

Generative Network For Semi-supervised Sea Ice Classification

James Imber, *Member, IEEE*

Abstract—The classification of sea-ice as a key element in monitoring the polar regions. There exists an extremely limited amount of ground truth information from direct observations on the ice surface. Furthermore, with the introduction of each new sensor system the process of generating labelled data for model development must be repeated. Generative network architectures with suitable constraints are powerful tools for minimally-supervised learning. Auto-encoding can also provide a strict constraint on network behaviour. This paper presents a novel extension to a network architecture concept that combines these elements known as an adversarial auto-encoder. Regularisation of the latent space representation using adversarial network elements produces a classification layer that can be directly applied to the task of semantic image segmentation. Here this method is employed specifically for the classification of sea-ice in SAR images. Results from training conducted on three TS-X scenes show the network is able to classify the unlabelled portions of the training images and completely independent scenes, despite very limited supervision from labelled data. Acquisitions collocated with ICESat-2 tracks confirm that the network is performing as intended.

Index Terms—Adversarial Auto-encoder (AAE), Sea-ice.

I. INTRODUCTION

Sea-ice is a key feature of polar regions and, given the rapid changes currently taking place there, the monitoring of sea-ice has become particularly urgent, as one of the indicators of the progression of global climate change [1]. It has been demonstrated that the reduction in the surface extent of sea-ice has been compounded by a decrease in the average thickness of the remaining ice as the onset of melting occurs earlier in the spring and lasts longer in to the autumn [2]. This has further motivated more detailed analyses to extract information on the distribution of the different types of sea-ice. Sea-ice also continues to pose a significant threat to shipping at high latitudes, the level of risk also depending on the sea-ice type distribution. Synthetic aperture radar (SAR) has long been favoured for sea-ice monitoring, combining wide swaths with high spatial resolution and all-weather capabilities, and perhaps most importantly, the ability to probe the nature of the sea-ice surface and by extension the processes that lead to its current state. This study employs dual-polarisation data (HH and VV) acquired by a single SAR sensor, however the basic methodology is transferable to other datasets. Sea-ice classification can be thought of as a classic semantic image segmentation problem. Efforts to extract sea-ice types from SAR images have developed from pixel-wise classification using purely radiometric and polarimetric features, to analyses

incorporating spatial texture information, in large part driven by the adoption of machine learning methods. An extensive and detailed review of the work done in this field can be found in [3]. The classification of sea-ice in dual-polarisation X-band imagery using selected polarimetric features has already been demonstrated [4] and compared with that of C-band and L-band sensors [5]. Local textural features such as those from gray level cooccurrence matrix (GLCM) analysis have also been shown to distinguish sea-ice types [6]. In both cases the optimal combination of these features was reached using shallow neural networks. Many deep convolutional neural network architectures have been proposed to solve this kind of problem, however they commonly rely on a large number of labelled examples for training, otherwise known as supervised learning. The amount of labels available to train sea-ice classifiers is very limited given the difficulty and expense of operating in remote regions at high latitudes. What labelled data there is therefore tends to be one of two kinds. The first are direct measurements taken during an expedition, with high accuracy and high spatial resolution but very limited in extent. This kind of data was employed as ground truth in [6] and [5]. The second kind is from sea-ice charts, derived from the manual interpretation of observations made by experts, such as those of the Canadian Ice Service (CIS) [7] or Norwegian Ice Service (NIS) [8]. These charts cover very large areas and are regularly updated, but with comparatively coarse resolution. In [9] CIS charts were used as labels to apply a single class to (5,5) km squares from within a set of SAR images, which were then used to train a DenseNET.

Semi-supervised learning strategies capable of extrapolating from a limited set of ground truth may offer a solution that maintains high resolution and broad domain applicability. Auto-encoding is a frequently used element of networks for learning with limited supervision. The encoding layers must learn to extract the essential information about the input into the latent space representation and training can be performed by comparing the input with the decoded image produced at the end. In this case the latent representation is an arbitrary compression of the input image. However, by the imposing of suitable constraints it is possible to regularise the latent space representation and thereby render it directly interpretable. One such method is derived from generative networks. The concept relies upon the application of auxiliary constraints to restrict the network behaviour. This study is based upon a particular generative network architecture known as an Adversarial Auto-encoder, the details of which are described in section II, that has been adapted for the purpose of sea ice classification.

The standard method of training a neural network through gradient decent relies upon the back-propagation of the gradi-

(Corresponding author: James Imber.) The author is with the Maritime Safety and Security Lab, Remote Sensing Technology Institute, German Aerospace Center, Bremen, 28359, Germany (e-mail: james.imber@dlr.de).

ent of a loss with respect to the values passed to each node. The function used to calculate the loss depends upon the nature of the output. Classification training is commonly achieved using the cross-entropy loss, whereas for regression the loss function might be the mean squared error. However, there are training tasks for which no such function can be constructed, for example, if the desired output is to have particular statistical properties without recreating any specific output that might already exist. The solution is to replace the loss function with another neural network known as a discriminator. The primary network can then be referred to as the generator. This family of network architectures are known collectively as Generative Adversarial Networks (GAN) [10]. As discussed above, with a GAN the loss determination no longer limited to direct output-to-truth comparison. The discriminator is also a neural network and thus arbitrary and optimisable and must be trained. The discriminator is trained to separate outputs produced by the generator and output-like results of the desired form provided. Simultaneously the generator is rewarded if it can fool the discriminator. As a result, learning is no longer a high dimensional optimisation path, but instead can be thought of as a two player non-cooperative game, the new goal being a Nash equilibrium between the two actors. Stable learning can be more difficult to achieve, and oscillations and mode collapse are well known hazards. Such guidelines that do exist to inform the construction of generative networks tend to be heuristics learned from experience which, while valuable, have limited theoretical underpinning. The GAN method can be used to apply constraints to the generator during training without the need for truth information, as long as similar results are available, and it is this ability that makes it useful here.

