Towards automated identification of ice features for surface vessels using deep learning

IOP Conf. Series: Journal of Physics: Conf. Series **1357** (2019) 012042 doi:10.1088/1742-6596/1357/1/012042

Towards automated identification of ice features for surface vessels using deep learning

E Kim[1], **N Panchi**[2] **and G S Dahiya**[3]

[1] Norwegian University of Science and Technology, Trondheim, Norway
[2] Indian Institute of Technology Kharagpur, Kharagpur, West-Bengal, India
[3] Dahiya ENK, Trondheim, Norway

E-mail: ekaterina.kim@ntnu.no

Abstract. Ship traffic in ice exposed areas is increasing, and ice navigation is largely a manual task. Despite the progress in machine learning and computer vision algorithms, little focus has been given to computer-aided scene understanding in icy waters to help modernize navigational support systems. This work lays the foundation for the automated identification of ice for surface vessels using modern deep learning (DL) algorithms. The focus is on locating and classifying multiple ice objects within images from a surface vessel travelling through icy waters. The following categories of surface ice features are considered: level-ice, deformed ice, broken-ice, icebergs, floebergs, floebits, icefloes, pancake-ice, and brash-ice. In the first phase, we used DL algorithms to classify the ice objects in an image. For this task, seven state-of-the-art residual network (ResNet) models have been tested and include ResNet18, ResNet34, ResNet50, SE_ResNet50, Xception-Cadene, Inception-v4, and Inception-ResNet-v2. During the second phase, we used DL algorithms to locate and classify ice objects. For these tasks, we used the UNet architecture combined with conditional random fields (CRFs) and analysed the effects of using fully connected CRF and convolutional CRF. We have trained and validated the models using the close-range optical ice imagery, and then the promising models were used to classify and locate the different ice features in images from the bridge of the US Coast Guard icebreaker Healy and the nuclear-powered icebreaker 50 Let Pobedy. This paper provides the main findings and lessons that were learned from the execution of this study.

1. Introduction

Climate change affects the Arctic. Surface vessels operating in icy waters have become a common activity, and cruise ship traffic has increased in ice-exposed areas. In these areas, situational awareness at all times is essential for safe and efficient navigation. Ice navigation tasks include reliably detecting ice, interpreting ice conditions, finding a safe navigation route, and manoeuvring in the ice. A summary description of these tasks is provided in the following paragraphs.

Ice detection. Ice information in the form of ice charts is considered to be 'theoretical', and marine radar is used as a primary means of finding ice. For navigational purposes, a radar's range is typically set to 10 – 12 nm, but for finding ice and ways through the ice, it is set to 2 – 4 nm. The ice without snow cover can be detected at approximately 5 – 7 nm using the radar; however, it is more challenging to detect ice that is covered by snow and melting or to detect small isolated chunks of floating ice (e.g., growlers) in high sea states. Another indirect way of knowing that an ice edge is near is to monitor the water temperature and/or waves. The water temperature and swell are lower closer to the ice edge,

provided there is no wind change. The presence of ice can be detected as a light strip on the horizon. In this way, it is possible to detect ice at a distance of two times the visibility range. Furthermore, the presence of sea birds and animals can indicate an ice edge within approximately 12 nm. In the dark, a search light and deck illumination can be used for ice edge detection. However, in convoy operations, the search light can blind the icebreaker captain. In the recent years, increasingly more operators have used satellite data that have been input into a navigation system, and the challenge is to retrieve ice information at the right levels of detail at the right time, considering that ice conditions can rapidly change, especially within straights. Visual surface observation is still the best means of detecting hazardous ice and hazardous ice characteristics (ice pressure, ice adhesion, ice drift and icing).

