiliary wind speed data and the other does not. Both CNN's will have the same satellite images and oil spills to detect such that a direct comparison can be made between them. This work will also be a proof of concept to an automated oil spill detection process that specifically uses wind data in addition to the satellite images.

To measure any difference in validation loss, precision or recall by using wind data, both convolutional neural networks are tuned to the same recall such that the false negatives are as low as possible for both neural networks. The comparison between the two neural networks shows that the neural network that includes wind data has 15 % lower validation loss and a slightly higher precision than the neural network that does not include wind data. However, this result is achieved by using wind data generated from the satellite image itself, which metrological wind data is not. A comparison test like this but with metrological wind data instead of wind data generated from the satellite image is considered future work that is worth exploring.

Keywords
Convolutional neural network, Oil spill detection, Synthetic Aperture Radar, Wind speed, Auxiliary data

Table of Contents

1 Introduction

Oceans are vulnerable to the activities of trafficking and energy production like oil, gas and wind that can harm the environment. Environmental damages from leaks that creates oil slicks (also called oil spills) can be large if not dealt with [1]. Monitoring oceans for oils slicks can be done in several ways like patrolling the sea and taking photos from the air, but some of the more efficient ways are using satellites that are able to take images of large areas from their altitude. More specifically, images from Synthetic Aperture Radar appear to be able to detect oil slicks well, since the radar does not get affected by weather and lack of light due to their penetration of clouds and active illumination [2].

1.1 Background and motivation

Oil spills are a source of hydrocarbons in the ocean. Hydrocarbons are molecules of solely carbon and hydrogen. They usually come from human-made sources that spill them in oceans, but they occur naturally as well [3]. Human-made sources like vessels and oil rigs are the most common to identify whereas natural sources are from seepage, usually around rivers along coastlines [3]. In addition, there exists look-alikes to oil slicks that occur from still waters and biological sources. These look-alikes and actual oil spills appear very similar and differentiating between them can be problematic. False alarms of oil spills are expensive due to redundant use of resources to further monitor and combat such an oil spill. To combat false alarms, the amount of wind in the same area can reveal spilled oil slicks from look-alikes [4]. Thus, both radar satellite images and wind data are used to conclude which of the detected oil slicks are real or not, which makes it possible to avoid more look-alikes. Currently, the work of detecting oil slicks is manual work carried out by a trained crew at a company named Kongsberg Satellite Services (KSAT). There is also other firms that do the same work in the world.

KSAT [5] is a partly state-owned company in Norway. The company monitor and report activities in oceans by analyzing images retrieved from satellites. One of the company's objectives is to report any oil slicks detected. Detecting oil spills from different sources like vessels and oil production platforms are done manually today with the help of computer software. KSAT produces reports with oil spills from areas monitored by satellite images. These reports include where, if any, spilled oil slicks appear and suggest possible oil slick sources from auxiliary data, like the automatic identification system (AIS). The work of the department has time constraints from when images are received to when marked oil slicks are to be delivered. This is because oil detections are time sensitive to discovery, as earlier warnings will provide valuable time to limit the damages.

KSAT is in the process of automating oil slick detection from satellite images to provide an oil slick detection service to their customers. To automate the oil detection from satellite images, a convolutional neural network (CNN) is a possible solution. CNN's [6] are one of several types of artificial neural networks in the field of artificial intelligence and deep learning.

An artificial neural network (ANN) has a collection of artificial neurons that are inspired by biological neurons in the human brain. Artificial neural networks work by giving some input values and expect some output values from the neurons. The neural network are trained by giving penalty to wrong output values in each neuron. These properties of an ANN are useful to create some automatic behavior from a given input. A subgroup of artificial neural networks called deep neural networks, DNN, has a property of learning a greater amount of information than simpler artificial neural networks. This is useful in image analysis tasks such as detection and classification of objects, but also in other fields such as text analysis and speech recognition. Deep neural networks consist of several layers of neurons, where each layer feeds the next with their output value. One use case of deep neural networks is image classification, where objects are identified and recognized. A CNN is a type of deep neural network which is particularly good at image detection and classification. CNN's use and combines ANN-components in such a way that CNN's excel in tasks with grid-like structures, like images. The CNN has revolutionized the field of computer vision the past decade.

Because of these properties a CNN is a candidate to detect oil spills from the satellite images provided by KSAT. Furthermore, a study [7] of using CNN's to detect oil spills shows that these networks can be utilized to implement efficient oil spill detectors. Thus, a CNN is expected to play a part in the solution to the automation of oil spill detection.

1.2 Problem

Identifying oil slicks and discriminate between spilled oil slicks and look-alikes are both difficult problems. To solve these problems, the wind data can be treated as an input value alongside satellite images to a CNN. However, from this proposal another problem arises. For example, the data are not in the same format. Moreover, the wind datapoints of speed and direction are not structured datapoints with high resolution like satellite images, and structured datapoints is a requirement to CNN's [8]. Then, the problem is that the structure of wind data has different structure than the satellite images. Thus, it is demanding to put together these two data types.

1.3 Hypothesis

The most direct way of comparing two structurally equal neural networks against each other is by comparing the validation loss. The validation loss is the penalty of wrongly classifying oil or background from a given loss function. This is the main measurement used in this project.

Another typical measurement of performance of the detection application is the use of precision and recall. The precision and recall are defined below:

- Precision is defined as the fraction of actual oil slicks detected among detections regardless of actual or not.
- Recall is defined as the fraction of actual oil slicks detected among all actual oil slicks.

Wind speed is a determining factor in the detection process of oils slicks [4]. By comparing a CNN with wind speed as input against a CNN with identical structure without wind speed as input, the hypothesis is that the neural network with wind speed as input has a lower validation loss of 0 to 15 %. An additional hypothesis is that the same CNN with wind speed as input has a higher precision and recall by 0 to 5 percent points and 0 to 1 percent points.

1.4 Purpose

The main purpose of this thesis is to present and describe the development of two CNN's. One CNN can handle a combination of satellite images and wind data to detect oil spills and the other CNN can handle only satellite images with the same task.

1.5 Goals

There are several goals in this project. These goals are listed below.

- An algorithm that transforms wind data into an input that can be accepted by the CNN.
- A CNN that can detect oil spills while ignoring look-alikes to some degree.
- Provide a proof of concept of automated oil spill detection with wind data along satellite images as inputs to a CNN.

The proof of concept is valuable to KSAT as they are exploring possibilities of automating oil spill detection. A successful automated oil spill detection system can benefit society by reducing costs of monitoring and possibly decrease detection time by advancement in efficiency.

1.5.1 Benefits, Ethics and Sustainability

The project and its results will show a potential proof of concept, or not, to KSAT and other interested parties in the oil detection business. This can potentially provide a positive impact of protecting the environment by more easily preventing disasters related to oil pollution. In addition, this can potentially lower the cost of earth observation service of detecting oil slicks. In overview, the potential benefits to society are less oil pollution at sea with a lower cost of preventing it in the long run.

The ethics of this project is on topics such as potential loss of jobs. If this project is working towards an eventual automated oil detection service, then several jobs in that service today are at risk. However, even if those jobs were at risk, the benefits are that resources used on manual work can be directed to other sectors that are important and potentially create more jobs. Other topics are for example loss of sensitive data to the public or foreign intelligence. However, this is unlikely to happen due to the usage of that sensitive data is performed exclusively at the facility of KSAT, which has the appropriate security standards. The data will never leave the building, and only image examples used in this thesis are a trace of it and is approved by KSAT to be released.

The sustainability of this project is about how a potential better automated oil spill detector can affect economics, society, and environment where it is used. There are potential economic benefits because the cost of operating an oil spill detection service is probably cheaper with less humans and more working computers. In addition, the work can potentially be done faster, giving other positive side effects to those who need the service. This is usually companies that need to monitor themselves and governmental agencies that monitor the ocean. Society can benefit as well since oil spills can potentially be detected faster and dealt with quicker, limiting the burden of dealing with oil covered shores and further implications. This can potentially benefit the environment as well, since dealing with it faster can limit the damages to shores, wildlife, and ecosystems.

The task of detecting oil spills is an important service to the environment and the consequences of not detecting actual spills are greater than detecting false positives. This is taken into consideration in this project.

1.5.2 Preventing release of confidential material

There will be used confidential material in the form of data and tools which are not to be publicly available. Thus, this material will not be included in this thesis if published and will remain a goal during the work with this project to obscure it to any other party than KSAT. Usage of this material will be exclusively on the location of KSAT in Tromsø with the necessary security measures in place. KSAT has agreed to usage of their material, which is in form of image data and products of these, as well as tools.

1.6 Methodology / Methods

In general, there are two methods of researching:

- *Quantitative research method*, which is in general in support of deductive and experimental methods. It supports measuring variables, experimenting, and testing hypothesizes and usage of statistics to verify the results [9].
- *Qualitative research method*, which is in general in support of inductive and theoretical methods. It supports discovering and understanding meanings, opinions, and behaviors. The understanding of such things can help the development of tentative hypothesizes or new inventions [9].

A triangulation of the two is possible, but some methods of each direction may not be compatible to achieve efficient research [9]. This project will use the *quantitative research method*.

1.6.1 Research method

There are several methods of researching that can apply to the topic of this project. However, not all are appropriate to the purpose of this project. There is *experimental research*, which studies causes and effects by manipulating variables and study their relationships. In contrast to experimental research there is *non-experimental research*, which draws conclusions from observations of already existing scenarios. In addition, there is *descriptive research*, which studies new observations and document them without drawing any conclusions. This method is useful to document phenomena which not necessarily has an established explanation yet. There are other research methods as well, such as *analytical research*, *fundamental research*, *applied research*, *conceptual research,* and *empirical research* [9].

The project is about detecting oil slicks from the source material of satellite images and wind speed of the area, which demands a large amount of data to achieve a network that can detect oil slicks with a precision and recall that is similar to the manual work. The CNN contains several variables that will need to vary to experiment. Thus, the research method of this project will be *experimental research*.

1.6.2 Research approach

The research approach will guide in concluding statements from research [9]. There are several approaches to this, which includes:

- *Inductive approach*, where one formulates theories from observed data usually collected with qualitative methods and in quantities that is enough to support the theories [9].
- *Deductive approach*, where one test theories or hypothesizes from usually quantitative methods with large datasets [9].
- *Abductive approach*, where one uses both an inductive and a deductive approach to conclude which hypothesis that best explains the data available and result in a plausible explanation [9].

The research approach of this thesis will be to test the hypothesis previously stated about the effect of wind data in a CNN with a quantitative approach of experimental data. CNN's are testable and comparable with the expected output (ground truth) from KSAT, which delivers services on oil-detection and -classification. Thus, a *deductive approach* is most appropriate.

1.6.3 Research strategy

Research strategy is guidelines of researching [9]. The most common research strategies to quantitative research are:

- *Experimental research*, where the strategy verifies hypothesizes by altering factors that affects experiments and provide an overview of causes and effects [9].
- *Ex post facto research*, where the strategy verifies hypothesizes by observing already existing experiments and provide an overview of causes and effects [9].
- *Surveys*, where the strategy assess attitudes and characteristics of a subject. It describes phenomena of frequency and relationships between variables that are not directly observed [9].
- *Case study*, where the strategy are empirical studies of phenomena where there are blurry boundaries between phenomena and context [9].

This project will use large amount of data to verify or refute the hypothesis previously stated about the effect of wind data in a CNN. The research strategy will be to alter factors that affects the results to test the hypothesis. This is expected to lead to an overview of causes and effects. Thus, an *experimental research* strategy is most appropriate.

1.7 Delimitations

The delimitations of this project are the information stored in the training data, which will limit the capabilities of the trained model, and the tools available when this project is done, which will limit the performance of the trained model.

The data does not cover every scenario of an oil slick. The effect of this is that the trained models will not be able to detect every oil slick in every condition. Regardless, it is assumed that the large amounts of data will cover most known scenarios and thus has the potential to guess other scenarios that is close to them. In addition, the model can train on other scenarios should they appear later.

Other delimitations are the work of KSAT to present the ground truth to where actual oil slicks are in the satellite images. The data stems from operational services, where speed is a higher priority than pixel perfect segmentation labels. It is assumed the data are good enough for this project to succeed.

1.8 Contribution and stakeholder

The main stakeholder to this project is Kongsberg Satellite Services, KSAT for short. KSAT expects to collect and interpret the results from this master project and how it can contribute to their current work in the field of oil detection. The contribution of this project to them is to show how a convolutional neural with satellite images and wind data performs compared to CNN with only satellite images and show one way of including wind data into a CNN.