II. METHODOLOGY

The design of the network described here is based on the concept of Adversarial Auto-encoders (AAE) [11] which are a development of the GAN concept. The aim of an auto-encoder is to force the network to encode the input image in a compact representation. An encoder E maps the input x into the encoded space $z = E(x)$, and a decoder D then reverses the operation, albeit perhaps imperfectly, mapping z to $x' = D(z)$. An AAE incorporates adversarial elements which are used to shape the nature of the encoded space, specifically to force it to match a set of predefined source distributions. In other words the encoder plays the role of the generator network in a standard GAN. The effect is both to regularise the encoding and to cause it to behave in a predictable manner. The input layers to the encoder are the two polarisation channels, HH and VV and their ratio HH/VV. The encoding takes the form of a five-type classification with six additional layers as continuous variables to account for intra-class variation of the input (including speckle) as shown in figure 1. Performing semantic segmentation using this method is made substantially more complex by the need to provide latent-space distribution examples to the discriminators with two spatial dimensions. The intra-class variation can contain some local coherence from real image features but also truly

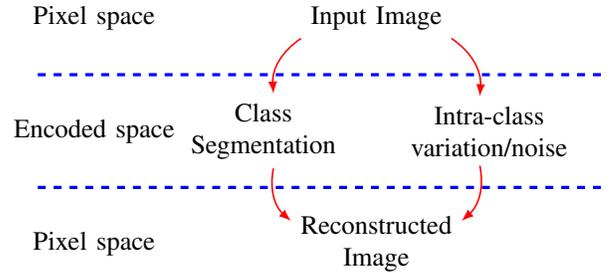


Fig. 1. A schematic illustrating the idealised flow of information through the encoding and decoding processes. The discriminators encourage the encoding to separate the left and right paths rather than having everything mixed together. In practise the separation is imperfect.

random fluctuations. The first three continuous variable layers are trained to try to match the behaviour of the distribution that describes the local fractional standard deviation of the input layers. The distribution describing the remaining three continuous variables is Gaussian with no pixel-to-pixel correlations. In general the class distribution within a patch is also random, but with manifest local correlations. Here, the discriminator is fed simulated 2D segmentations derived from contours drawn on randomly generated surfaces. The surfaces were generated by interpolating between randomly generated points in a 3D space. The number of contours (and hence the number of different classes) and the class labelling was randomly assigned. It is essential that the shape information related to the class boundaries be channelled to the class segmentation layer. Some examples of generated patches are shown in figure 2. For both discriminators, the input samples are approximations of the desired encoder response, and serve in effect as priors for the final latent-space distributions developed during training.

The encoder consists of a series of convolutional layers. In each case no padding is applied, thus the edge pixels are removed. Alternate layers have a stride length of two, thereby reducing the spatial dimensions. This is the preferred method over pooling layers for adversarial networks. The number of filters is simultaneously increased. The decoder uses inverse convolutions to increase the spatial dimensions back to recreate the input. Schematics can be found in figure 6. Batch normalisation is used in the encoder [12]. The nominal LeakyRelU activation function is used throughout.

Cyclical losses are incorporated into the training phase to facilitate the auto-encoding. The first pixel-wise loss measures the difference between the original image and the reconstructed image, constraining the behaviour of $x' = D(E(x))$. A second element-wise loss measures the difference between the first image encoding and a second encoding performed on the reconstructed image i.e. $z' = E(D(z))$. In this manner consistency of the representation in the image and latent spaces is encouraged. The encoding process reduces significantly the degrees of freedom and thus limits the accuracy of x' compared to x , and results in x' having pixel values closer to the average amplitude. To compensate, additional cyclical losses comparing the local mean, standard deviation and skew of x and x' are included.

A diagram illustrating the various network components can

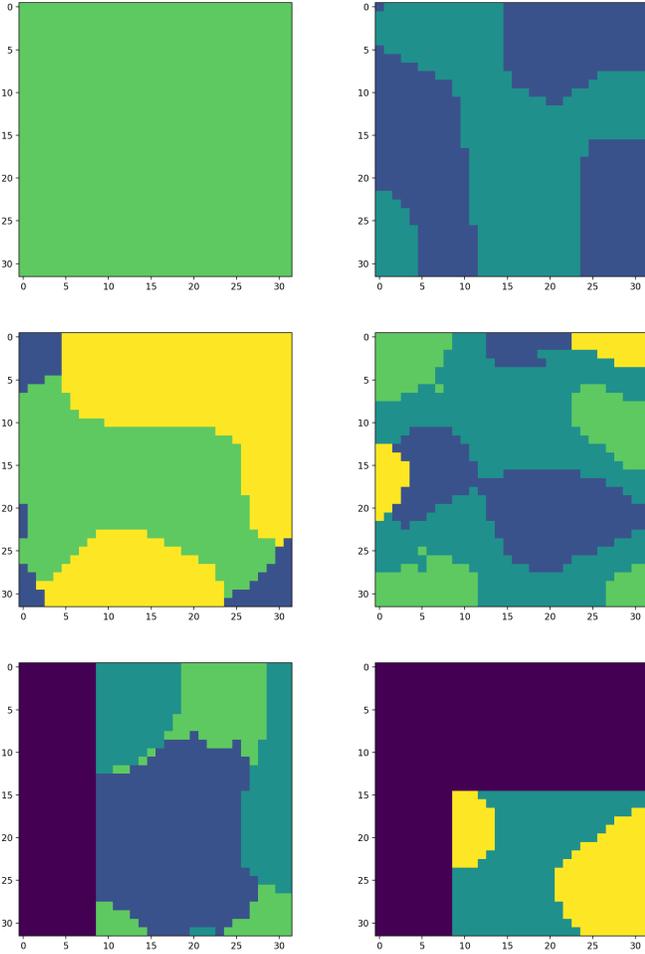


Fig. 2. Examples of randomly generated 2D class patches, from top-left; One class, two classes, three classes, four classes, and image edge and corner boundaries.