Interpretation of ice conditions and finding weak ice. The World Meteorological Organization [1] developed well-established nomenclature for classifying sea ice. A trained ice navigator is required to identify various shapes and forms of ice, to reliably recognize ice and, when possible, to avoid the most dangerous forms of sea ice (e.g., grounded or stranded ice). Differentiating first-year, second-year and multi-year ice is challenging in good weather [2], but fog, snow and darkness make this task even more difficult. The age of the ice may be determined by using its colour, thickness, freeboard, floe shape, size, ponding/drainage, hummocking, etc., and the details can be found in the information that is reported in [2]. New ice is grey, and water can be seen through the ice. Light blue coloured ice indicates stronger ice (multi-year ice), and contact with this ice should be avoided. It is also recommended to avoid ice that is covered by snow since it causes additional friction on the hull. Flat and even ice cover is preferable over hummocks. The wind direction should also be considered as important since it can cause the ice to come closer to the ship at any moment in time.

Manoeuvring in ice. Manoeuvring in ice includes tasks such as safe passage planning and turning in ice, estimating the place where you can meet ice, following an icebreaker, turning in an ice canal, mooring operations, overtaking, and towing in ice. It is recommended to follow open water and leads, and avoid contact with ice as much as possible, especially multiyear ice or deformed ice. When approaching the ice, it is recommended to change the engine to manoeuvring mode, decrease the speed, and make ballast adjustments. It is better to enter the ice at a right angle with a minimum steering speed. Arctic surface vessels are usually supplied with a polar water operation manual and an ice navigation certificate where the information on safe vessel speeds is provided. The manoeuvring tasks in ice are complex and are not further discussed.

Currently, all navigation of surface vessels in ice-infested waters is based on the experience of ice pilots and captains. To support ice transit decision process, it is of great advantage to be able to automatically identify ice types within an overall ice regime. Although considerable research has been devoted to autonomous shipping technology for aspects, such as navigation support, equipment monitoring, and safety improvements, rather less attention has been paid to computer-aided scene assessment from ships that are travelling in ice. Examples of relevant works are given in [3–10] and include techniques and sensor systems for analysis of ice concentrations, floe-size distributions, and drift speeds using shipborne images from optical cameras and/or marine radars. However, to the best of our knowledge, no algorithm can discriminate between ice types using optical images from a surface vessel.

Motivated by the recent progress in deep machine learning and computer vision in the field of scene assessment, we assume that optical cameras can let a surface vessel see its ice environment, and the associated software can learn to recognize different ice types and present the information to the bridge for decision support. To test the capabilities of modern deep learning algorithms as well as to identify the challenges that are associated with the automated identification of ice, the following experiment has been carried out.

First, we trained and validated the classification and segmentation algorithms using the close-range optical ice imagery, and second, the most promising models were used to classify and locate the different ice features in images from the bridge of the US Coast Guard (USCG) icebreaker Healy and the nuclear-powered icebreaker 50 Let Pobedy. We limit this study to multilabel ice object classification and segmentation tasks, which form the core of autonomous navigation and navigation assistance systems.

The identification of ice parameters, such as ice concentration, stage of development (age), freeboard, size, drift speed and direction, etc., is not addressed in this study but has been planned as a future work. A brief description of the experimental setup and of the most important results is provided in the following sections. We also highlight opportunities for future work.

2. Experimental setup

The following categories of surface ice features have been considered: *level-ice*, *deformed ice*, *broken ice*, *icebergs*, *floebergs*, *floebits*, *icefloes*, *pancake-ice*, and *brash-ice* (Figure 1). This choice was limited by the available data that were gathered during a cruise to the Fram Strait on the RV Lance in March 2012 and by querying the Google, Yandex, and Baidu search engines in different languages. The dataset contains optical images that are predominantly taken from on board the vessels. These images were generally captured in daylight under good visibility conditions. Only close-range optical images of ice surface features have been considered, and these images were manually labelled in accordance with the definitions from [1]. Additional features (leads, fractures, cracks and polynyas) were not labelled. All the deep-learning based models in this study were built using the *fastai* [11] and *PyTorch* libraries [12]. Almost all of the computations for the image classification task (Section 2.1) were performed on an NVIDIA TITAN X (12GB) GPU, and for the segmentation task (Section 2.2), the computations were performed on an NVIDIA V100 (16GB) GPU that was located at UNINETT/Sigma2 AS in Norway.