1.9 Outline

In chapter 2, 3 and 4 several topics are clarified to create a foundation for the original work of this project in chapter 5 and onwards. These topics include what type of data is used, what type of techniques are used in detection, and scientific methods. Chapter 5 includes the design and requirements of the project. Chapter 6 will go into details of the implementation of this project. The results are shown and discussed in chapter 7. These results include wind interpolation, validation loss, and detections. The evaluation of the project is described in chapter 8 and the possible next steps after this project are described in chapter 9. Finally, the thesis is concluded in chapter 10.

2 Synthetic Aperture Radar and Related Data Types

Topics related to this project is presented in this chapter. This chapter describes what oil slicks do to the surface of the ocean, the type of wind speed data KSAT has, type of satellite images and why they are used in oil slick detection. It also describes deep neural networks that are used and why they are appropriate for oil slick detection. Additional topics like interpolation and measurements techniques are described to clarify how they work. Related work includes previous attempts to detect oil slicks and how to automate the process.

2.1 Oil slicks and look-alikes

One key aspect of oil detection is to distinguish between oil spills and look-alikes. Oil spills have usually recognizable patterns and does not vanish easily by higher wind speeds. However, they come in different shapes and can resemble look-alikes when the wind is low enough. Sources of oil spills are usually oil rigs, oil tankers, and oil pipes and seeps from the bottom of the ocean [10]. Look-alikes are usually more diffuse with less defined edges but can in some cases resemble oil spills. They usually appear with lower wind speeds and in the wind-shadow of islands. The sources of such look-alikes are several, such as rain, grease ice, wakes, and natural films from fish oil, vegetable oil and algae [11].

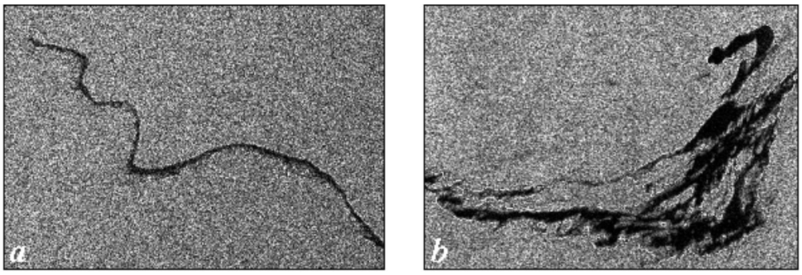

Figure 1: Example of oil spill and look-alike from a paper called *Observing marine pollution with Synthetic Aperture Radar* [12]

The figure above shows examples of an oil spill to the left and a look-alike to the right. The oil spill to the left is more defined and has trail that likely follows the path it has taken from the source. The look-alike to the right resembles oil but is more diffuse in its shape.

2.2 Damping of ocean waves by oil slicks

Oil slicks on the surface of the ocean dampens ocean waves in rough seas and the dampening increases quadratically by higher wind speeds [13]. This fact makes it possible to differentiate oil slicks from the surface of the ocean by observing ocean waves given that the wind speed is sufficiently high enough. In addition, this strong dampening on wave lengths larger than some maximum value is not possible by any known substances that exists in the ocean surface naturally [13]. Thus, it is highly probable that any substance that causes strong dampening in waves lengths larger than this maximum value (that is correlated with higher wind speeds) is of human origin.

From this it is apparent that wind speed has an important role in detecting oil slicks. However, wind speed data must be collected such that it is useable with associated satellite images.

2.3 Types of wind data at KSAT

KSAT supplies two types of wind data. These are under the names of MET-wind and SAR-wind internally at KSAT and the same names will be used in this thesis as well.

MET-wind is a result of an advanced process of data from sensors at sea, weather balloons and region and global weather-systems. This wind-data is like a weather report and creates a grid-structure of wind speeds and directions over the sea monitored. This wind-data has a lower typical resolution of wind points per area covered than SAR-wind.

SAR-wind is a result of a combination of the associated radar image and data from MET-wind [14]. The algorithm that creates the SAR-wind data is made by NORCE [15] but the algorithm is not public knowledge. The general idea is that wind speed is retrieved from the features of waves in the ocean. However, since SAR-wind has a 180-degree ambiguity a wind direction cannot be determined. Thus, MET-wind is used to determine wind direction at KSAT. This wind-data has a higher typical resolution of wind points per area covered than MET-wind.

Only SAR-wind was used in this project since this wind type was sufficient to test if there are any difference in results between including wind data and not. In addition, MET-wind has a connection to the satellite image like SAR-wind. Thus, results from MET-wind are not expected to deviate much to results from SAR-wind. That said, it is not known what the difference might be.

2.4 Synthetic aperture radar

The satellite images used in this project are created by a *synthetic aperture radar*, or SAR [2]. It is a form of radar that can create images from radar pulses that illuminates areas where the echo of those pulses is received and processed [2]. SAR placed in satellites takes advantage of the motion of a satellite to simulate a larger antenna. From this larger antenna it is possible to achieve higher resolution images than physical antennas [2]. However, there is a tradeoff in resolution and area covered, such that resolution is sometimes not favored to cover larger areas [2].

2.4.1 Polarization channels

The relevant SAR images to this project comes in different types of combinations of polarization, with a total of 4 combinations. These combinations are made of either vertical or horizontal polarization in transmission and receival [16]. These combinations of different polarizations are divided into two different categories from their best use cases, like oil [17] or ships [18]. One category is altering polarization where the transmitted and received polarization alters from horizontal and vertical or vice versa [16]. These polarization combinations appear to reveal ships and other objects in the ocean better than the other category [18]. The other category of polarization combinations is non-altering polarization where the transmitted and received polarization are the same [16]. These polarization combinations appear to reveal oil and waves in waters better than the other category [17]. The non-altering polarization combinations category will be the most relevant in this project since those combinations are superior in detecting oil slicks.

2.4.2 The project's satellite imagery and wind data

In this project it is used satellite images from SAR in combination with wind data over the given area to be monitored.

The satellite image from SAR shows features from the ocean where features that resemble oil are searched for. The benefits of using SAR images are the availability of vision at night and through clouds. The satellite actively sends radar signals that are reflected by the ocean such that features of waves and other motions of the ocean are visible.

The wind data from KSAT shows geographical points where a wind speed is measured. These geographical points are then transformed into an image of same size as the satellite image, where the empty space between wind points is estimated.

By combining satellite images with wind data, the CNN receives information from both data sources to decide if the features that resembles oil are to be classified as an oil slick. The wind data has the role of giving a signal to discarding look-alikes since they often appear with lower wind speeds.

3 Deep Learning and CNN's

This chapter describes deep learning that is used in this project and why this is chosen for oil slick detection. Additional topics are overfitting, interpolation, and measurements of accuracy in detecting oil slicks. In the end of the chapter, related work is presented with previous attempts to detect oil slicks and how to automate the process.

3.1 CNN's

Convolution neural networks from deep learning provide the opportunity to specialize machine learning models to grid-like structures and with the opportunity to scale the model to large sizes for enhanced learning capabilities [8]. This subchapter is going to discuss the relevant deep CNN's for this project.

ResNet, or Residual neural networks, are deep neural networks that can extend to greater depths than other conventional deep neural networks to achieve higher accuracy in object detection by using skip-connections between its layers. ResNet won the ILSVRC object detection task of 2015 [19]. ILSVRC is short for ImageNet Large Scale Visual Recognition Challenge and is about evaluating algorithms for object detection and image classification at large scales. The object detection challenge specifically penalizes algorithms for objects not detected and duplicate detections for several classes of objects to generate an accuracy score. ResNet had the highest accuracy score of this challenge. Thus, ResNet proves useful in image-based detection and has the advantage of a variable size to fit the need of the data available to learn the model.

An option considered is the VGG network, which is also a CNN. This network proves useful in image-based classifications as it also won the object localization challenge in ILSVRC 2014 [20]. The object localization challenge is a slightly modified detection challenge where only a single object for each class is detected even though there are more objects of the same classes. However, the VGG network has worse error rate than ResNet in an Image Net comparison done by the ResNet paper [19].

There are multiple ways of using such CNN's, like detection, classification, and semantic segmentation. The latter is the most relevant to this project because oil spill data from KSAT is arranged for semantic segmentation specifically.

3.2 Semantic segmentation

Semantic segmentation is a technique of classifying each pixel in an image. It differs from object-detection by classifying every pixel instead of a focused area of the image. By using CNN's to achieve semantic segmentation, there usually is an end-to-end learnable architecture that uses deep CNN architectures as an encoder, where information is compressed, and a mirrored version as a decoder, where the information is decompressed again. This combination is illustrated in figure 2, which is from the SegNet-architecture paper [21]. This figure shows a typical CNN architecture used in semantic segmentation, where different colors represent the different types of layers used in such an architecture. The flow of information goes from left to right, starting with the input image and ending up with the segmented image.

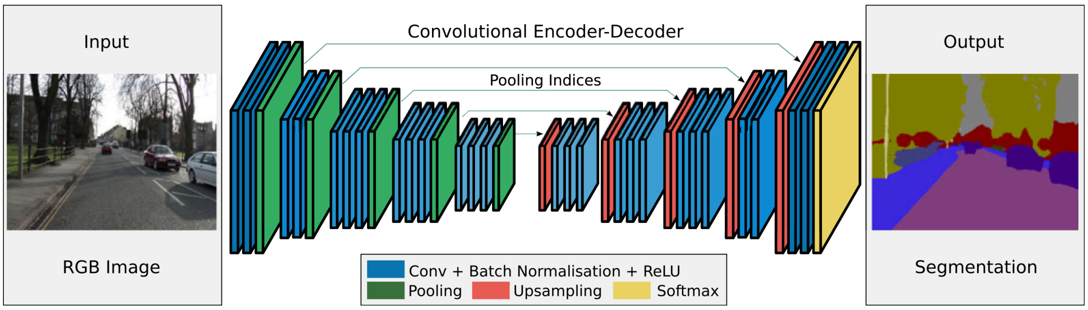

Figure 2: Semantic segmentation illustration from the SegNet-architecture paper [21]

3.2.1 U-Net

The architecture of a CNN that is mirrored but without fully connected layers in the middle is often called a U-Net architecture [22]. The convolutional layers to the right increase the resolution to achieve a replication of the important information at a higher resolution. By using this architecture with semantic segmentation, an output image of the same resolution as the input image is possible.

3.2.2 Training CNN's

The training of CNN's works by giving some input value, such as an image. This input value is forwarded through the CNN and finally gives some output value. The CNN gives predictions in this output value, for example where an oil spill is detected. The output value from the CNN during training is then compared against some preferred output value, such as a two-dimensional grid with whole numbers for each class, marking the actual class of each pixel. This comparison is used to alter weights in the CNN in such a way that it corrects the error. This comparison is specifically calculated by a given loss function.

When there is imbalance in training data, the training of semantic segmentation networks handles the imbalance by using a specific loss function.

3.2.3 Loss function

In semantic segmentation tasks with images, one of the commonly used loss functions is pixel-wise cross entropy loss, where each pixel has its own loss from all possible classes, and the loss value for the whole image is calculated from the average of all pixels. Since the data for this project is very unbalanced between each class, the training with this loss function is likely to result in the most abundant class dominating the learning progress [23]. There are several ways of combating this. One way is weighting the loss for each output channel (class) as done in the *Fully CNN's* paper [24]. Another way is pre-computing a weight map where there is a higher weight at borders of objects in the segmentation as proposed in U-Net [22].

Another loss function is Focal Loss [25]. In focal loss, the loss function addresses imbalance between detections and background during training by down-weighting easy classified negatives. Focal loss comes with a parameter that tunes the down-weighting depending on the class-imbalance in the data.

There are other challenges such as overfitting, which applies to machine learning models in general.

3.3 Overfitting

When training a network, a model needs to make both the training error and the gap between training and test error small to perform well. Consequences of where the gap between training and test error is too high are challenges like overfitting. Overfitting tends to apply with models of too high capacity of learning. To combat this either the amount of training sessions needs to decrease, or the data set needs to increase. [6]

To generate any data sets of wind data that match the resolution of satellite images, a technique called interpolation is worth considering.

3.4 Interpolation

Interpolation [26] is a technique to estimate new data points from a set of known data points. This technique is possible to use with images and thus increase resolution based on a combination of estimated and known pixels. In this project two ways of doing interpolation are considered, bilinear and bicubic interpolation. Both are described and compared.

Bilinear interpolation is linear interpolation in a two-dimensional grid of data points. Linear interpolation is interpolation by creating new data points on a linear line between two known data points. By combining two dimensions of data points, the linear interpolation of both dimensions combined is bilinear interpolation. Bilinear interpolation in images takes a total of 4 pixels into account for every pixel.

Bicubic interpolation is cubic interpolation in a two-dimensional grid of data points. Cubic interpolation is interpolation by creating new data points on a polynomial curve between four known data points. By combining two dimensions of data points, the cubic interpolation of both dimensions combined is bicubic interpolation. Bicubic interpolation in images takes a total of 16 pixels into account for every pixel.