be found in figure 3. Some elements are shown twice as they are used in two places in the data flow and have additional constraints applied as a result.

Where small patches of ground truth labels exist, they are employed to provide a low level of supervision to the classification task through the standard categorical cross-entropy loss. The network should then extrapolate this labelling scheme to the remaining data based upon the global consistency of the encoded layer, enforced by the auto-encoding which receives feedback from the entire image.

Losses are accrued via the auto-encoding, the discriminators and the supervised pixels. These loss elements are summed as shown in (1), with coefficients tuned to reflect their effective credibility. The auto-encoding and supervised loss are direct learning elements and thus have larger coefficients than the discriminator derived losses. This imbalance acts to stabilise the learning process. The strength of the feedback to the network from the adversarial units increases with each step as the discriminators become more adept. This should be matched by the ability of the encoder. This behaviour is quite different from normal loss functions and optimising the

loss coefficients is less straightforward for a network in a mixed loss training scenario, than when using either direct loss functions or adversarial loss alone. To use both types of loss together requires careful adjustment, in particular modifying the loss coefficient using the epoch number N as shown in (2).

$$L_{tot} = \alpha L_{auto} + \beta L_{disc_c} + \gamma L_{disc_s} + \delta L_{sup} \quad (1)$$

where β and γ are constants and α and δ have the form,

$$f(N) = a(b - \exp(-N/c)) \quad (2)$$

Normally the inputs to a discriminator are given labels 0 and 1 for "fake" and "real" respectively. One-sided label smoothing changes the real labels to take a range of values, in this case a Gaussian distribution with $\mu = 0.95, \sigma = 0.1$. This adaption has been shown to improve learning performance in generative networks. The network parameters are updated by stochastic gradient descent using the Adam optimiser [13] which adjusts the learning rate during training. The initial learning rate is 0.0006 with the remaining parameters left at default values. Training is terminated after 250 epochs. This is a somewhat arbitrary figure at present, limited to permit the repeated training of an ensemble of models in an acceptable time frame with the hardware available.

The encoder reduces the spatial dimensions by a factor of eight, while the decoder naturally reverses that process to recover the full resolution. Given that the input layer only has three channels (the HH and VV polarisations and the ratio HH/VV), the encoding process must reduce the spatial dimensions to some extent or else no compression will have been achieved. This goes hand-in-hand with the goal to classify the input image based upon multi-pixel spatial features, rather than pixel-by-pixel independently.

The network was realised in Tensorflow 2.0 [14] and executed using an Nvidia RTX2080 GPU. With this setup, the training of each model instance required ~ 4 hours to complete.

III. DATASET

TerraSAR-X (TS-X) has been acquiring high resolution, X-band images for over a decade. This study is focused on data acquired in stripmap mode with HH and VV polarisations, multi-looked and ground range detected with a pixel size of 4 m and a resolution 8.3 m [15]. Three images acquired in May 2014 have been employed for training the network. The locations of these acquisitions (2014-05-22 A and B, and 2014-05-25) can be seen in figure 4. The images themselves are shown in the left-hand side of figures 7, 8 and 9 as reduced-resolution composites of the two polarisation channels, with more water-like pixels shaded blue. The images 8 and 9 overlap somewhat reducing the total number of independent input patches but enabling a consistency check. The following preprocessing steps are applied. The input scenes are first calibrated to σ_0 , truncated at $\sigma_0 = 0.4$ and then normalised to the range [0,1]. An additional layer, the ratio of the two polarisations HH/VV is generated and also normalised to [0,1]. The images are expanded at the edges to construct an integer

number of (256,256) pixel patches. Cropping out overlapping (352,352) pixel extended patches allows for reduction during the convolution layers of the network to guarantee uniform response across the patch. This process is visualised in figure 5. To pass the patches through the encoder network a second time, the patches are padded back out to (352,352) using a reflection of the pixels at the edge.

Five classes are defined to segment the images and describe the ice distribution. The "Water" class contains all regions of open water. "Smooth Ice/Floes" refers to ice surfaces with very low backscatter such as ice floes but potentially also very new ice. "Uniform Ice" describes regions with a moderate backscatter but with little variation. "Rough Ice" describes ice containing bright features and substantial variation. Finally, the "Empty" class contains pixels with no image information at all. The labelled pixels are shown in the central images of figures 7, 8 and 9 and summarised in table I. In total they cover 0.646% of the training images. Even for sparse ground truth, with the supervised loss being only one of many terms, class imbalance can cause problems, however in this study there is currently no evidence that the degree of class imbalance was detrimental.