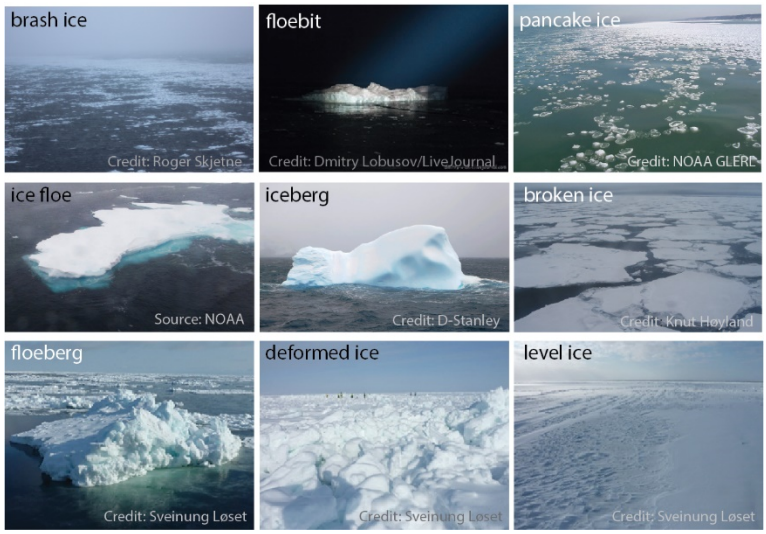


Figure 1. Typical optical images of surface ice features.

2.1. Ice object classification from an image

Convolutional neural networks (CNNs) have been chosen for this task, as a growing number of studies report similarities in the way CNNs and the human visual system processes objects, and [13] provides an overview. We have started with pretrained residual neural network models, namely, ResNet18, ResNet34, and ResNet50 [14], SE_ResNet50 [15], Xception-Cadene [16], Inception-v4 and Inception-ResNet-v2 [17]. We trained the model in two phases. First, we trained only the upper fully connected layers using the collected ice imagery, whereas all other layers of the model had no change in their weights (since we froze the weights of the bottom layers). This is because the bottom layers have their weights pretrained by using the ImageNet dataset [18] and we did not want to change them during the initial phase. Then, we unfroze the remaining layers of the model and fine-tuned the whole model using differential learning rates [19]. The optimum learning rate was determined by finding the value at which the learning rate is the highest, and the loss is still decreasing. For the upper layer, it was set to 10^{-2}, and that for the whole model ranged between 10^{-6} and 10^{-3}. Smaller learning rate values were used for the lower layers of the model and larger values were used for the upper layers. To handle a small amount of

training samples (a total of 370 raw images), we exposed the model to more aspects of the data. More training data were generated from the available training samples using data augmentation. For each image in the dataset, we followed standard practices [14 – 17] and performed random geometric and colour augmentations, including zooming, cropping, horizontal flipping, rotations, lightening and contrast changes, and symmetric warping. These data augmentation increased the breadth of information that the model can analyse to learn from a given dataset. In addition, an adaptive synthetic sampling approach was adopted for imbalanced learning by artificially generating additional data for categories with few images, such as pancake ice, ice floes, floebergs, floebits, and icebergs. The model was exposed to different image sizes (resolutions of 224 × 224 and 299 × 299), which enables the model to better generalize and handle the small dataset size. The training stopped after 40 epochs for the two different image sizes (20 epoch per size), which took approximately 30 minutes on a single NVIDIA GPU (12 GB). The learning rate was manually adjusted throughout the training when the validation error plateaued. The weight decay and dropout rates were adjusted to improve the model's performance and generalization.