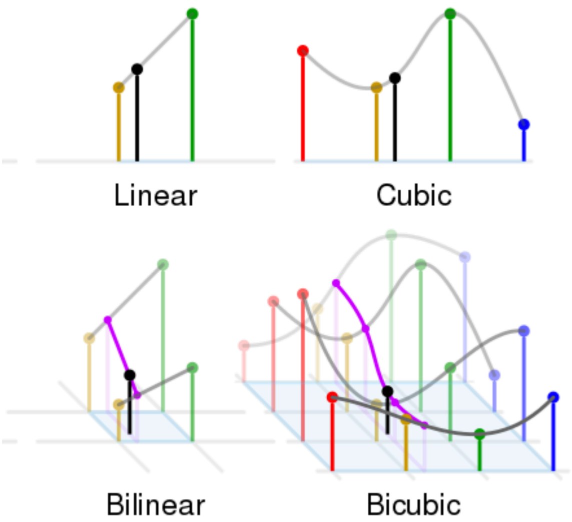

Figure 3: Illustration of linear, cubic, bilinear, and bicubic interpolation from the Wikipedia page of cubic interpolation[1] [27].

Figure 3 illustrates linear and cubic interpolation in both one dimension and two dimensions. The black dot is the interpolated value (pixel), while the other coloured dots are neighbouring values (pixels). The height of the dot is the scalar of the pixel value in the image.

Bilinear interpolation, compared to bicubic interpolation, has the advantage of being less complex in calculations and thus fitting in time-sensitive situations. However, bilinear interpolation is known to cause more interpolation artifacts than bicubic interpolation. Bicubic interpolation is known to have the least number of artifacts compared to both bilinear interpolation and simpler techniques like nearest-neighbor interpolation.

3.5 Precision, recall and F-measure

Precision and recall are calculated from the number of false positives and false negatives compared to the number of correct detections. The correct detections include only the oil spill class and not the background class. False positives are detections made when there was nothing to detect. False negatives are failures to detect something which was present.

Precision is a measure of the false positive rate. It is the ratio of correct detections among all detections, where a higher ratio is better. The definition of precision is as follows:

$$Precision = \frac{Correctly\ detected}{Correctly\ detected + False\ positives}$$

[1] Made by user: Cmglee, under CC BY-4.0 at: https://creativecommons.org/licenses/by-sa/4.0/ (9.9.2021). This image is cropped from the original by excluding an additional method of interpolating.

Recall is a measure of the false negative rate. It is the ratio of correct detections among all expected correct detections including false negatives, where a higher ratio is better. The definition of recall is as follows:

$$Recall = \frac{Correctly\ detected}{Correctly\ detected + False\ negatives}$$

F-measure is calculated from a combination of precision and recall values and thus says something about both in a single value. The definition of F-measure is defined below:

$$\text{F measure} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

3.6 Related work

There is no known (to the author of this thesis) related works that explores interpolating wind speed as an auxiliary data source to be used in a convolutional network. However, other related work to this project is known to be based on a model of automatic or semi-automatic detection of oil slicks based solely on images with image-altering techniques to discover and separate them from natural occurring slicks.

One example from a paper [28] with automatic oil detection shows auxiliary data like wind speed used as a filter on the output data from a neural network, not as an asset inside it. This paper suggest wind speed is the most important parameter [28]. The paper validates oil spills from wind speed between 2 and 14 m/s, where detections with above 10 m/s wind speed are unmistakably oil spills when detecting from SAR images [28].

In addition, one comparison of different algorithms such as clustering, logistic regression, and CNN's that detects oil slicks with image segmentation concludes that CNN's are more advantageous in certain areas. These areas are more automatic segmentation, better balance between precision and recall of detected oil spills and a minimal computation time [29].

4 Scientific methods

This chapter describes the scientific methods used in this project. This includes data collection, data analysis, and quality control. Different methods are described, and the method chosen for this project is discussed. The validity, reliability and replicability of the quality control are also discussed. Finally, software engineering methods are briefly presented.

4.1 Data collection

Some data collection methods used in quantitative research are experiments, questionnaire, case studies and observations, while other data collections methods used in qualitative research are questionnaire, observations, interviews and language and text [9].

- *Experiments* collects large data sets from changing variables [9].
- *Case studies* collects data sets from some number of participants in case study research [9].
- *Observations* collects data sets from observing behaviors with a focus on different situations [9].
- *Questionnaire* collects data from questions that are either closed like multiple choices or open like reviewing questions [9].
- *Interviews* collects data from a somewhat structured deep understanding of a problem through a participant's point of view [9].
- *Language and text* are usually collected through queries of sources of text and recordings of verbal language such as libraries and databases [9].

The experiments in this thesis need large amounts of data to test the hypothesis. This applies to both satellite images, wind data and markings of oil slicks (ground truth). This is a natural consequence of using deep neural networks to solve a problem. The data is expected to be available from KSAT. Thus, the data collection method of this thesis is through *experiments*.

4.2 Data analysis method

The method of analyzing data is the method that analyses the collected data by inspecting, cleaning, transforming, and modelling it. Some data analysis methods used in quantitative research are statistics and computational mathematics, while other data analysis methods used in qualitative research are coding, analytic induction, grounded theory, narrative analysis, hermeneutic, and semiotic [9].

- *Statistics* analyzes data through calculating results for a population and evaluating the significance of the results [9].
- *Computational mathematics* analyzes data through numerical methods, modelling and simulations [9].

- *Coding* analyzes interviews and observations through transcription to transform qualitative data to quantitative data [9].

- *Analytic induction* analyzes data through iterations between collecting and analyzing data. The iterations stop when the hypothesis ends with a validated theory when no cases dismiss it [9].

- *Grounded theory* analyzes data in the same way as analytic induction by iterations until a validated theory emerges [9].

- *Narrative analysis* analyzes literacy by discussing meaning of text (*Hermeneutic*) or meaning of sign and symbols (*Semiotic*) [9].

The project is about finding one or several algorithms that has two input sources and detects oil slicks as the output. The algorithm will be a type of deep neural networks, which are based on mathematical calculations to find the optimal weights. Thus, the data analysis method of this project is *computational mathematics*.

4.3 Quality control

Since this quantitative research with a deductive approach, the following topics must be discussed:

4.3.1 Validity

To achieve valid results, the algorithms will be applied against statistical demands when tested against an independent dataset (validation set). For example, false positives or negatives will decrease the validity of the results. Thus, this project will actively prevent these, while at the same time prioritize which is more important to prevent.

To get a deep learning algorithm to learn the correct behavior, it is important to have ground truth that represents the expected results. The ground truth of oil slicks is available in earlier oil detection services, which are produced by hand of KSAT employees.

4.3.2 Reliability

A common problem with deep learning algorithms is overfitting, which can make the results less reliable. There are different methods to combat this and they will be considered during the design process of the neural networks.

4.3.3 Replicability

The replicability of this project is dependent on the availability of the data used. Thus, only parties with access to the same data can do the exact same experiment. KSAT is the owner of the data of detections and decides who can use it for what purpose. Data from satellites alone are open to the public.

4.4 Software engineering method

Determining which software engineering method is most appropriate depends on several factors, such as the topic and the requirements of the project. Some of these software engineering methods are for example waterfall, agile, prototyping, incremental development, and extreme programming.

4.4.1 The waterfall method

The waterfall method is a software process model that is a simplified representation of how software is developed. The method requires planning of the whole development process, and each stage of the development process to be completed before moving on to the next stage [30]. Each of these stages has usually a schedule of completion. When the development is finished, the remaining operation and maintenance is done at the stage necessary. Repeating stages after the maintained stage may be necessary to ensure the often-strict requirements of the software developed.

4.4.2 Agile methods

The agile methods are an umbrella term for several methods that follows the philosophy of agile software development. This philosophy is about a more incremental development process with looser stages [31].

4.4.2.1 Test-driven development

One of the agile methods are the test-driven method. This method approaches software development and testing by interchanging these stages for each incremental feature developed [32]. One advantage from this is the method allows features to be tested before the whole system is up and running. This test-driven development can catch issues sooner than a later testing stage in development.

4.4.3 Prototyping method

The prototyping method involves engineering an executable model of a system, which can be a proof of some concept. This executable model is then tested to see if it meets the requirements of the stakeholders. Stakeholders will then give feedback of any requirements changes or if the development should proceed.

4.4.4 Chosen method for building two CNN's

The chosen software engineering method for building two CNN's is a combination of several methods. This combination includes methods such as agile, prototyping and proof of concept. Since one of the goals is to provide a proof of concept of automated oil spill detection with wind data along satellite images as inputs to a CNN, the strict nature of the waterfall method is unnecessary and may hinder development during the time of this project. Thus, a more agile method was chosen with test-driven development steps. This development include testing and acquiring proof of the developed code works as intended. However, only one version of the code was developed since the software developed in this project is intended for a proof of concept, not in production at KSAT.

The agile method, used in this project, includes prototyping of implementations of architectures necessary to fulfil the goals of this project. Several implementations were expected to be altered or even discarded during the development process. In addition, several key resources have not been available at the ideal time. This includes resources such as office space to handle data during the development process because of the strict security with codebase and data at the facility of KSAT. In addition, the follow-up from KSAT was not ideal, which was directly and indirectly cause by the ongoing covid-19 pandemic due to prioritization of home offices of employees and other issues. Thus, the development process was improvised at times and tasks such as further planning, design and documentation was done in the meantime. Implications from this was some chaotic development process at times. However, most software engineering was done in an incremental fashion where parts were built on top of each other as the project moved forward. For example, the transformation of wind data was implemented and tested before the neural networks since one of the two CNN's depended on it.

5 Project Design

This chapter will go into detail about the design of this system. There are roughly three different parts that are designed.

- A process of transforming unstructured wind datapoints into structured datapoints
- A neural network that detects oil slicks with satellite images and wind datapoints
- A neural network that detects oil slicks with satellite images but without wind datapoints

The type of neural network will determine the requirements of the transformed wind datapoints.

5.1 Requirements for CNN

The CNN's, built for this project, have a requirement where one CNN shall handle satellite images and the other CNN shall handle both satellite images and wind data. Consequently, the number of input channels shall be different between the two CNN's. A channel is defined as a slice of an image with same resolution but only shows the intensity of one color. This can also be described as a two-dimensional array of values.

5.1.1 CNN that includes wind data

The requirements of the CNN were that it shall include both satellite images and wind data as an image with three channels. These channels shall have same resolution and consist of one satellite image channel, one land mask channel, and one wind speed data channel. The output shall be of equal pixel-size to the input image, but with only one channel. This channel shall display the probability of each class, oil spill and not oil spill.

5.1.2 CNN that does not include wind data

The requirements of the CNN shall only accept satellite images as images with two different channels. These channels shall have same resolution and consist of one satellite image channel and one land mask channel. The output is required to be of equal pixel-size to the input image, but with only one channel. This channel shall display the probability of each class, oil spill and not oil spill.

5.1.3 Commonalities

Both CNN's are required to be of same base type of CNN, including length and architecture for each input channel. Thus, whatever architecture is chosen, it must be the same for both CNN's.

5.2 Choice of base neural network

CNN's excel at processing grid-like structures and makes it possible for fully connected layers later to process the most important features.

Satellite images are already grid-like structures, which do not need major alterations before giving it to the CNN. In addition, CNN’s can accept variable input sizes. This means that the input images do not necessarily need to match in resolution.

Wind datapoints, however, is not granted to be in a grid-like structure but is likely to be of some pattern over the same area the satellite image covers. Thus, these datapoints need some form of transformation to achieve a guaranteed grid-like structure before the model will accept it.

5.3 Transformation of wind data

The wind data is originally presented as geographical points with a wind speed in zonal and meridional directions and absolute wind speed. Only absolute wind speed is used in this project. Wind direction is rather used in detecting sources of oil spills, such as oil platforms and ships, which is outside the scope of this project. Given that the geographical points already resemble a semi-structured grid when visualized in combination with the associated satellite image, they can be treated like an image with pixel values as absolute wind speed. Thus, each datapoint needs to be ordered in such a way that an image can be created and match spatially with the satellite image.

However, the wind datapoints are not the same resolution as the satellite image. Resolution of wind datapoints must be equal to resolution of satellite image to present them as separate channels of an image to the neural network. This is solved by upscaling the resolution of absolute wind speeds to match the dimensions of the satellite image. The upscaling technique used in this project is interpolation by either bilinear or bicubic algorithms. Both algorithms are tested to see if there is any significant time difference. If not, only the cubic algorithm will be used.

Wind data is converted after the interpolation to absolute values since wind speed is measured in only positive values.

There are areas of the image where wind speed data is misleading. This has the potential to happen over land areas such as islands and coasts. Thus some data about land masses are included as well.

5.4 Land mask

Every satellite image provided by KSAT supplies a land mask of potential land area the image covers. This land mask consists of binary values that tells where land is covered or not.