IV. RESULTS

One of the difficulties that arises when using unsupervised or semi-supervised learning strategies is how to evaluate the performance of the network. In the case of ice classification,

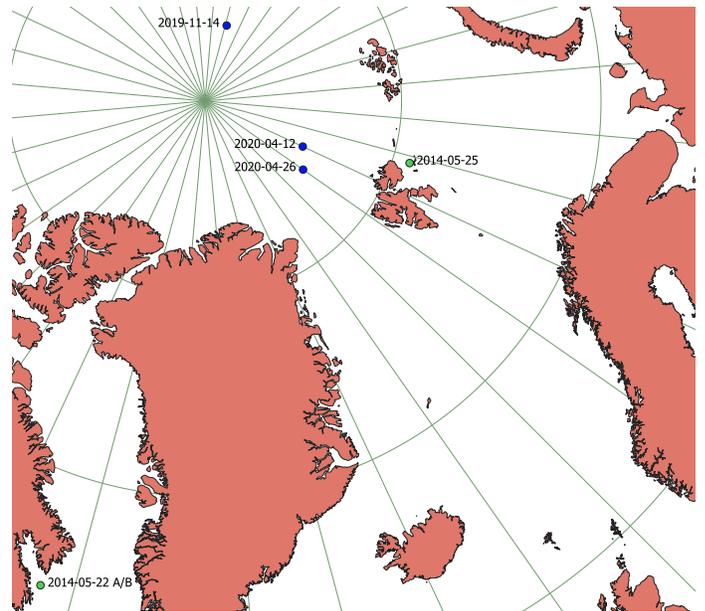


Fig. 4. An overview of the positions of the TS-X images used for training (green) and the TS-X image and ICESat-2 collocations introduced for validation in section V (blue).

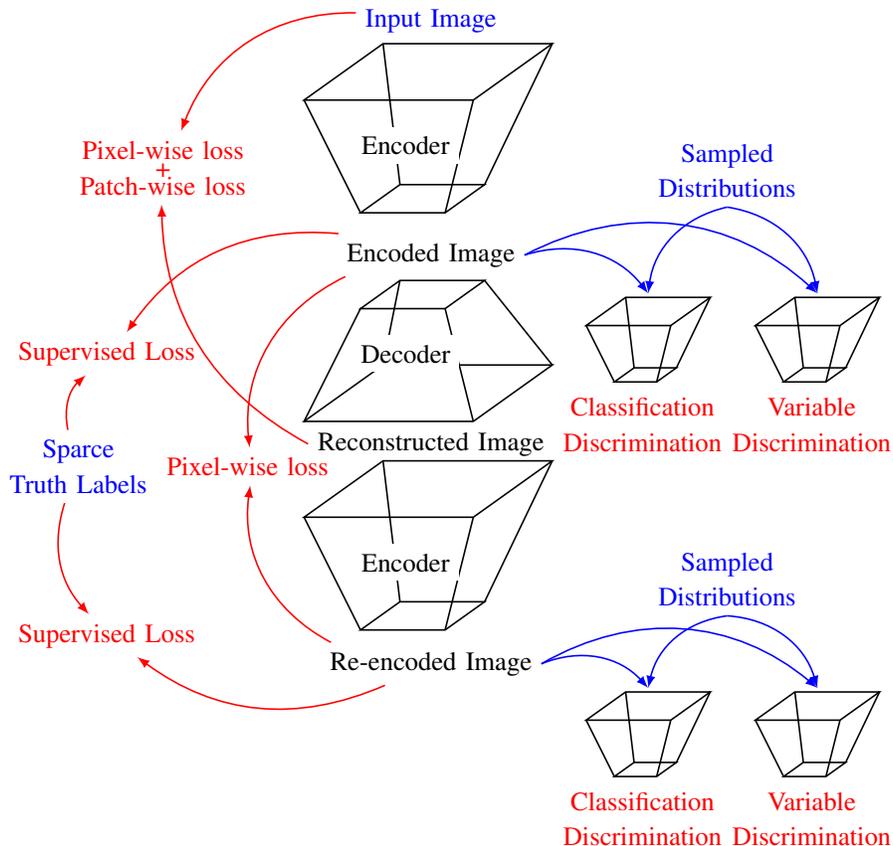


Fig. 3. A full network schematic. Inputs are highlighted in blue and losses appear in red.

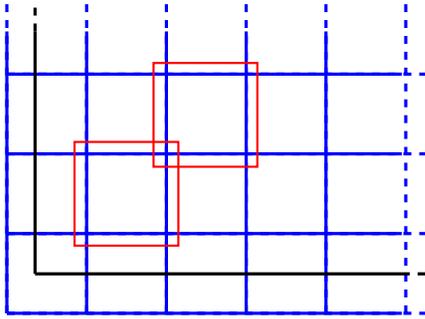


Fig. 5. An illustration of the input data patch formation scheme. Original image (black), padded with zeros to produce a grid with an integer number of (256,256) squares (blue). The input to the network to classify each region is the larger (352,352) patch aligned at the centre, for which two examples are shown in red.

TABLE I
A SUMMARY OF THE LABELLED PIXELS

Class	Total Pixels
Water	260,000
Smooth Ice/Floes	265,548
Uniform Ice	217,523
Rough Ice	145,556
Empty	270,000
All Labelled	1,158,627
Full Training Images	179,315,901

it is possible to a large extent to grade the results by eye. This approach is made easier by the markedly less subjective nature of the results presented here than those of other GAN-based work in the literature, such as producing works of art from photographs in the style of a particular painter. The accuracy of the classification on the labelled regions is $\sim 90\%$, however, given the sparse nature of the labelling, this should be regarded as a very loose estimate of the performance. A more compelling assessment can be found in the section V.