To evaluate the performance of the model and to compare the prediction qualities of different models, two metrics were used: The F-beta score [20] (β=2, library default) and the accuracy. Overall, both SE_ResNet50 and ResNet34 performed better at classifying broken ice, deformed ice, and brash ice, whereas the performance for minority classes (floebits, floebergs, icebergs, and pancake ice) was lower. The F-beta score was up to 0.75, and the accuracy was as high as 92%. Analysis of the misclassified samples has also been carried out. ResNet34 outperformed the other tested models. The most challenging tasks were to distinguish between pancake ice and brash ice, as well as between broken ice and the partially melted level ice. Further testing of ResNet34 on images with distorted ice textures and colours revealed that the trained model is biased towards textures. The model could classify ice objects from greyscale images with poorer performance on the significantly distorted images with a small degree of local ice texture.

2.2. Segmentation of ice objects in an image

The training and validation datasets consisted of 390 optical images and their corresponding segmentation masks (Figure 2). The major part of the images (370) was the same as in the above classification task except for the additional 20 images containing a man-made channel in ice and five additional classes (open water, melt ponds, underwater ice, sky, and shore). Each image was manually labelled to obtain the ground truth segmentation masks, as shown in Figure 2. Our goal is to create a model that will generate these masks for an unseen image as accurately as possible. The CNN architecture that we have used is a modified version of UNet [21], which is a neural network architecture commonly used for image segmentation. Experiments were run with replacing a standard backbone of UNet with ResNet34, ResNet50, and ResNet101 [15]. The UNet architecture consists of a down-sampling encoder and an up-sampling decoder. For the down-sampling encoder, ResNets that are pre-trained on ImageNet [18] were utilized. With the help of transfer-learning we ensured that the common features in images such as circles and edges are pre-learned by the network and when training on our data, the model learns the high-level features that are specific to our dataset. The learning rates for the training were chosen by plotting the loss versus the learning rate curve over a training iteration (epoch). The learning rate at which the loss decreases at the highest rate before increasing again was chosen. Like the procedures in Section 2.1, the training of the model was done in multiple phases with different learning rates. A smaller learning rate was chosen for the second phase of the training to ensure better model performance. The best weights (the weights from the epoch at which the mean intersection-over-union (IoU) score of the classes is the highest) from phase 1 were used as the starting weights for the next phase of the training. The IoU metric is the number of pixels common between the target and prediction masks divided by the total number of pixels present across both masks.

The predictions from the models were post-processed to remove the noise. Two different conditional random fields (CRFs) based post-processing techniques were used: fully connected CRFs [22] and convolutional CRFs [23]. The parameters of the CRFs were adjusted to get the better post-processing

performance. The weight decay rate and the dropout rate were chosen using iterations to improve the model performance. The models were trained on two image sizes, 256 × 256 and 512 × 512, and the best model from the 256 × 256 images was chosen as a starting point for training on 512 × 512 images. This approach helps the model to better generalize when using the 512 × 512 images.