5.5 Tiles

The satellite images to be used in this project are too large for a CNN with the available GPU memory to handle at once. Thus, both satellite images with their associated wind data will be divided into tiles of a size the network can handle. Wind data tiles will still match in dimensions with associated satellite image tiles.

5.6 Neural network design

The design of the neural network that accepts wind datapoints accepts them as a separate channel of an image in combination with a satellite image channel and a land mask channel.

5.6.1 Channels

The satellite image channels consist of radio amplitudes of a polarization combination. These channels show the potential oils slicks.

The wind data channel consists of an interpolated image of absolute wind speeds in the same area the satellite image covers. This channel will verify the oils slicks shown in the satellite image channels.

5.6.2 Purpose of neural network

The neural network has the purpose of imitating the output of the manual work from TEOS. A simplified flowchart of the manual work at TEOS can be seen in figure 1. This flowchart is not exhaustive in factors deciding an oil detection but shows the most common approach without edge cases.

The simplified flowchart of manual TEOS work shows the usage of both SAR-image and wind data. The SAR image contains features that potentially can be classified as oil slicks. If any oil slick features are found, then wind speed is used to verify the detection. There is a lower and an upper threshold of wind speed where detections must be in between as a requirement to be classified as spilled oil slick.

Other sources like AIS-position of vessels are left out of this simplification and will not be used in this project as well. This is because they are more relevant to finding sources of any spilled oil slicks rather than confirming any suspicious oil slicks.

5.6.3 Dataflow

The data flow of the neural network is visualized in figure 2. Instead of using a filter, it is expected that the neural network will learn itself what wind speed is valid for a detected spilled oil slick. The lower and upper thresholds of wind speed are expected to be abstractions rather than actual filters in the neural network, which is expected to learn the thresholds in some way.

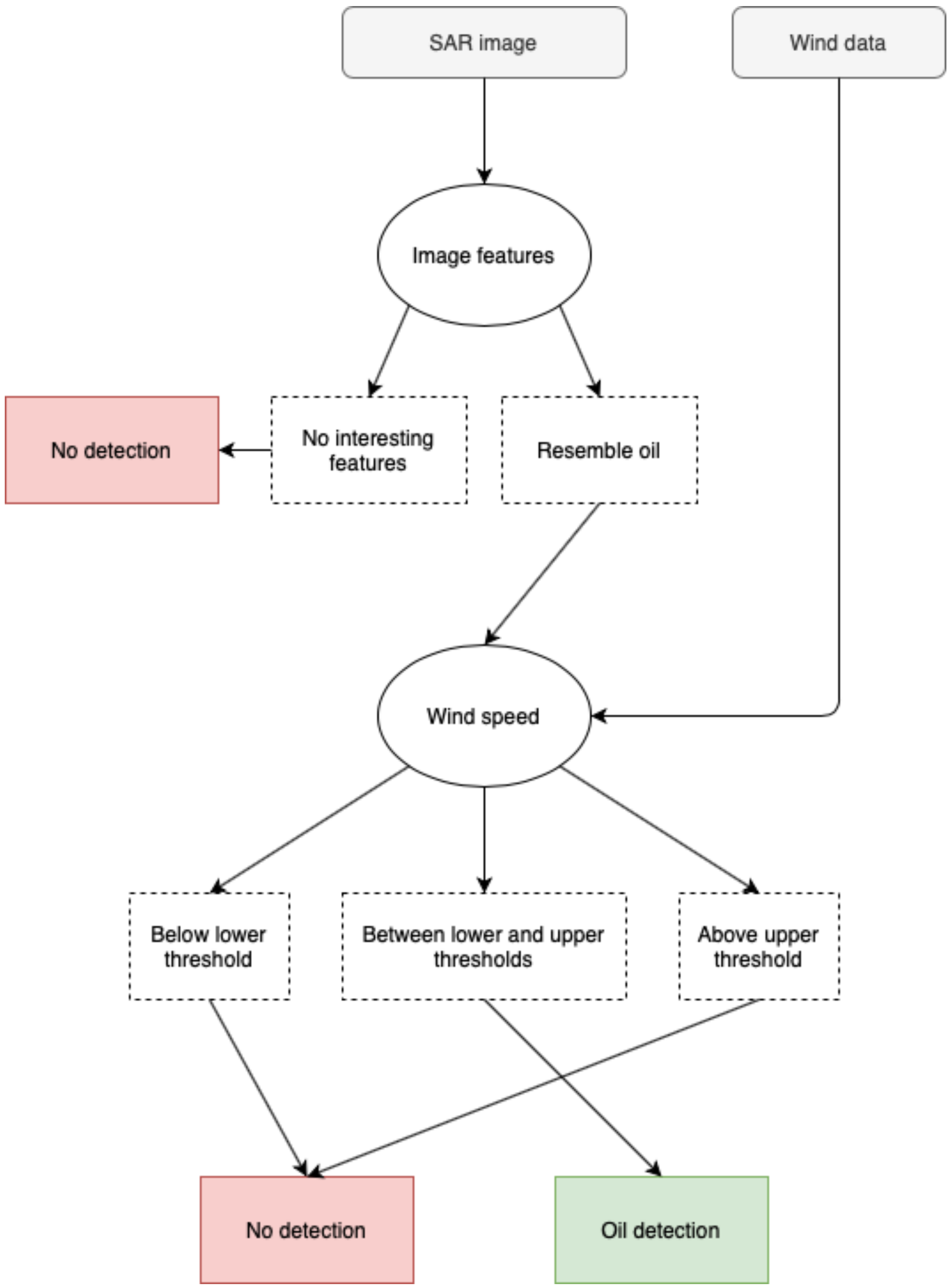

Figure 4: Simplified flowchart of manual TEOS work

Figure 4 shows the simplified flowchart of the manual work at TEOS. The manual work is roughly divided into two sequential tasks. The first one is to identify any oil spill resembling features. The second task is to verify these oil spill resembling features by studying the wind speed of that area.

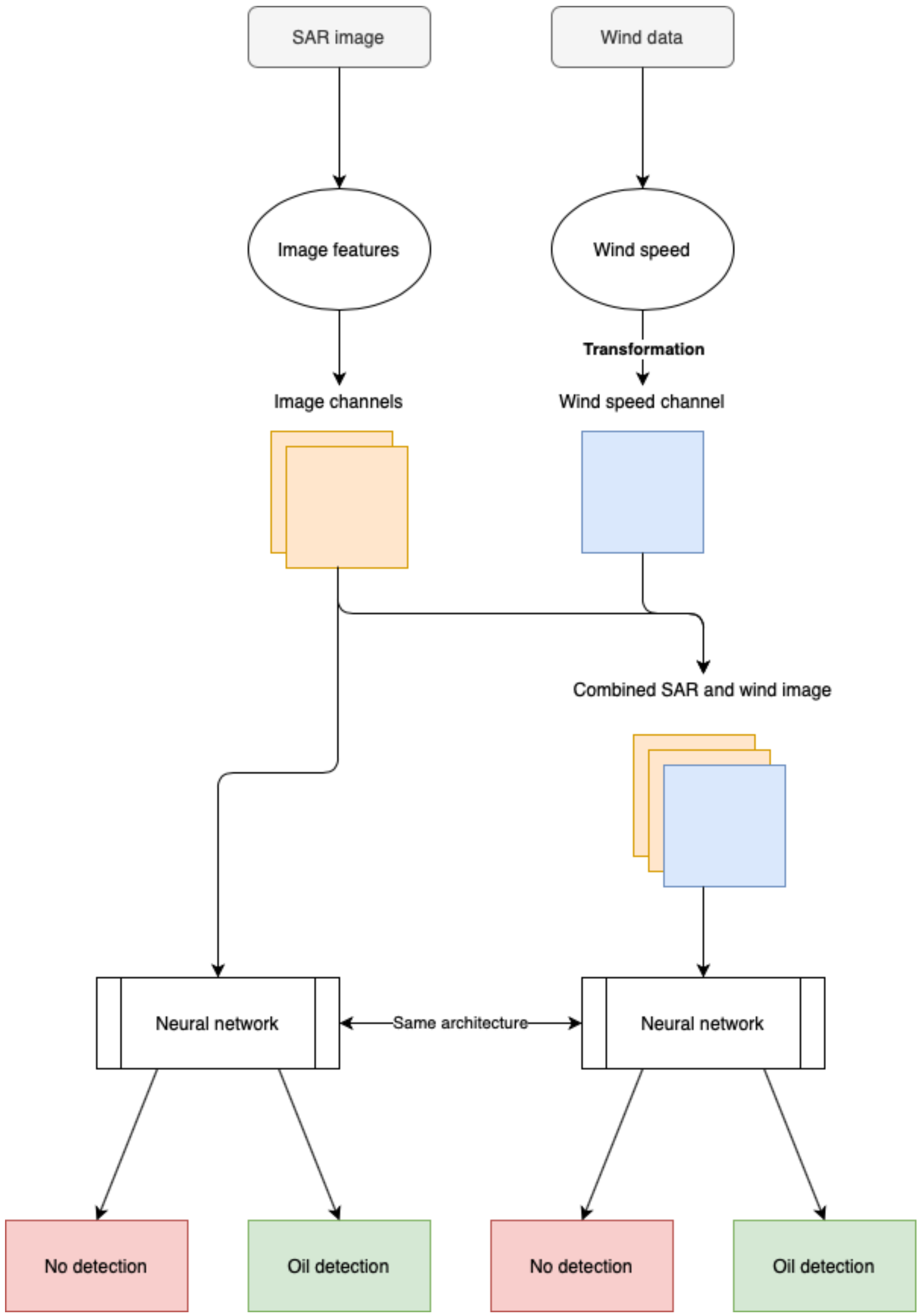

Figure 5: Neural network flowchart

Figure 5 shows the neural network flowchart, including both CNN's. Wind speed data is transformed into and image and thereafter combined with satellite images as inputs to one CNN, while only satellite images are given as input to the other CNN.

5.6.4 Components

The design of the neural network itself is a U-Net with the components of ResNet with 34 layers. The neural network classifies the satellite images with semantic segmentation, as illustrated in figure 2.

5.6.5 Model with only satellite imagery

The design of the neural network that has only satellite imagery as input will be like the other network. This is because its purpose is to remove the effect of wind data to the detection process. As seen in figure 5, the removal is done by removing the wind data channel from the input image such that the remaining input data of only satellite images will predict oil slicks. Thus, the only difference between the neural networks are the wind data as input. A comparison between the two CNN's will be a measurement of the difference in validation loss, visual interpretation of output images, and precision and recall.

6 Implementation details

6.1 Interpolation of wind data

The conversion and interpolation of semi-structured wind data into the structured data of an image is done by using SciPy.org interpolate methods. These methods are the LinearNDInterpolator, which is based on linear interpolation, and the CloughTocher2DInterpolator, which is based on cubic interpolation. These methods do both steps of converting unstructured data into structured data and interpolating missing data to get higher resolution.

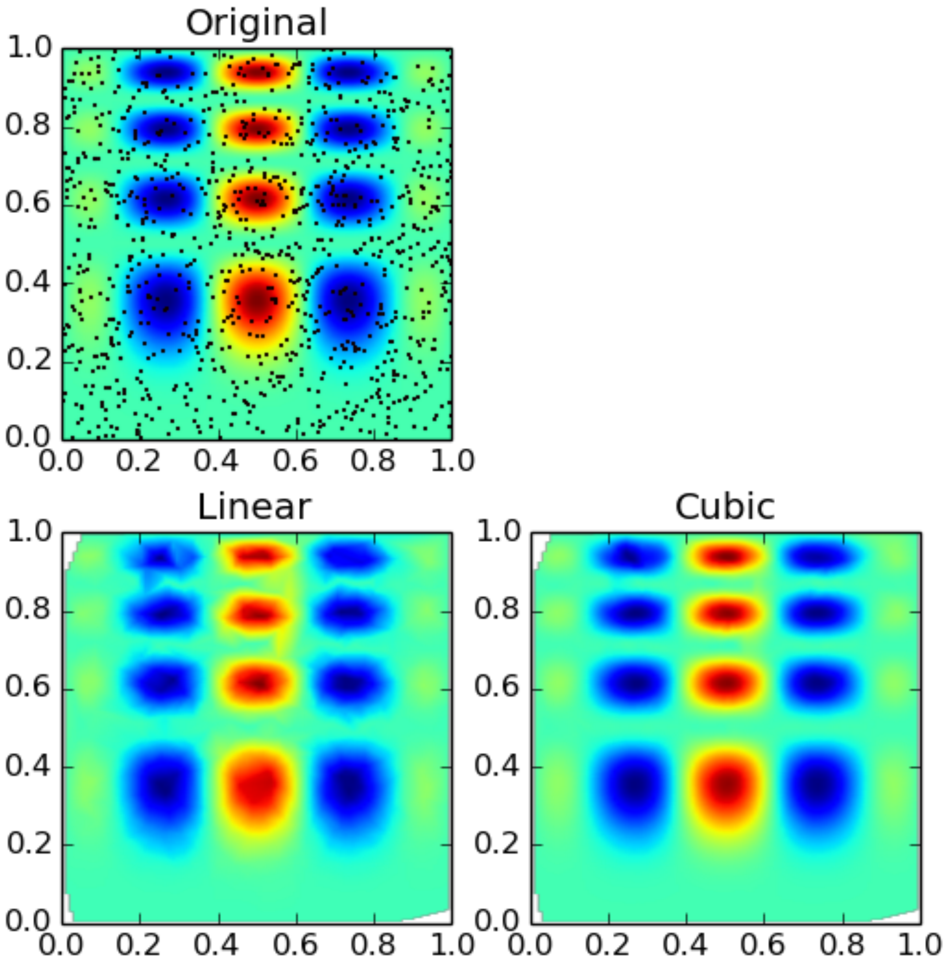

Figure 6: Examples of resulting structured data from linear and cubic interpolation methods compared to the original unstructured data points. This example is replicated from the original illustration at SciPy documentation page of interpolation [33].