The right-hand images in figures 7, 8 and 9 are the output classification layers for each of the three scenes used to train the network. The encoding has clearly captured the salient features of the input images whilst largely ignoring irrelevant detail. The inter-class boundaries show small variations from run to run. This effect is to be expected as a natural consequence of the remaining freedom in the constraints used during training. The class discriminator acts to guide the output class shapes and the auto-encoder forces the accurate encoding of the input image, limited only by the available degrees of freedom. However, in precisely where to define the class boundaries, particularly given the reduction in resolution, there remains ambiguity. To reduce the number of outliers, an ensemble of independently trained networks provides smoother results but potentially at the expense of some of the finer details. The combination is formed by averaging the softmax output from each network or by selecting the most popular class from the final classifications. The results shown use the latter method, however both give very similar results.

V. VALIDATION

In this section is presented a comparison with an independent and wholly different means of investigating sea ice from

space, in order to verify the performance of the AAE. The NASA Ice, Cloud and land Elevation Satellite-2 (ICESat-2) is an operational mission to measure the height profile over the Earth's surface, and in particular to measure the height of ice sheets and sea ice in the polar regions. To this end the satellite carries the Advanced Topographic Laser Altimeter System (ATLAS), a space-based laser altimeter. The laser beam is split to allow measurements along three paired tracks (left and right each for tracks 1, 2 and 3) along the surface. The paired tracks are formed by one weak beam and one strong beam, the strong beam pulses containing approximately four times the number of photons as the weak beam pulses. Each pair is separated by $\sim 90\text{m}$ and each track by $\sim 3\text{km}$. The height is determined by measuring the time-of-flight (ToF) of the photons within each pulse. The level 3A sea ice height product (ATL07) was used for this study [16]. At this processing level, the detected photons are grouped into segments with 150 photons. The ToF for each photon is converted into a distance and from there a calculated surface height with respect to the WGS84 reference ellipsoid. A Gaussian function is fitted to the photon surface height distribution from each segment which gives the final surface height for each segment. The width of this distribution (twice the standard deviation), is a measure of the height variation within each segment. A surface type classification, particularly for ice-water discrimination, using the surface fit parameters and background (solar) photons as an indication of the nature of the surface scattering is also provided in the data. The performance of the ICESat-2 ATL07 product has already been validated against data acquired by the Airborne Topographic Mapper [17] flown by Operation IceBridge [18]. In addition, the surface classification, particularly open lead detection, have been evaluated using Sentinel-2 optical imagery [19]. Variations in the mean and standard deviation of the height of the segments contain information on the structure of the surface geometry. This can then be compared with the SAR response and the AAE derived classification over the same region as a means of validation.

Using ICESat-2 has a number of advantages. It is a completely different technology from SAR and thus the sources of measurement uncertainty are either entirely different or at least uncorrelated. For ICESat-2 perhaps the most significant source of error is deviation of the photon distribution from a Gaussian shape. This may be a result of large surface features like ice ridges, or multiple scattering causing additional delay in the ToF. SAR response uncertainties can arise from snow cover and snow layer boundaries, surface melt water, small scale surface roughness down to $\mathcal{O}(\text{cm})$, the surface wind speed (over water), and from the change in incidence angle across the reconstructed image.

Collocations were found using TS-X images to define a region of interest (expanded that region slightly to ensure coverage) and the python package *icepyx* [20] to select the relevant data from the ICESat-2 archive. The ice is not stationary, but slowly drifting, and thus a time window of $\pm 6\text{hrs}$ was also applied. TS-X acquisitions in stripmap mode are relatively small, and the ICESat-2 tracks also cover a narrow band, thus the highest probability for collocations comes at high latitudes, close to the pole, where their respective orbits

cross more frequently. In addition, ICESat-2 requires cloud-free skies for effective operation. The footprint of each ICESat-2 beam is approximately 17m wide which corresponds to ~ 5 pixels in the TS-X image. Slices were made to extract the equivalent information from the calibrated TS-X image and classification layer. A five by five pixel moving average over this slice was also generated for comparison. The positions of the three collocations considered in this section are shown in figure 4.

An overview of the collocated data selected from 14th November 2019 is shown in figure 10. The time difference between the acquisitions is 3:40:16, which is large enough for significant drift of the sea ice to have occurred. It was therefore necessary to apply a simple translational correction using the open leads as a guide, 270 pixels along-track and 30 pixels across-track. As can be seen in the figures, the open leads represent the starkest features in both datasets. However, this kind of translation is not sufficient to accurately model the drift vector-field and hence it remains possible to make a qualitative comparison only. Note also that this TS-X image has a slightly higher resolution than the training images with pixel size 3.5 m and resolution 7.5 m.

As already mentioned, ICESat-2 segments are composed of 150 photons, thus the higher photon rate of the strong beams means that strong-beam segments span about a quarter of the along-track distance of weak-beam segments. This means that the segment height for weak-beam segments is an average over a longer distance, approximately four times that for strong-beam segments, leading to a difference in the sensitivity to smaller scale surface features of the ice surface. This difference in segment length is demonstrated in figure

11. Note that the strong beam of track 2 has a photon rate of $\sim 85\%$ of that for the strong beams of tracks 1 and 3, leading to longer segments, although by how much depends on the exact nature of the surface and how far track 2 is from the nadir.

In the ICESat-2 data, the sea ice height and the width of the Gaussian fit to the photon scattering surface height distribution are clearly correlated as shown in figure 12. The correlation coefficients are inset on each plot. This suggests that the ice segments with a higher measured surface have a rougher geometry on scales smaller than the segment lengths. This correlation agrees with the expected behaviour that regions with thicker ice should also produce brighter SAR backscatter.