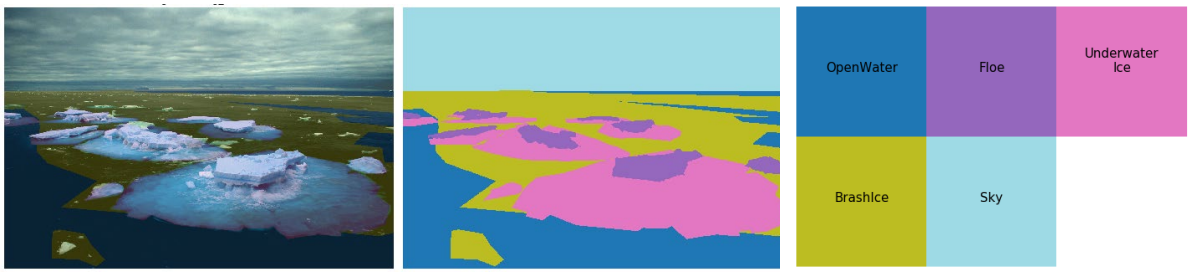


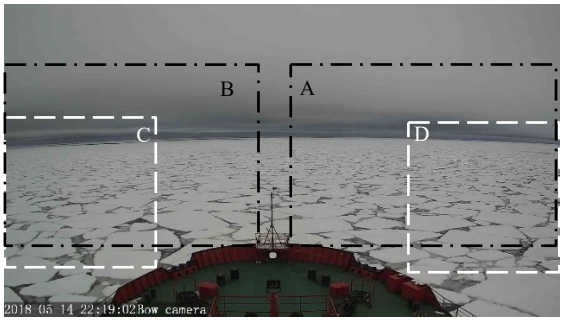
Figure 2. (from left) Image (Credit: Liliane Arizmendi) overlapped with manual segmentation mask, manual segmentation mask and classes in the segmentation mask.

IoU, which was averaged over all classes, was chosen as an evaluation metric for the models. The UNet model with ResNet34 as the backbone outperformed other models on the 256 × 256 images, with a mean IoU score of 0.51 without post-processing and 0.70 after post-processing using the convolutional CRF. An average increase of 0.19 was seen in the mean IoU over 30 epochs as a result of post-processing using the convolutional CRF and an average increase of 0.10 was seen in the mean IoU as a result of post-processing using the fully connected CRF. A comparison between the time consumption of the convolutional CRF and the time consumption of the fully connected CRF was also performed. The fully connected CRF took significantly more time (an additional ~60 seconds per epoch) compared to the convolutional CRF with little average difference in the mean IoU. Further analysis of the misclassified samples revealed that the model was most confused between floebergs and deformed ice, as well as between pancake ice and brash ice. There were difficulties in identifying small parts of the shore near the horizon. The sky was the easiest part to predict, while melt ponds were almost always missed and classified as a part of the ice floe. In some occasions, less concentrated brash ice was classified as open water.

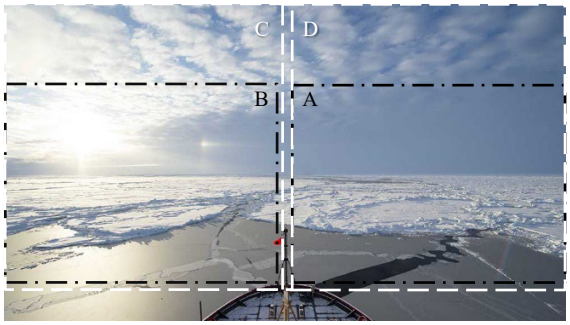
3. Models testing and main results

The most promising approaches in Section 2 (i.e., ResNet34, UNet with ResNet34 with and without the convolutional CRF) were used to detect and classify the different ice features in publicly-available optical images from the bridge of the US Coast Guard icebreaker Healy and of the nuclear-powered icebreaker 50 Let Pobedy (Figure 3). The images from the Healey contain young ice, first-year ice, ridges, rafted ice and some open water, whereas the images from the 50 Let Pobedy predominantly contain first-year ice floes with few inclusions of brash ice and some open water. This identification of ice is based on definitions in [1]. The vessels' bows were excluded from the classification- and segmentation tasks by manually defining regions A–D as shown in Figure 3.I. The A and B parts have been used for the classification task, whereas the C and D parts were used for the segmentation task. Figure 3.II – 3.IV shows the models predictions for these regions. Performance of the models was assessed only on the A–D regions and is summarized in the following paragraphs.

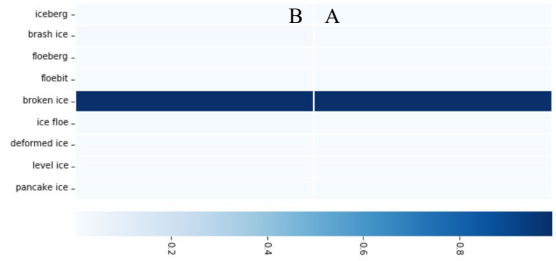
The trained ResNet34 model correctly classified the ice objects in the optical images (two images per view) from the 50 Let Pobedy and Healey (images A and B in Figure 3.II). Note that in this study, we have used general terms, such as *broken ice* and *deformed ice,* rather than specific subdivisions, such as close ice, very close ice, rafted ice and ice ridges. In particular, the general term *broken ice* have been used to define predominantly flat ice cover that is broken by gravity waves or melting, and the term *deformed ice* have been used for ice that has been squeezed together and, in places, forced upwards (and downwards).