6.2 Combining wind data and satellite image

The combination of wind data and satellite images are done through concatenating the relevant layers in the satellite image and wind speed image into one image with several layers. The framework used to create the CNN and the algorithms that feeds it data is called *fast.ai*. This framework has built-in functions for creating data loaders that feeds the network by providing it paths to the images and labels.

6.3 Data setup

The validation data is picked randomly with a ratio of 0.1 from all the data, where the rest will be training data. The validation data picked is replicated exactly with each neural network and training session by a given seed to the splitter. This is done to avoid accidentally picking easier validation data for one result compared to the others.

6.4 Neural networks

The neural networks are both built using a framework called *fast.ai*. The neural network is created by a combination of U-Net and ResNet architectures, where the U-Net contains the components of Res-Net with 34 layers.

7 Results and discussion

This chapter describes the results of this project. This includes the interpolation of wind data, training the neural networks and how the neural networks perform on a given dataset of oil slicks. In addition, this chapter compares and discuss the results.

The wind interpolation, probability maps and detections are tested by judging the visual output. The validation loss is tested by comparing validation losses from several training sessions against each other.

7.1 Wind interpolation

The wind interpolation worked as intended and the results resemble the associated satellite image of the area covered. As seen in figure 7, the image to the right is the resulting wind interpolation with cubic interpolation and SAR-wind, while the radar satellite image is to the left. The color bar at the right represents absolute wind speed. The visible land area is nullified since the data is irrelevant to oil spill detection and might confuse the neural network. The nullification area is retrieved from the provided land mask of the product, which is seen in the satellite image to the left in green color. Linear interpolation was tested and worked as well but was not used further in learning the neural networks since cubic sufficed.

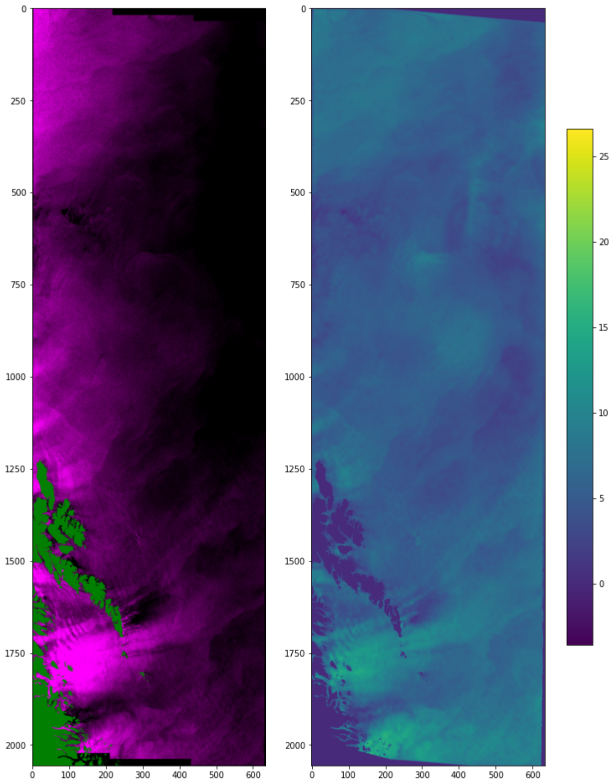

Figure 7: Example of wind interpolation

This result means that interpolation of a grid of geographical wind speed points, which were not perfectly lined, is possible and is useable to a CNN. This does not implicate that including wind data will improve the detection of oil spills. Further testing is done by comparing the two neural networks that are either fed with wind data or not.

7.2 Training and validation loss

Training the neural networks was done several times to achieve lowest possible validation loss. Both neural networks train and validate with the same data by using the same seed while splitting training and validation data. Both neural networks were trained three times and the trained network with the lowest loss is used in the results later described. Training three times was done to rule out random spike values that does not reflect the typical loss and get an understanding of how much such loss varies in these models. The lowest loss is chosen because the lowest loss implies a better performing model that is not penalized as much by the loss function. The better performing model is obviously chosen to do the given task, not the average performing model.

Validation loss table

Training iteration	Model with wind data	Model without wind data
1	240.98	288.03
2	251.04	333.57
3	260.22	283.00

Validation loss metrics table

Metrics of validation loss	Model with wind data	Model without wind data
Lowest	240.98	283.00
Mean	250.75	301.53
Standard deviation of population	9.623	27.868
Confidence interval of 95%	240 to 262	270 to 333

The results show a difference in validation loss between the neural networks. The neural network including wind data has a 14.85 % lower validation loss if choosing the lowest validation loss for both neural networks. This difference shows that including wind data in the neural network has a measurable effect on the resulting validation loss, which implies a plausible lowering of validation loss in neural networks of similar design when training with this kind of data.

In addition, the neural network including wind data has a 16,84 % lower validation loss if choosing the mean of validation losses for each neural network. This implies any low spike value did not create the difference between the lowest validation losses. However, the slightly higher difference in mean compared to lowest value can be attributed to the higher spike in the second training iteration of the model without wind data.

Furthermore, the results show the model without wind data has a 2.90 times higher standard deviation than the model with wind data. This is because the second iteration of the model without wind data show a spike value in validation loss that is 15.81 % higher than the second highest validation loss of all iterations. The other two iterations of the model without wind data do not stand out. The source for this spike is unknown.

7.3 Probability map from neural networks

The output from the neural networks consists of a probability map of each class in the semantic segmentation. Since there are only two classes, each class has a map that is the inverse probability of the other. Thus, the image results below will only show the probability of oil slicks.

The validation set of tiles that is used in these results is the same as the validation set in the training of both neural networks. The total amount of validation tiles is 100, where 10 of them have oil slicks in them.

The figures below show in this order from left to right the output images from both neural networks, the ground truth, the wind channel, and the SAR-channel. Land mask is not shown but can be seen indirectly from wind- and SAR-channels. Each row represents one tile in the validation set.

7.3.1 Ground truth

The ground truth image shows where the actual oil slicks lie according to the data from KSAT. The yellow color is oil slicks, and the purple color is background.

7.3.2 Wind channel

The wind channel has a brighter color with higher wind speeds. Darker colors are lower wind speeds. The range chosen for these images are from 0 to 10 meters per second.

7.3.3 SAR channel

The SAR-image shows how it is visible to the neural network. It is typical for darker patches to represent either low wind speed or oil slicks.

7.3.4 Output figures

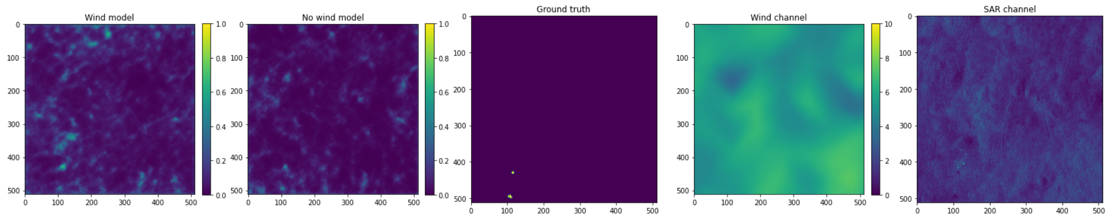

Figure 8: Tile 1

Figure 8 shows small differences in probability output from the two models. Both oil slicks are clearly visible in the output on both models but is not obvious unless you already know where to look. Other areas show similar probability of oil slicks which would probably result in some number of false positives. That said, the model that includes wind data has some areas that has a more probability of oil slicks than the model without wind data. This additional area of higher probability shows no clear pattern from wind channel. However, it shows a clear pattern from the SAR channel. Thus, it is assumed that the wind channel somehow triggered higher probability of oil slicks from some features in the SAR image. Wind channel is clearly biased by the SAR image.

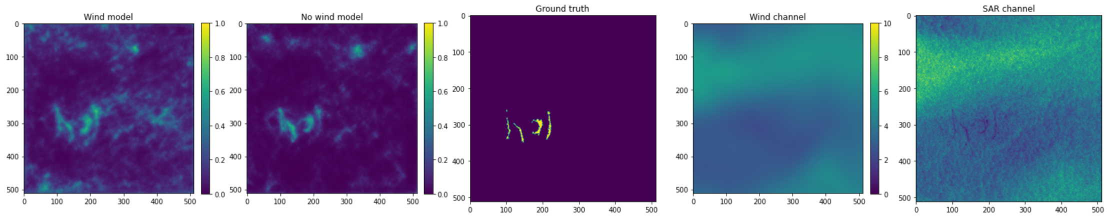

Figure 9: Tile 2

Figure 9 shows small differences in probability output from the two models. The oil slick group is clearly visible in the output from both models and does stand out from the rest of the image except for one small area in the top right of both output images. This small area would possibly trigger a false positive in both models. Otherwise, the model that includes wind data has some more area that show a slightly higher probability of oil slicks. This area has some correlation with the lower wind speed in the area, which is also clearly visible in the SAR image by difference in contrast. Wind channel is again clearly biased from SAR image.

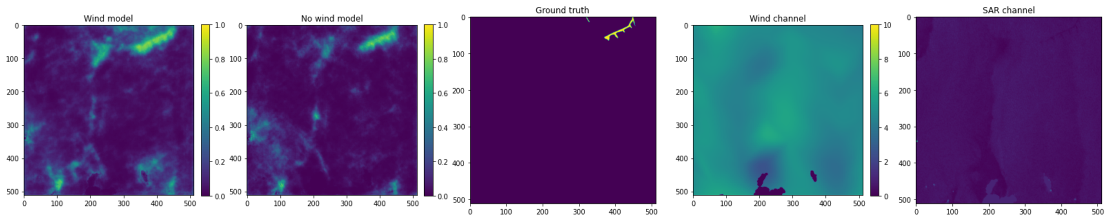

Figure 10: Tile 3

Figure 10 shows small differences in probability output from the two models. The oil slick is clearly visible in the output from both models and does stand out from most of the image. There are some additional spots that indicates oil slicks to the left of the actual oil slick and in the bottom left of the image in both outputs. The output from the model that includes wind data has a slightly higher probability for those spots. This would possibly result in some number of false positives in both outputs. The areas of higher probability in the model that includes wind data correlates to lower wind speeds in the wind channel but also features from the SAR image. There is still clearly biased from SAR image in the wind channel.

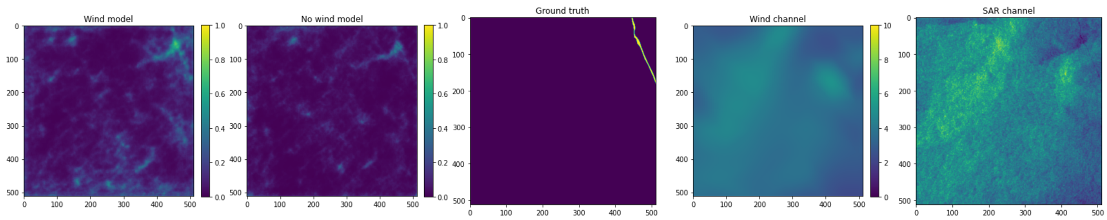

Figure 11: Tile 4

Figure 11 shows small differences in probability output from the two models. The oil slick is visible in the output from both models but is more clearly in the model that includes wind data. There are no other spots that indicates oil slicks in both models, but there are some areas of slightly higher probability in the model that includes wind data. These areas are not likely to trigger any false positives. There are no correlation between these areas with higher probability of oil slicks to the wind channel, but some correlation to the SAR image. Wind channel is again clearly biased from SAR image.

In addition, this tile might be a good example that it is not easy do determine dark spots as oil spills since other larger dark spots in the SAR image are more dominating.