An open lead has been identified along track 1 in figures 15 and 16, at $x \approx 3500$ m which coincides with open water in the classification output. Both datasets agree that the lead branching off to the right has refrozen, leaving a smooth ice surface. The open lead at $x \approx 6200$ m in figures 17 and 18 has also been correctly identified by the AAE. However, along track 3 in figures 19 and 20, of the four open leads in the ICESat-2 data, only the widest lead at $x \approx 8000$ m has been classified as open water. This may be the result of refreezing of these narrow leads during the time between the acquisitions, or because the leads are otherwise too dark in terms of radar backscatter to produce the necessary amplitude in the HH/VV ratio to be classified as open water.

Figures 13 and 14 show overviews of two more examples of collocated data, this time acquired in April 2020. Figure 21 shows data from 12th April with a collocation time interval of 1:56:02. A translational correction of 270 pixels along-track has been applied. Figure 21 shows data from 26th April

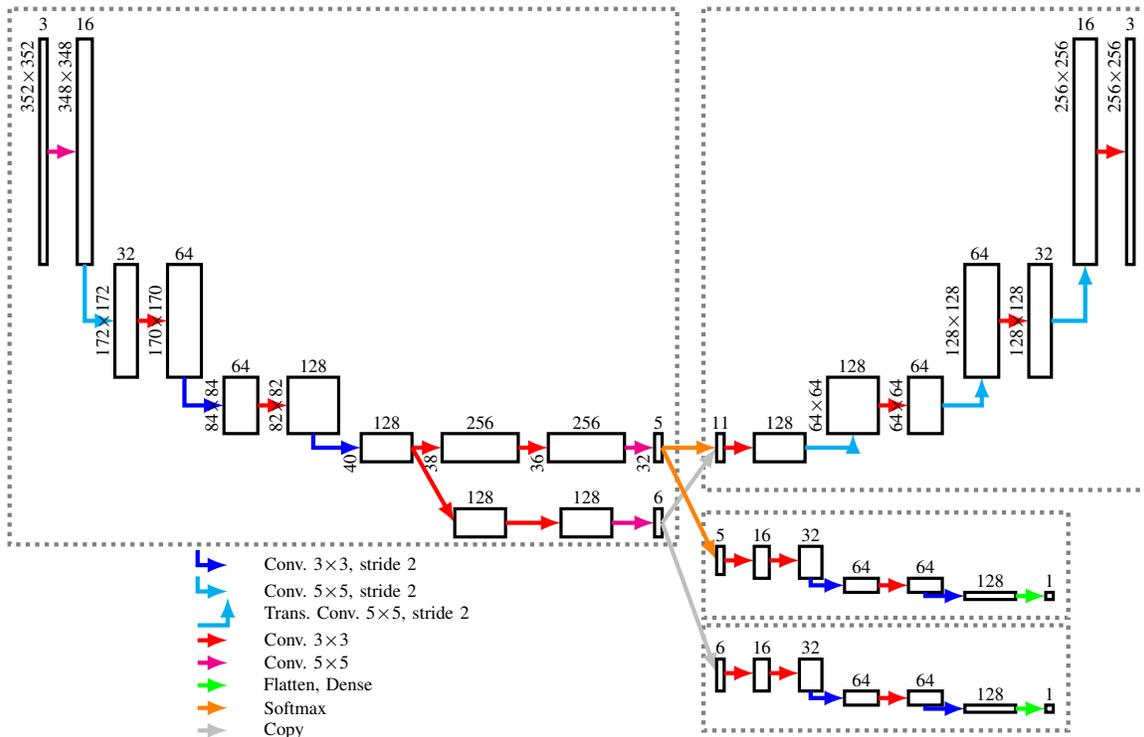


Fig. 6. A schematic of the network elements, left; the encoder, right from top to bottom; the decoder, class discriminator, variable discriminator.

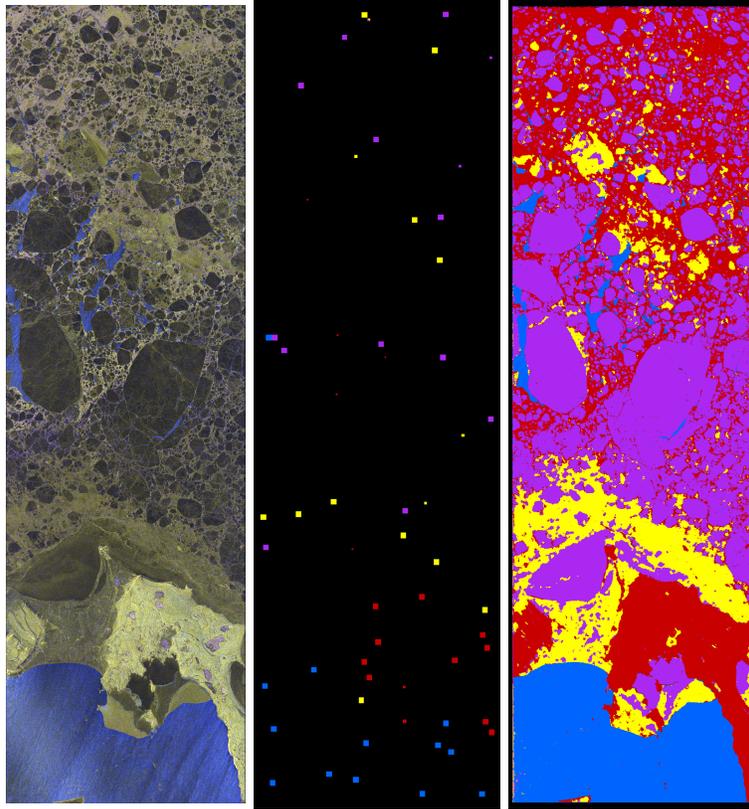


Fig. 7. Results for training image 2014-05-25 from left to right; Dual polarisation composite SAR image, labelled pixels and classified image.