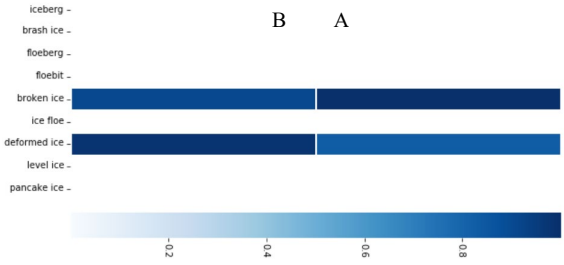
I. Original image from 50 Let Pobedy and the regions A – D exposed to the models (source: LiveJournal)



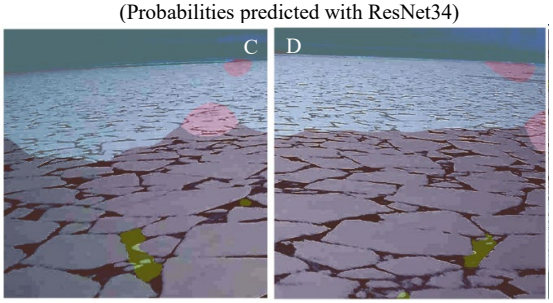
I. Original image from Healey and the regions A – D exposed to the models (source: coastguard.dodlive.mil)



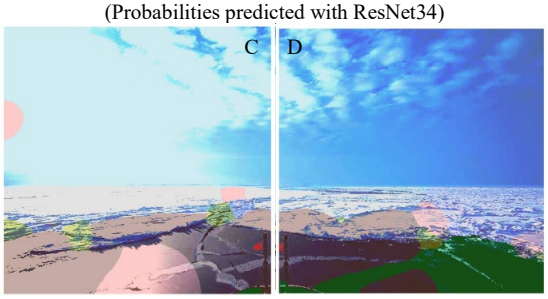
II. Ice objects classification (Probabilities predicted with ResNet34)



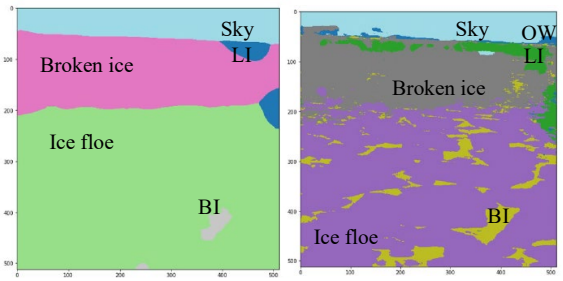
II. Ice objects classification (Probabilities predicted with ResNet34)



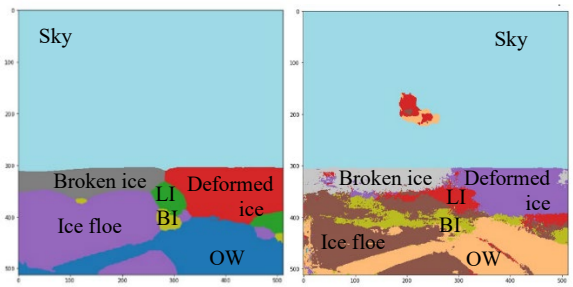
III. Image overlapped with segmentation mask (UNet with ResNet34 + convolutional CRF)



III. Image overlapped with segmentation mask (UNet with ResNet34 + convolutional CRF)



IV. Segmentation mask for D with CRF (left) and without CRF (right)



IV. Segmentation mask for D with CRF (left) and without CRF (right)

Figure 3. Classification and segmentation results of the views from the bridge of the USCG icebreaker Healy and from the 50 Let Pobedy (LI – Level Ice, BI – Brash Ice, and OW – Open Water).

