Operationalizing Ship Detection Using Deep Learning

Torgeir Brenna, b, Lars-Petter Gjøvik, Gunnar Rasmussen, Tony Bauna, Michael Kampffmeyer, Robert Jenssen, and Stian Anfinsen

aKongsberg Satellite Services, Prestvannveien 38, Tromsø, Norway; bUiT The Arctic University of Norway, Hansine Hansens veg 18, Tromsø, Norway

ABSTRACT

In this work, we showcase use of an extensive radar earth observation image dataset based on a long history of vessel detection services, with the goal of operationalizing a next generation service using deep learning. The extent of the data represents an opportunity for unprecedented performance, but the data was not gathered with deep learning in mind, and challenges of weak labels and extremely unbalanced classes must be overcome. We achieve this through a mix of techniques, including a novel approach within the framework of weakly supervised learning. Additionally, we recognize that the image augmentations which have become ubiquitous in computer vision is often not physically valid in the radar domain. We ask whether such augmentations can still be pragmatically valid, i.e. if they can be justified by increased model performance.

Keywords: Deep learning, vessel detection, weakly supervised learning, extremely unbalanced dataset, image augmentation, earth observation, synthetic aperture radar

1. BACKGROUND AND MOTIVATION

Kongsberg Satellite Services (KSAT) has an extensive history of delivering near real-time (NRT) vessel detection services based on synthetic aperture radar (SAR) earth observation (EO) images. This 24/7 service is centered on human verification of automatically detected vessel candidates. Due to the SmallSat revolution, we anticipate an exponential increase in available data in the near future. It is not feasible to increase staffing proportionally, i.e. the automated part of the processing chain must be improved to detect a bigger proportion of the vessels while raising fewer false alarms.

2. DATASET CHALLENGES

These historical vessel detections are recorded as point labels (single-pixel), and they are not necessarily perfectly centered at the ships center of mass. A typical SAR image can easily contain > 10^9 pixels, and in our example dataset there are on average ∼ 10 vessels per image. With single pixel labels, the class imbalance on the pixel level becomes extremely high, in the order of 10^8. We view this as two separate challenges: How to automatically account for this weak labeling, and how to adapt to such extreme class imbalance.

Further author information: (Send correspondence to T.Br.)
T.Br.: E-mail: torgeir.brenn@ksat.no
L.P.G.: E-mail: larspg@ksat.no
G.R.: E-mail: gunnar.rasmussen@ksat.no
T.Ba.: E-mail: tonyb@ksat.no
M.K.: E-mail: michael.c.kampffmeyer@uit.no
R.J.: E-mail: robert.jenssen@uit.no
S.A.: E-mail: stian.normann.anfinsen@uit.no

*The advantages of SAR for surveillance and situational awareness purposes include invariance to clouds and sunlight, giving a reliable all-weather all-night capability.
2.1 Dealing with Weak Labels

Weakly supervised learning (WSL) is a sub-field of deep learning which is gaining momentum and receiving increased attention.\(^1\) In general, WSL is used to train models on data where the labeling is of a lower level than what would be ideal - for instance an image-level notation (“there is at least one car in this image”) while asking for bounding boxes of all cars in that image. The promise of WSL is to reduce the manual effort required compared to the size of the dataset. KSAT is in the quite typical situation that such techniques are crucial in order to leverage our historical dataset for deep learning.

Figure 1 demonstrates our novel WSL strategy. Originally, the dataset contains only single-pixel labels (red), and often there’s an offset from the center of mass due to time pressure in the service delivery. From an operational standpoint, this is a perfectly fine delivery; it is usually not interesting to get a segmentation mask, when it is the vessels position which is of interest.\(^1\) These labels are grown into small circular regions which are intended to match small vessels, and as such depends on the satellite and image type used. To account for larger ships, we also use a bigger region in training, where all non-vessel pixels are ignored by zeroing their loss. In other words, we require that all pixels within the small circle be classified as vessel pixels, and all pixels outside the big circles be classified as background, but the network is free to designate the pixels between the circles as either vessel or background, since we admit to not knowing the vessel size due to our weak labels.

2.2 Counteracting the Extreme Data Imbalance

Recall that the original dataset has around \(10^8\) times more background pixels than vessel pixels. To overcome this, we employ several tools:

- Label growing as described above. This reduces the background-to-vessel ratio by a few orders of magnitude.
- The weighted focal loss.\(^5\)
- The discarding of a proportion of tiles containing no background.
- Augmenting only those tiles where a vessel is present.