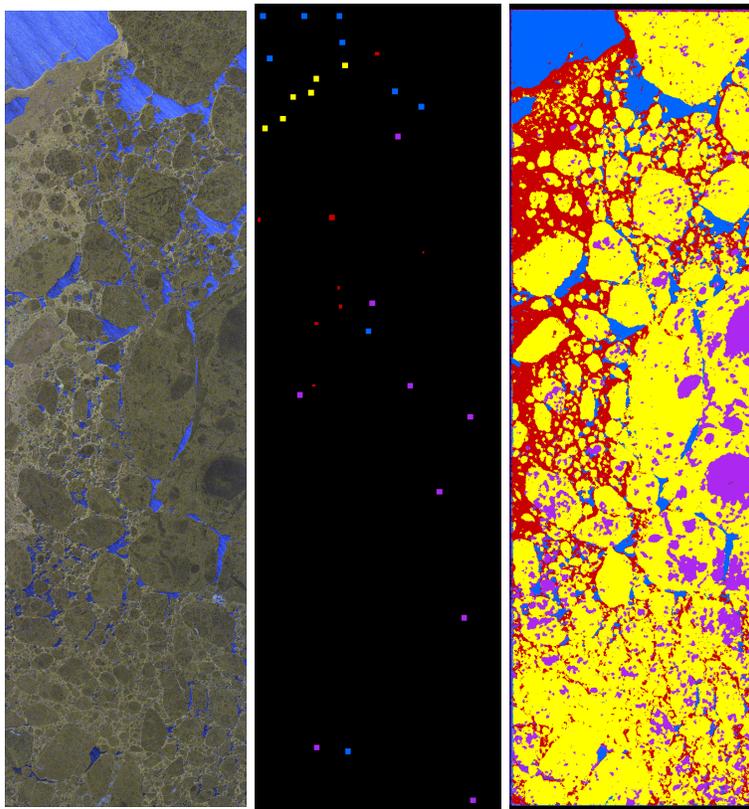


Fig. 8. Results for training image 2014-05-22_A from left to right; Dual polarisation composite SAR image, labelled pixels and classified image.

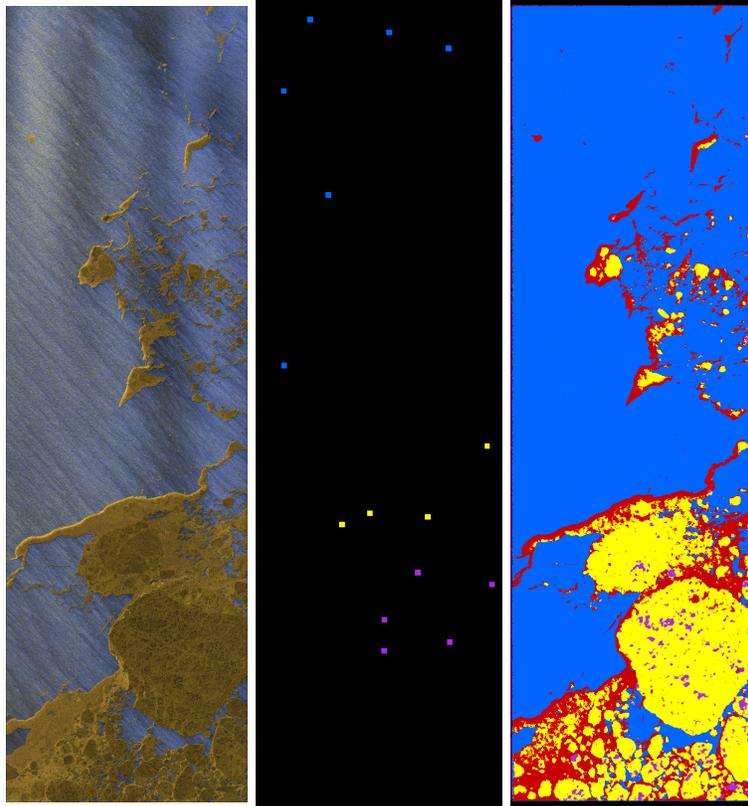


Fig. 9. Results for training image 2014-05-22_B from left to right; Dual polarisation composite SAR image, labelled pixels and classified image.

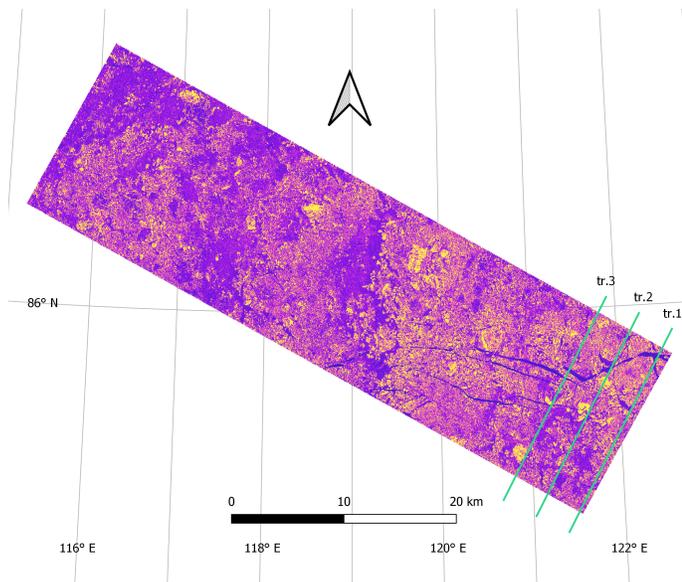


Fig. 10. An overview of the TS-X VV channel image and ICESat-2 collocated strong beam tracks (width not to scale) acquired in November 2019.