The image segmentation results (Figure 3.III – 3.IV) using UNet with and without convolutional CRF post-processing are less perfect but promising. Overall, the predicted class labels with UNet do not contain false positive or false negative labels that could challenge the identification of hazardous ice features. There are limitations to detecting isolated ice floes, especially when using CRF post-processing. Some artefacts in the predictions with UNet are visible in Figure 3.IV. Another limitation that is presented in Figure 3.IV is that the grey ice was partially identified as open water, and open water was partially identified as brash ice and ice floe.

Future work should focus on addressing these limitations, on model testing and on improving the overall accuracy of the predictions.

4. Concluding remarks

Ice navigation is a complex task that includes the identification of hazardous ice. Currently, the navigation of surface vessels in ice-infested waters is largely a manual task that requires singnificant training and experience. To support the development of the data-driven automated identification of ice features from a surface vessel, we have presented the first results on classifying multiple ice objects within an image as well as locating ice objects in the image with the corresponding ice category. In this work, close-range optical images of ice have been used to 'train' models to locate and classify nine different ice types. After training the models, we have tested their capabilities to recognize the ice features in optical images that was taken in daylight from the bridge of the USCG icebreaker Healy and the icebreaker 50 Let Pobedy. The results are limited to a few images of a certain resolution and show that despite the small amount of data (370-390 original images for training and validation), the ResNet34 model was able to classify the broken and deformed ice in the optical images. Locating surface ice features in images from the bridge was shown to be more difficult and has limitations, but we expect that the imperfect predictions of the models will improve as more systematically labelled data become available. The model performance on images of higher resolutions should also be checked.

The presented work opens many opportunities for future research, as the amount of optical ice imagery that is collected from vessels is rapidly increasing (refer, for example, to the near real-time web camera feeds from icebreakers and research vessels, such as the USCG Healy, the 50 Let Pobedy and the Aurora Australis). Below, we highlight issues that are important to consider when planning to automatically interpret the ice conditions for a surface vessel.

Data collection. The currently available optical images of ice are predominantly taken under favourable weather conditions with good visibility. Images of ice that are taken using artificial light or in poor lightning conditions, heavy snow, rain, and fog are scarce, and efforts are needed to collect and reliably label such data. Web cameras seem to be a good way to collect such data. A web camera should have a wide angle with minimum distortion, a high enough resolution, and an adequate sampling frequency. We recommend using a variable sampling frequency. More images could be collected during difficult ice conditions or in poor visibility conditions. Lower sampling frequencies can be used in open waters or in low ice concentrations. Not all ice going vessels are equally good for collecting ice data. On some vessels, cranes and other equipment may obstruct the view, thus reducing the clarity of the images. It is recommended that ships should place optical cameras on the bridge to avoid icing and malfunctions due to weather conditions. It is also recommended to collect the elevation of the mounted camera and a scale for identification of the size of the ice features as meta information along with the images. Without a reference, it is difficult to identify the sizes of the features for labelling purposes, this is especially true in case of the ice floes.

Data labelling. Data labelling is one of the most time-consuming tasks, especially for broken ice fields with high ice concentrations, for which each and every ice floe, including the areas with brash ice and melt ponds, need to be labelled. None of the currently available labelled datasets are perfect with regard to the ice features. Often, there are few ice categories (mainly icebergs and ice fields), and many ice features are incorrectly labelled. For the multi-class image segmentation task, we recommend following a set of labelling rules because the human identification of ice depends on experience, and there will be images where it is impossible to reliably identify ice features.

We expect that future work will result in an open ice-scene dataset for the training and benchmarking of models and in future improvements of the quality of optical ice images. We hope that the presented findings can stimulate and support the development of ice navigation support systems.

Acknowledgements

The authors thank Professors Sveinung Løset, Knut Høyland, and Roger Skjetne sharing their photographs of ice and UNINETT Sigma2 AS, which is the national infrastructure for computational science in Norway, for granting access to their data storage and processing resources.