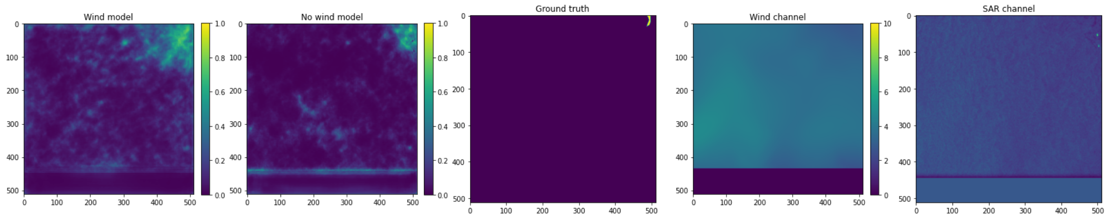

Figure 12: Tile 5

Figure 12 shows some difference in probability output from the two models. The oil slick is visible in the output from both models. The difference between the models is the area of high oil slick probability is larger on output from the model that includes wind data, and a higher probability of oil slicks in the bottom of the output from the model that does not include wind data. There are no other spots that indicate oil slicks. The area of higher probability in the output from the model that includes wind data correlates with the lower wind speed in the wind channel. Again, we see the correlation between the wind and SAR data.

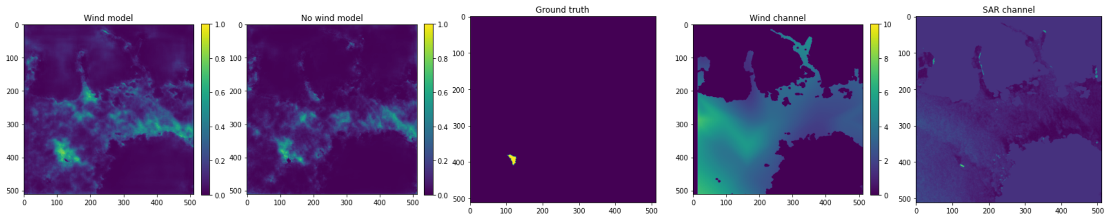

Figure 13: Tile 6

Figure 13 shows some difference in probability output from the two models. The oil slick is visible but is not the only high oil slick probability area of both outputs. False positives are likely in both outputs. The difference between the outputs is a slightly higher probability intensity around the oil slick in the output from the model that includes wind data. This area is correlated to lower wind speed in the wind channel. Wind channel is again clearly biased from SAR image.

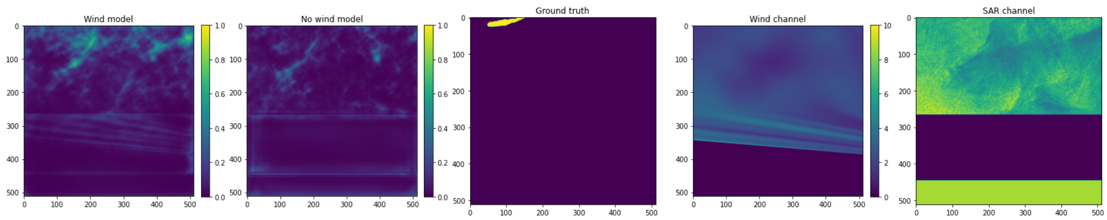

Figure 14: Tile 7

Figure 14 shows some difference in probability output from the two models. The oil slick is not visible in neither model. However, both models indicate false positives in a nearby area. The model that includes wind data has a slightly higher oil slick probability in this area, but false positives are likely in both models. The slightly higher oil slick probability area in the model that include wind data has only correlation with the SAR image. Wind channel is again clearly biased from SAR image.

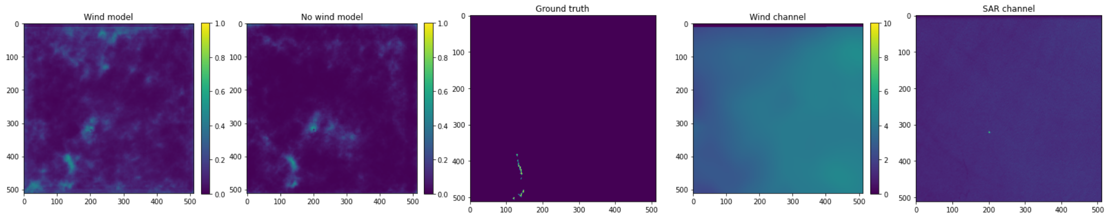

Figure 15: Tile 8

Figure 15 shows little difference in probability output from the two models. The oil slick is visible, but not the only high probability spot in the output from both models. The most notable spot is around the light dot in the SAR image, which is assumed to be some stationary object like an oil rig. False positives are likely. In addition, there are some spots that has a probability between 0.4 and 0.6 in the output from the model that includes wind data, but not likely to trigger false positives. These spots are correlated to lower wind speed in the wind channel. This tile also shows the wind channel is clearly biased from SAR image.

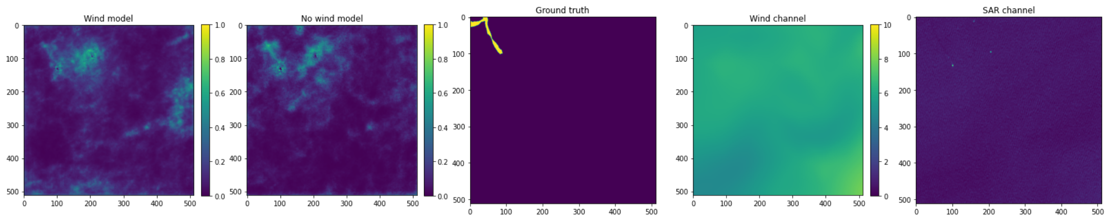

Figure 16: Tile 9

Figure 16 shows little difference in probability output from the two models. The oil slick is visible in both outputs but does not stand out from surrounding high probability area. These surrounding areas are most notable around the light dots in the SAR image, which are assumed to be some stationary objects like oil rigs. False positives are likely. In addition, there are some spots in the right middle side of the output from the model that includes wind data that has probability between 0.4 to 0.8. False positives are likely here. These spots are correlated to lower wind speed in the wind channel. Again, wind channel is clearly biased from SAR image.

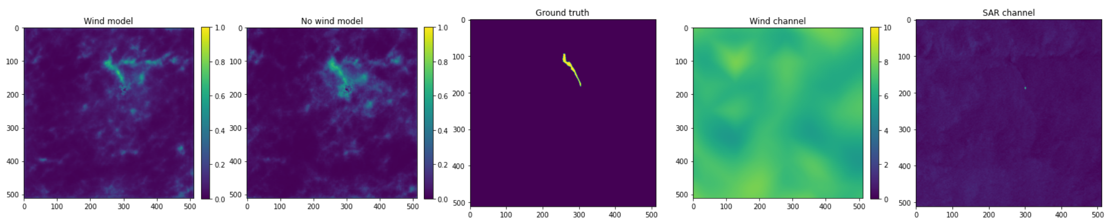

Figure 17: Tile 10

Figure 17 shows some difference in probability output from the two models. The oil slick is clearly visible in both outputs. No false positives are likely. The model that does not include wind data has a slightly larger area of high oil slick probability. This area is correlated with the SAR image, and especially around the light dot, which is likely a stationary object like an oil rig. Wind channel is clearly biased from SAR image.

In summary the model without wind seem to perform slightly better with less false positives. This might be because the wind-model favor hitting oil spills more than avoiding false positives and explain the contradiction from the validation loss which favored the model with wind data. However, these false positive areas are weak and may not trigger false positive detections from a detection algorithm that is based on some threshold.

Not all validation tiles are in favor of the model without wind data. Tiles 4 and 10 seem to favor the model that includes wind data by less false positive areas. The other tiles are by comparing false positive area either roughly equal or in favor of the model that does not include wind data.

The areas where the model that includes wind data has more false positives seem to have a correlation with low wind. Darker spots, which might indicate oil spill, and areas with lower wind speed are similar in the SAR image. This might explain some of the correlation. However, the correlation has yet to be verified in a potential future work where lower wind speeds are compared to higher probability of oil slicks by including wind data in the convolutional neutral network.

7.3.5 Output figures with detection algorithm

The following figures show the same output images with the addition of marked detections from a detection algorithm.

The detection algorithm used is created at KSAT. Detecting oil slicks is done by evaluating the output probabilities of the semantic segmentation neural network. The places with highest probabilities that is higher than a given threshold is classified as oil slicks. This threshold is tuned individually for each of the two neural networks such that they each have the lowest number of false negatives as possible. The lowest number of false negatives possible is two, and both seemed impossible to detect for both models.

False negatives are worse than false positives since the consequences of missing actual oil spills are higher than false alarms. Thus, it is necessary to adjust the detections threshold for both models such that each model detects as many oil spills as possible, where each model may have a different threshold.

The detections are marked with a red circle such that bodies of continuous slicks are not counted as several oil slicks. In addition, an attempt of region growing is done at the nearest oil like figure to the red circle of the detection. The result from this region growing is marked in black.

To simplify the detection process, some of the oil slicks are grouped together to represent a single oil slick or oil slick group. This is done because the importance of detecting an oil slick is more important than separating nearby oil slicks from each other and that nearby oil slicks have a higher probability of discovery once one of them are detected. This is a subjective exercise, which may be biased.

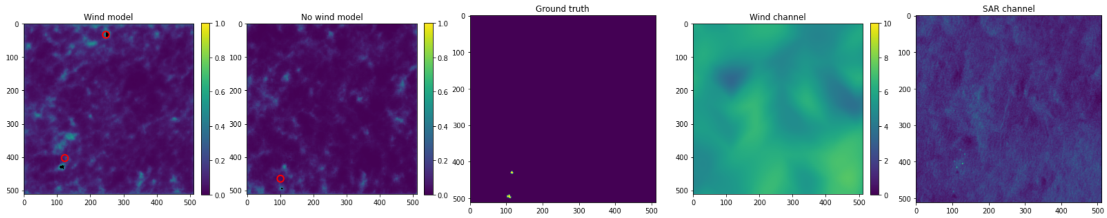

Figure 18: Tile 1

Figure 18 shows validation tile 1. This tile shows a group of oil slicks in the lower left corner. There are two slicks in this group, but since they are in proximity, they are defined as a single group. Both models detected the oil group but appear to detect different slicks in the group. This is a subjective judgement as the detections are not far away from each other. The wind model does detect a false positive as well in the upper middle part of the tile. Thus, the model without wind has a slightly better precision in this case. However, recall remains the same for both models.

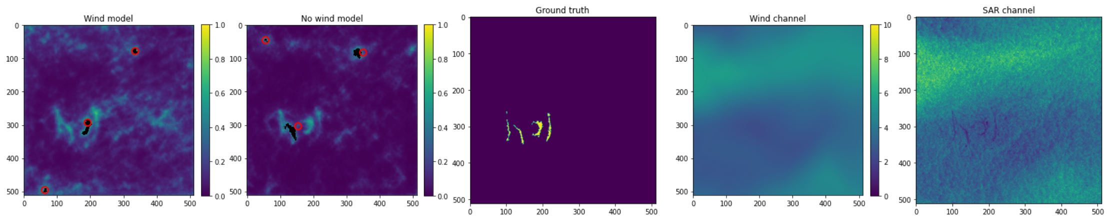

Figure 19: Tile 2

Figure 19 shows validation tile 2. This tile shows a group of oil slicks roughly in the center. There are four large bodies of oil slicks in this tile with proximity. Thus, they are defined as a single oil slick group. Both models detected the oil group but detects different slicks in the group. Both models detect two false positives as well. Precision and recall are the same for both models in this case.

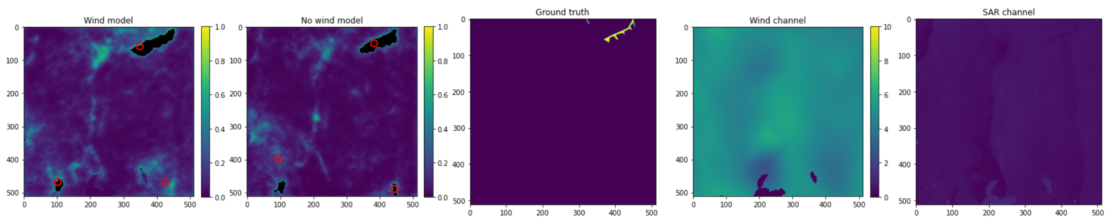

Figure 20: Tile 3

Figure 20 shows validation tile 3. This tile shows an oil slick group in the upper right corner. There are one large body of oil slick and a smaller body in proximity to it. Thus, they are defined as a single oil slick group. Both models detect this oil slick group, but both have two false positive as well. Precision and recall are the same for both models in this case.