\(^1\)KSAT also estimates vessel size and heading, but for relatively low resolution SAR this is often difficult and inaccurate. Estimating these quantities and also vessel type is future work w.r.t. deep learning at KSAT.
Figure 2. The physics of image augmentation: Human perspective images can be zoomed/scaled, flipped horizontally and augmented with additive Gaussian noise, while still being physically valid. Optical EO images can also be rotated and flipped vertically, but what about SAR images?

E.g., a model trained for 20 epochs will have seen augmentations of each detection 20 times, but each background tile perhaps only once on average. Additionally, the weighted focal loss focuses the training on the hard examples (typically vessels and vessel look-alikes), also giving the loss much more weight when the true class is "vessel". This combination of approaches allows us to train the model outlined in section 4.

3. SAR IMAGE AUGMENTATION

Image augmentation is a popular way to inflate a dataset by transforming the training images in different ways each time they are presented to the neural network. A question which, to our knowledge, has yet to be properly addressed is the physics of SAR image augmentation, and how this relates to model performance. As figure 2 demonstrates, it does not make sense to vertically flip an image taken horizontally (unless of course we are expecting upside-down cats in our test data). For optical EO images the flexibility is increased, since accurate corrections for incidence angle means that any flip or rotate operation corresponds to the same operation on either the sensor or the scene, both of which are physically valid.

Radar on the other hand, works in a fundamentally different way, as it is an active rather than passive sensor. What this means is that a rotation does not correspond to either a rotation of the sensor or the scene, as the axes have distinct physical interpretations. Flipping the image in the cross-range direction represents opposite satellite flight direction and as such is physically valid, while the other flip is dubious at best due to the complexities of SAR range corrections. Even zooming is problematic, as the active nature of the SAR sensor yields a "noisy texture" called speckle, meaning that zooming is not equivalent to changing the distance between sensor and scene. This same property also excludes Gaussian additive noise, although other noise schemes might be valid.

Nonetheless, these physics observations do not automatically disqualify SAR image augmentations for training operational deep learning models, we must be pragmatic enough to do whatever gives the best possible model. This pragmatic approach might put physics in the back seat, especially for smaller datasets.

4. SATELLITE MODES, MODEL AND RESULTS

We designed a fully convolutional segmentation model with learned upsampling and skip connections, see figure 3. By now this is a quite standard approach, popularized by e.g. U-net.  

---

6 A SAR image has range and cross-range axes, with the former denoting distance from the sensor.
Figure 3. Our model: This a work in progress so none of the hyperparameters are final. Due to the fully convolutional nature of the network, the input can be any multiple of $256 \times 256$. Note that the receptive field of the most downsampled layer is an important design choice as it dictates the size of the region which can contribute to the detection decision for each pixel. The final two upsampling blocks required to recover the input resolution have been omitted, since our labels are weak we see no operational downside to a $4 \times$ downscaled classification output.

4.1 Preliminary Results

Based on an assessment of the available data and operational needs, we chose the Copernicus Sentinel 1 mission, satellites A and B, IW GRDH mode, as our pilot. That dataset was divided into 1358 training images and 129 validation images, containing respectively 19316 and 927 manually verified vessel detections. Each image is several gigabytes, meaning that training is highly time-consuming.

Table 1. A comparison between KSAT’s in-house classical constant false alarm rate (CFAR) ship detector and a preliminary version of our convolutional neural network (CNN) ship detector.

<table>
<thead>
<tr>
<th></th>
<th>CFAR</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detections</td>
<td>879</td>
<td>803</td>
</tr>
<tr>
<td>Misses</td>
<td>48</td>
<td>124</td>
</tr>
<tr>
<td>False Alarms</td>
<td>1021</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 1 compares the deep learning approach with the currently used classical ship detector. Since these algorithms both are input to a human-in-the-loop which manually assesses each vessel candidate, removing a false alarm represents significantly less work than finding a missed vessel. However, for potential fully automatic services these results are particularly encouraging, especially considering this is only a pilot study on a part of our dataset, with limited training time, no model optimization etc.

4.2 Future Work

To put KSAT in a leading position in terms of developing maritime situational awareness solutions, we are currently building an infrastructure of software and hardware capable of training and deploying models based on neural networks. Concurrently, we must develop novel methods to use our vast but weakly labelled dataset,
and conduct research into the validity and effectiveness of SAR image augmentation and other fundamental questions.

REFERENCES