References

[1] WMO 1970 *Sea ice nomenclature; Terminology, Codes, Illustrated Glossary and Symbols* (World Meteorological Organization WMO no. 259)
[2] Johnston M E and Timco G W 2008 *Understanding and identifying old ice in summer* (Canada: Canadian Hydraulics Centre, National Research Council CHC-TR-055)
[3] Muramoto K, Matsuura K and Endoh T 1993 Measuring sea-ice concentration and floe-size distribution by image processing *Ann. Glaciol*. 18 33
[4] Hall R J, Hughes N and Wadhams P 2002 A systematic method of obtaining ice concentration measurements from ship-based observations *Cold Reg. Sci. and Tech*. 34 2 pp 97-102
[5] Lu P and Li Z 2010 A method of obtaining ice concentration and floe size from shipboard oblique sea ice images *IEEE Trans. on Geosci. and Rem. Sens*. 48 pp 2771-2780
[6] Ji S, Li H, Wang A and Yue Q 2011 Digital image techniques of sea ice field observation in the Bohai Sea, *Proc. POAC11-077*
[7] Zhang Q and Skjetne R 2015 Image processing for identification of sea-ice floes and the floe size distributions *IEEE Trans. on Geosci. and Rem. Sens*. 53 pp 2913–2924
[8] Lu W, Zhang Q, Lubbad R, Løset S and Skjetne R 2016 A shipborne measurement system to acquire sea ice thickness and concentration at engineering scale. *In Proc. Offshore Tech. Conf.* doi:10.4043/27361-MS
[9] Heyn H-M, Knoche M, Zhang Q and Skjetne R 2017 A system for automated vision-based sea-ice concentration detection and floe-size distribution indication from an icebreaker *In Proc. Int. Conf. Ocean, Offshore & Arctic Eng*. 8
[10] Kjerstad Ø K, Sveinung L, Skjetne R and Skarbø R A 2018 An ice-drift estimation algorithm using radar and ship motion measurements *IEEE Trans. on Geosc. and Rem. Sens*. 56 6 pp 3007–3019
[11] Howard J, Thomas R and Gugger S 2018 Fastai GitHub
[12] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L and Lerer, A 2017 Automatic differentiation in PyTorch *In NIPS-W*
[13] Geirhos R, Janssen D H J, Schütt H H, Rauber J, Bethge M and Wichmann F A 2017 Comparing deep neural networks against humans: Object recognition when the signal gets weaker arXiv:1706.06969
[14] He K, Zhang X, Ren S and Sun J 2016 Deep Residual Learning for Image Recognition arXiv:1512.03385
[15] Hu J, Shen Li and Sun G 2018 Squeeze-and-Excitation Networks arXiv:1709.01507
[16] Chollet F 2017 Xception: Deep Learning with Depth wise Separable Convolutions arXiv:1610.02357
[17] Szegedy C, Ioffe S, Vanhoucke V and Alemi A 2017 Inception-v4, Inception-ResNet and the impact of residual connections on learning arXiv:1602.07261
[18] Deng J, Dong W, Socher R, Li L-J, Li K and Fei-Fei, L 2009. ImageNet: A Large-Scale Hierarchical Image Database *IEEE* Conference on Computer Vision and Pattern Recognition pp 248-255
[19] Smith L N 2018 *A disciplined approach to neural network hyper-parameters: Part 1 --learning rate, batch size, momentum, and weight decay* (US Naval Research Laboratory, Technical Report 5510-026)
[20] Chinchor N 1992 MUC-4 evaluation metrics. *In Proc. MUC-4* pp 22–29
[21] Ronneberger O, Fischer P and Brox T 2015. U-Net: Convolutional networks for biomedical image segmentation arXiv:1505.04597
[22] Krähenbühl P and Koltun A 2011 Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials in *Proc. Advances Neural Inf. Process. Syst.* pp 109–117
[23] Teichmann M T and Cipolla R 2018 Convolutional CRFs for semantic segmentation arXiv:1805.04777