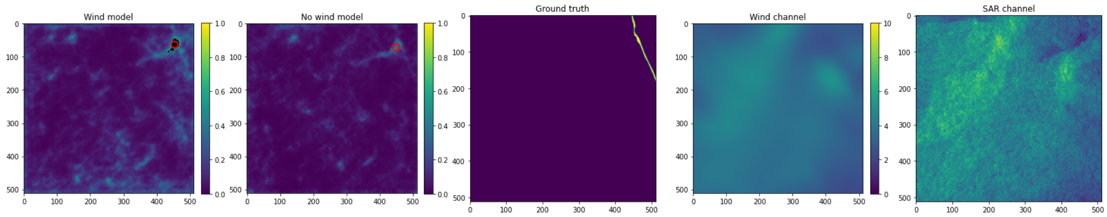

Figure 21: Tile 4

Figure 21 shows validation tile 4. This tile shows a single oil slick in the upper right corner. Both models detect this oil slick with no false positives. Precision and recall are the same for both models in this case.

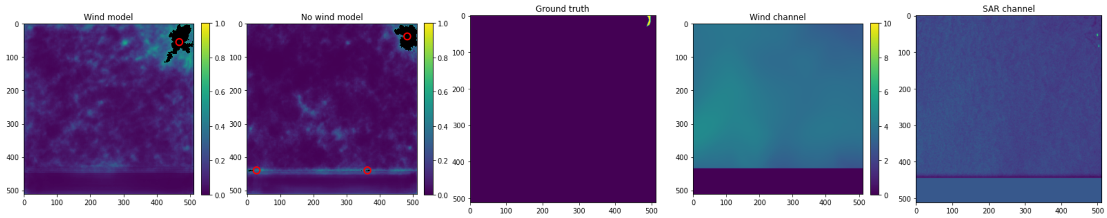

Figure 22: Tile 5

Figure 22 shows validation tile 5. This tile shows a single oil slick in the upper right corner. Both models detect this oil slick, but the model without wind detected two false positives. The wind model has better precision while recall is the same for both models in this case.

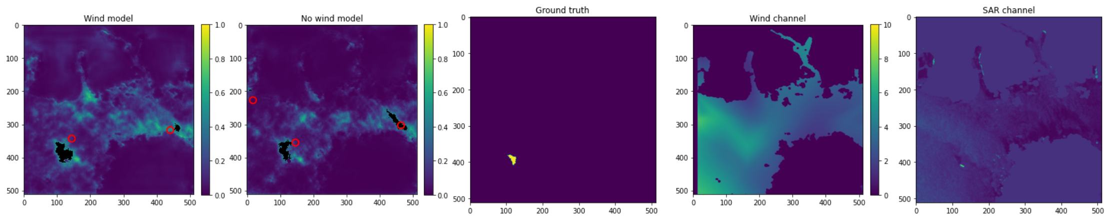

Figure 23: Tile 6

Figure 23 shows validation tile 6. This tile has a single oil slick in the lower left corner. Both models detect the oil slick. The wind model has a single false positive and the model without wind has two false positives. Thus, the wind model has better precision while recall is the same for both models in this case.

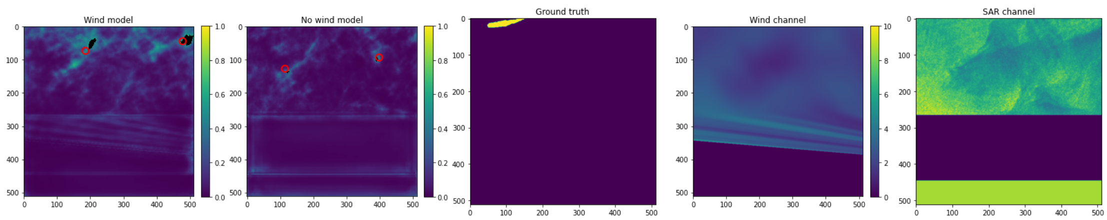

Figure 24: Tile 7

Figure 24 shows validation tile 7. This tile has a single oil slick in the upper left corner. Neither model detect the oil slick but detects two false positives instead. The nearest false positives from both models are considered too far from the actual oil slick to count as detections. Precision and recall remain the same for both models in this case.

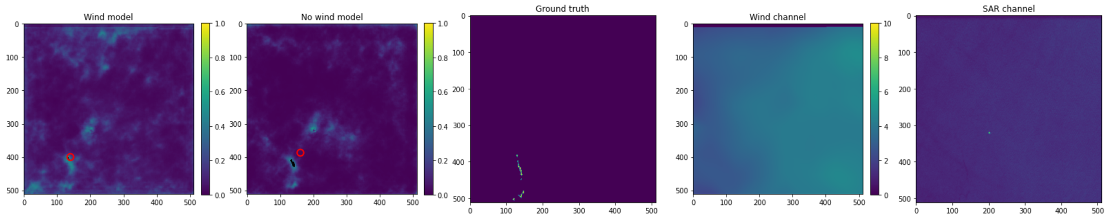

Figure 25: Tile 8

Figure 25 shows validation tile 8. This tile has an oil slick group in the lower left corner. Both models detected this oil slick group. Precision and recall are the same for both models in this case.

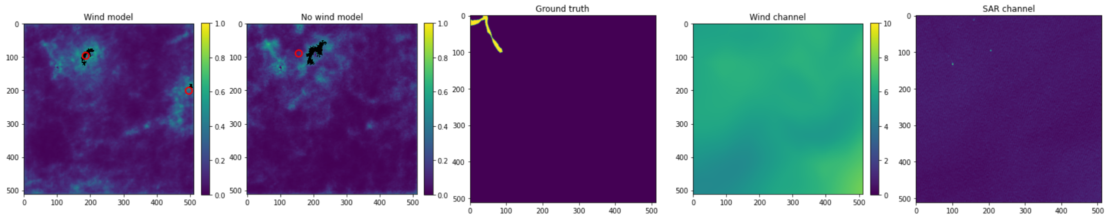

Figure 26: Tile 9

Figure 26 shows validation tile 9. This tile has a single oil slick in the upper left corner. Neither model detects this oil slick. The wind model detects two false positives and the model without wind detects a single false positive. The nearest false detections from both models are not close enough to the actual oil slick to count as detections. The model without wind has a higher precision while recall is the same for both models in this case.

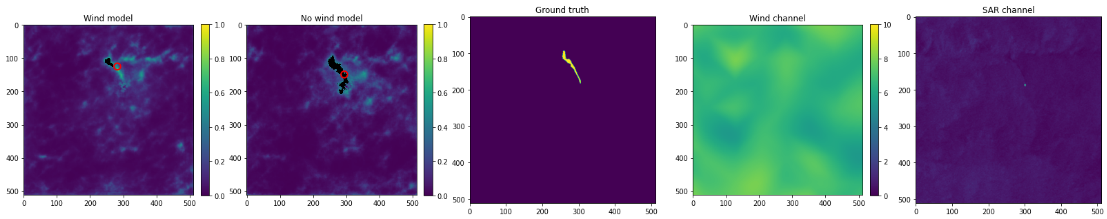

Figure 27: Tile 10

Figure 27 shows validation tile 10. This tile has a single oil slick in the center. Both models detect the oil slick with no false positives. Precision and recall are the same for both models in this case.

In summary the validation tiles are of varying challenge to the oil detectors, where the challenging tiles has oil slicks very close to the edges of the tile or are not easily visible on the SAR channel. Since these are randomly picked from a dataset of oil slicks, the varying challenge was expected and shows what an oil detector might expect. Both detectors seem to struggle with the same tiles but managed to detect most given oil slicks or oil slick groups. However, both had a high number of false positives, which was also expected since the number of false negatives was tuned to be as low as possible for both detectors. The resulting statistics from these detections are described and discussed in chapter 7.4.

7.4 Detected oil slicks

The detection results from both neural networks are described in the tables below.

7.4.1 Neural network with wind channel

Tile nr.	Correctly detected	Falsely detected	False negatives	Oil slicks or oil slick groups
1	1	1	0	1
2	1	2	0	1
3	1	2	0	1
4	1	0	0	1
5	1	0	0	1
6	1	1	0	1
7	0	2	1	1
8	1	0	0	1
9	0	2	1	1
10	1	0	0	1
Total	8	10	2	10

7.4.2 Neural network without wind channel

Tile nr.	Correctly detected	Falsely detected	False negatives	Oil slicks or oil slick groups
1	1	0	0	1
2	1	2	0	1
3	1	2	0	1
4	1	0	0	1
5	1	2	0	1
6	1	2	0	1
7	0	2	1	1
8	1	0	0	1
9	0	1	1	1
10	1	0	0	1
Total	8	11	2	10

As seen in tables 7.4.1, neural network *with* wind channel, and 7.4.2, neural network *without* wind channel, the total number of correctly detected and false negatives of oil slicks are the same. However, falsely detected oil slicks differ by one, as neural network without wind channel has one more falsely detected.

Correctly detected are identical between the two tables. Both neural networks managed to detect oil slicks in validation tiles 1, 2, 3, 4, 5, 6, 8 and 10. Both neural networks did not detect oil slicks in validation tiles 7 and 9.

Validation tiles 7 and 9 stand out from the other validation tiles for several reasons. One common feature of tiles 7 and 9 are that the oil slicks are very close to the edge of the tile. However, this is not unique to tiles 7 and 9 as other tiles such as 1, 3, 4 and 8 has oil slicks or oil slick groups near the edge as well. Another common feature between validation tiles 7 and 9 is that they both have falsely detected oil slicks near the actual oil slicks.

7.5 Precision, Recall and F-measure

The precision, recall and F-measure are calculated form the results described in chapter 7.4.1 and 7.4.2. Precision, recall and F-measure are defined in chapter 3.5.

7.5.1 Neural network with wind channel

Precision	Recall	F-measure
44.4%	66.7%	57.1%

7.5.2 Neural network without wind channel

Precision	Recall	F-measure
42.1%	66.7%	55.2%

The neural network with included wind data channel has a slightly higher precision than the neural network without wind data. There is no difference in recall, which was by design expected when the threshold for detection was adjusted to achieve the least number of false negatives for both neural networks.

One should clarify that these results only represent this specific dataset which these neural networks has trained from, and this specific validation set which these neural networks has validated from. By changing the dataset and the validation set of that dataset the outcome could be different. However, since the dataset and validation set are the same for both neural networks and the best (lowest) loss of three attempts from training is chosen for both, the outcome from both neural networks can be compared against each other.

7.6 Result comparison and discussion

The validation loss is the main result to evaluate since it is a direct comparison of raw output between the two models. This is because results from detections and visual interpretations of the probability maps are a subject to bias. The difference of 14.85 % in lowest validation loss and a difference of 16,84 % in mean validation loss is measurable and significant. However, the population size of 3 with each model is small, even though the confidence intervals of 95 % are not overlapping. Further testing is necessary to get any results that are certain of a lower validation loss in models that use wind speed in a CNN that has the task of detecting oil slicks. While the result implies a likely lower validation loss, it does not mean an implied likely increase in precision in detection. The validation loss only measures the penalty of wrongly classifying oil or background from a given loss function, which is focal loss in this instance. That loss can be distributed in a variety of ways in the probability images of oil slicks. When a detection algorithm is dependent on a threshold of the probability of an oil slick, the distribution of that probability might not benefit the detection algorithm even though it gives a lower computed loss.

The difference of 2 percent points in precision and f-measure is not large enough to predict a similar difference to a larger training data set than the small population size of 100 tiles of oil slicks. The difference of precision and f-measure was a result of only one false detection less by the wind model. False negatives and correctly detected was the same for both models. This small difference in precision and f-measure is not enough to know for certain that a wind data channel improves precision to oil detection. However, the number of detections does not show this model performs worse either.

The output figures of the validation tiles favor the model that does not include wind data. A slight increase in area of higher false positive probability is seen in most validation tiles from the model that includes wind data. This observation contradicts the results from validation loss and detection since they favor the model that includes wind data slightly. Also worth mentioning is that the loss function used in this project is not as strict with false positives, but rather focuses on false negatives. However, the interpretation of the output figures is subject to bias.

Overall validation loss and detection results favor the model that includes wind data while output figures favor the model that does not include wind data. Given that the validation loss is not subject to bias from interpretation, the hypothesis of an increase in validation loss of 15 % in models that include wind data is likely. However, a hypothesis of increase in precision and recall is inconclusive and at best still plausible given the low size of the dataset. The increase in area of higher false positive probability does not give high promise to a CNN that includes wind data specifically from SAR-wind. Thus, including wind data does not benefit oil detection in an obvious way since a lower validation loss does not imply anything other than itself in future experiments. The promise of a CNN that includes wind data from MET-wind is still unknown as it might show completely different results. Suspected changes to results are that MET-wind might give the model that includes wind data an advantage. This is because MET-wind should in theory be free of bias from the SAR image.

8 Evaluation of the project

This chapter presents evaluation of the work done in this project, including contributions and constructive feedback. The work done in this project includes planning, research, design, implementation, and testing. In addition, this chapter presents evaluation of the proof of concept and evaluation of this project by staff from KSAT.

8.1 Evaluation of the combination of Prototyping and Agile methods

8.1.1 Planning and research

The planning phase of this project was mostly done in the first half year, where decisions regarding approach of the problem and what previous work is used as a foundation for further development. Core design choices were planned and researched here, like how to transform wind data into an image and what type of artificial neural network is preferred. This phase became longer than expected due to restricted access to the offices because of the covid-19 pandemic. Access to workspace at the offices was required to handle internal data at KSAT. Thus, further design and implementation were delayed significantly.

8.1.2 Design

The design phase of this project was mostly after the planning and research phase. However, design choices were made during almost all phases due to the nature of unexpected problems and re-implementations. This phase had mostly design choices based on the knowledge of the codebase from KSAT, where several key elements of data handling and processing were borrowed and built upon.

8.1.3 Implementation

The implementation phase of this project was only started at the second half of the project due to missing workspace, and thus missing access to data and codebase to build upon. This phase had the implementation of data handling, data transformation and building the neural networks.

8.1.4 Testing

The testing phase of this project was the phase that came after the implementation phase. This phase had the testing of the neural networks, but also testing the transformation of the wind data. The testing of transformation of wind data was completed during the implementation phase since both implementing the neural networks and testing them depended on the results of it.

8.1.5 Contributions

The contributions of this work are showing how a CNN can handle satellite data and wind speed data while detecting oil slicks. The proof of concept was successfully implemented and compared against a CNN without wind speed data. This comparison gave some measurements that are valuable to predict any benefits of implementing this concept into production of an oil detection service at KSAT.

8.1.6 Critical reflection

The constructive feedback of this work are some issues that have potential of improvement. These are an earlier start of the design and implementation phases such that problems and necessary re-designs are detected earlier, and more cooperation with KSAT to avoid long problem-solving issues related to their codebase. Both were some consequences of the ongoing covid-19 pandemic because of missing workspace and lower than normal direct contact with employees at KSAT. It is not certain that these issues could have been avoided, and some of them resulted in a delay of this project. That said, some preparedness for available workspace and more cooperation could have been considered if this project is to be replicated.

8.1.7 Goals

Given the approved delay of this project's deadline, all three goals were achieved.

- An algorithm that transforms wind data into an input that can be accepted by the CNN was achieved.
- A CNN that can detect oil spills while ignoring look-alikes to some degree was successful with mixed results.
- A proof of concept of automated oil spill detection with wind data along satellite images as inputs to a CNN was achieved as well.

8.2 Proof of concept

The proof of concept for this project is about automated spill detection with wind data along satellite images as inputs to a CNN. This proof of concept has been achieved since the results from this CNN is able to detect some oil spills. However, the concept did not detect more oil spills than the other CNN without wind data. The comparison between the two CNN's had mixed results and it is not clear if this concept should replace current methods of oil spill detection.

8.3 Evaluation of project by KSAT staff

The evaluation of this project by KSAT staff is done by the head of the machine learning group that works with automated ship and oil detections. His evaluation is that this project has made significant contributions, showing how a CNN with both satellite images and auxiliary data such as wind performs against a similar CNN with only satellite images when detecting oil spills. In addition, this project provides valuable knowledge and lessons learned, guiding the further development of KSAT's automated oil spill detection algorithm.

9 Next step for the CNN

The next step for the CNN of this project concerns topics such as testing with more data, MET-wind, linear interpolation and a potential correlation between low wind speed and higher false positive probability. MET-wind and linear interpolation were not used in this project and has a potential of additional research. The potential correlation has yet to be verified.

9.1 Testing with more data

The results from this project can be explored further by including more data in training of these convolutional neural networks. This includes more oil spills, satellite images and wind data. This will also give a larger pool of validation data to validate the results from.

9.2 MET-wind

Since only SAR-wind was used in this project, a future work is testing the same neural networks with MET-wind. This will include transforming the MET-wind by interpolation to same resolutions SAR-wind was interpolated to even though MET-wind has a lower original resolution than SAR-wind. It is unknown if the interpolation will be a success with this MET-wind resolution.

9.3 Linear interpolation

For this project, the processing time has not been of significance. Thus, linear interpolation with its theoretically faster interpolation was not needed. However, should an oil slick-detection algorithm benefit from shorter processing times, testing this interpolation algorithm would be beneficial as a future work.

9.4 Correlation between lower wind speed and higher probability of oil slicks

From the resulting tiles in chapter 0, there seem to be a correlation between a higher false positive probability of detecting oil slicks where the wind speed is low. This has yet to be verified by calculating a correlation between the samples of pixels that show higher oil slick probability and pixels that show low wind speed in the wind speed layer. The choice between including or excluding actual oil slick pixels with this correlation is yet to be discussed, as it is false positives that is of concern to a model that includes wind data.

10 Conclusions

This thesis presents the work of implementing a CNN with satellite images and wind data for detecting oil slicks.

The main problem of detecting oil spills is avoiding classifying look-alikes as oil spills. This problem is addressed by including wind data in the CNN. Wind speed in the area monitored is important information to know the difference between actual oil spills and look-alikes. Therefore, the potential benefits from including wind speed data are an increase in precision and recall of detecting oil slicks. This improvement in precision and recall is measured by implementing a similar CNN that only accepts satellite images and compare the two networks against each other.

10.1 Discussion

The results from comparing precision and recall between the two neural networks, where one of them includes wind speed data and the other does not, only show a slight benefit of using wind data. However, the validation loss from training the models show around a 15 % decrease in the CNN that includes wind data. The resulting detection measurements using an algorithm at KSAT is showing an increase of 2 percent points in precisions when using wind speed data because of the fewer false alarms. The resulting probability map of oil slicks does not show an obvious improvement by including wind speed data but may worsen the potential falsely detected oil slick number. Given the conflicting results and use of wind data that is based on the satellite image, the conclusion is that including wind speed data from SAR-wind in a CNN in detecting oil slicks appear to be possibly beneficial. This can be explored further with a larger dataset of oil spills, satellite images and wind data. In summary, the results from including wind data in a CNN for oil spill detection are promising.

10.2 Future work

The potential benefits of using MET-wind are still not known and is encouraged to be explored in a future work. This includes replacing SAR-wind with MET-wind and do similar measurements in precision, recall and other interpretations. Other future work includes linear interpolation instead of cubic interpolation that may decrease interpolation time, and study a possible correlation between lower wind speed and higher probability of detecting oil slicks from the CNN in this project.

11 Bibliography

[1] P. F. Kingston, "Long-term Environmental Impact of Oil Spills," *Spill Science & Technology Bulletin,* pp. 53-61, 1 June 2002.

[2] NASA, "What is Synthetic Aperture Radar?," Earth Science Data Systems, 16 April 2020. [Online]. Available: https://earthdata.nasa.gov/learn/backgrounders/what-is-sar. [Accessed 27 Januray 2021].

[3] E. P. Myers and C. G. Gunnerson, "Hydrocarbons in the Ocean," Maritime Administration; NOAA Environmental Research Laboratories, Boulder, Colorado, 1976.

[4] H. A. Hovland, J. A. Johannessen and G. Digranes, "Slick detection in SAR images," in *IEEE - IGARSS*, Pasadena, CA, US, 1994.

[5] Kongsberg Satellite Services AS, "KSAT - About Us," KSAT, 1 January 2020. [Online]. Available: https://www.ksat.no/about-us/. [Accessed 9 Desember 2020].

[6] I. Goodfellow, Y. Bengio and A. Courville, "Capacity, Overfitting and Underfitting," in *Deep Learning*, Cambridge, Massachusetts 02142, The MIT Press, 2016, pp. 107-113.

[7] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis and I. Kompatsiaris, "Oil Spill Identification from Satellite Images Using Deep Neural Networks," *Oil Spill Remote Sensing,* pp. 1-22, 26 July 2019.

[8] I. Goodfellow, Y. Bengio and A. Courville, "Convolutional Networks and the History of Deep Learning," in *Deep Learning*, Cambridge, Massachusetts 02142, The MIT Press, 2016, pp. 359-361.

[9] A. Håkansson, "Portal of Research Methods and Methodologies for Research Projects and Degree Projects," *WORLDCOMP'13, FECS'13,* pp. 1-7, 22-25 July 2013.

[10] Effects Committee on Oil in the Sea: Inputs, Fates and National Research Council, "Input of Oil to the Sea," in *Oil in the Sea III: Inputs, Fates, and Effects*, Washington, National Academies Press, 2003, pp. 65-88.

[11] H. Espedal, "Detection of oil spill and natural film in the marine environment by spaceborne SAR," *IEEE International Geoscience and Remote Sensing Symposium,* p. 1478–1480, 28 June 1999.

[12] P. Trivero and W. Biamino, "Observing marine pollution with Synthetic Aperture Radar," *Geoscience and Remote Sensing New Achievements,* pp. 397-418, 1 February 2010.

[13] W. Alpers and H. Hühnerfuss, "The Damping of Ocean Waves by Surface Films: A New Look at an Old Problem," *Journal of Geophysical Research Atmospheres,* pp. 6251-6265, 15 May 1989.

[14] X.-M. Li and S. Lehner, "Algorithm for Sea Surface Wind Retrieval From TerraSAR-X and TanDEM-X Data," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING,* pp. 2928-2939, 1 May 2014.

[15] NORCE Norwegian Research Centre AS, "NORCE - Om oss," NORCE Norwegian Research Centre AS, 1 Januar 2021. [Online]. Available: https://www.norceresearch.no/om-oss. [Accessed 28 Juni 2021].

[16] NASA, "Polarimetry," 1 January 2021. [Online]. Available: https://nisar.jpl.nasa.gov/mission/get-to-know-sar/polarimetry/. [Accessed 25 January 2021].

[17] S. Angelliaume, P. C. Dubois-Femandez, E. C. Jones, B. Holt, B. Minchew, E. Amri and V. Miegebielle, "SAR Imagery for Detecting Sea Surface Slicks: Performance Assessment of Polarization-Dependent Parameters," in *IGARSS*, Valencia, Spain, 2018.

[18] T. N. Arnesen Hannevik, "Polarisation and mode combinations for ship detection using RADARSAT-2," in *IGARSS*, Honolulu, Hawaii, United States, 2010.

[19] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE CVPR,* pp. 770-778, 27-30 June 2016.

[20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Represenattions*, San Diego, California, United States, 2015.

[21] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* pp. 2481-2495, 10 November 2017.

[22] O. Ronneberger , P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015.

[23] M. Buda, A. Maki and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks,* no. 106, p. 249–259, 1 October 2018.

[24] J. Long, E. Shelhamer and T. Darell, "Fully Convolutional Networks for Semnatic Segmentation," in *Conference on Computer Vision and Pattern Recognition*, Boston, Massachusetts, United States, 2015.

[25] T.-Y. Lin, P. Goyal, R. Girshick and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE TRansatcions on Pattern Analysis and Machine Intelligence,* pp. 318-327, 7 February 2020.

[26] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Interpolation in Two or More Dimensions," in *Numerical recipes in C: the art of scientific computing*, New York City, New York, United States, Cambridge University Press, 1992, pp. 123-128.

[27] Cmglee, "Comparison of 1D and 2D interpolation," Wikimedia, 3 June 2020. [Online]. Available: https://commons.wikimedia.org/wiki/File:Comparison_of_1D_and_2D_interpolation.svg. [Accessed 9 September 2021].

[28] F. Girard-Ardhuin, G. Mercier, F. Collard and R. Garello, "Operational Oil-Slick Characterization by SAR Imagery and Synergistic Data," in *IEEE Journal of Oceanic Engineering*, Online, United States, 2005.

[29] D. Cantorna, C. Dafonte, A. Iglesias and B. Arcay, *Oil spill segmentation in SAR images using convolutional neural networks. A comparative analysis with clustering and logistic regression algorithms,* A Coruña, Galicia, Spain: Elsevier, 2019.

[30] I. Sommerville, "The waterfall model," in *Software Engineering Third Edition*, London, Pearson, 2016, pp. 47-49.

[31] I. Sommerville, "Agile Software Development," in *Software Engineering Third Edition*, London, Pearson, 2016, pp. 72-100.

[32] I. Sommerville, "Test-driven development," in *Software Engineering Third Edition*, London, Pearson, 2016, pp. 242-245.

[33] SciPy, "SciPy.org," 18 April 2021. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.griddata.html#scipy.interpolate.griddata. [Accessed 9 September 2021].

[34] F. Milletari, N. Navab and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *Fourth Interantional Conference on 3D Vision*, Stanford, California, United States, 2016.