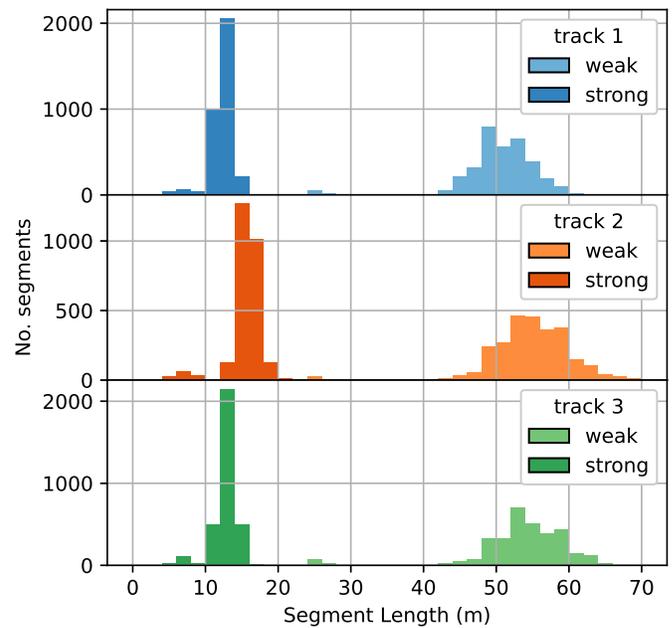


Fig. 11. Histograms of the length of the segments in the ICESat-2 data collocated with the TS-X acquisition from November 2019.

with a much shorter collocation time interval of 0:16:44. A translational correction of 15 pixels across-track has been applied. Note that this is a slightly coarser resolution image with pixel size 5.25 m and resolution 12.1 m.

In general it can be seen that the regions designated as

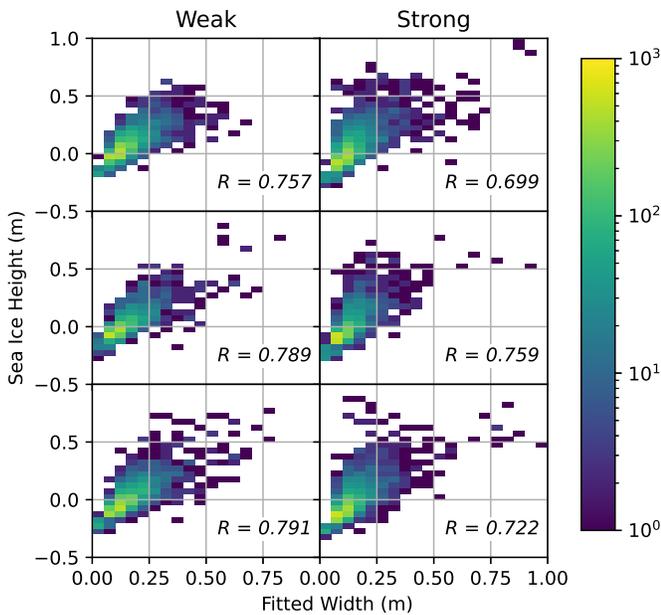


Fig. 12. Histograms of the measured sea ice height versus the width of the Gaussian fit to the photon surface height distribution for all segments in the ICESat-2 data collocated with the TS-X acquisition from November 2019. (top: track 1, middle: track2, bottom: track 3). Inset are the correlation coefficients.

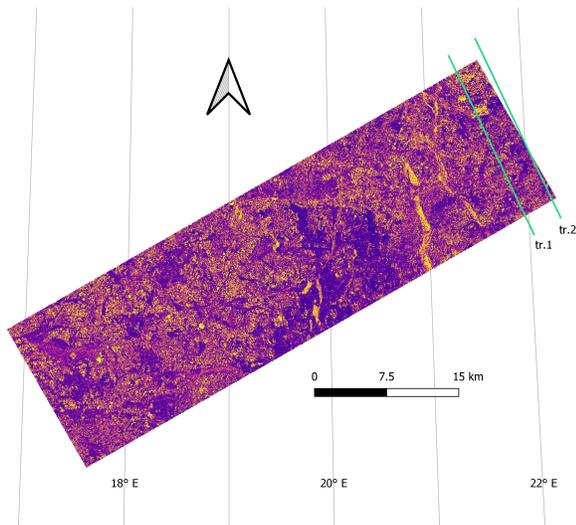


Fig. 13. An overview of the TS-X VV channel image and ICESat-2 collocated strong beam tracks (width not to scale) acquired on 12th April 2020.

smooth ice (purple) in the classification correspond well with the flattest sections of the ICESat-2 tracks. Likewise the rough ice (red) regions match with the sections with the greatest variation in the surface height and overall thickness. The agreement between the SAR backscatter and the ICESat-2 height measurement is particularly good in the data from Nov 2019 where many individual ridges and troughs can be matched at scales down to $\mathcal{O}(200\text{m})$. The surface roughness of the regions designated uniform ice (yellow) is likely at a scale below that fully resolvable with the ATL07 product, hence it appears at least superficially similar to smooth ice.

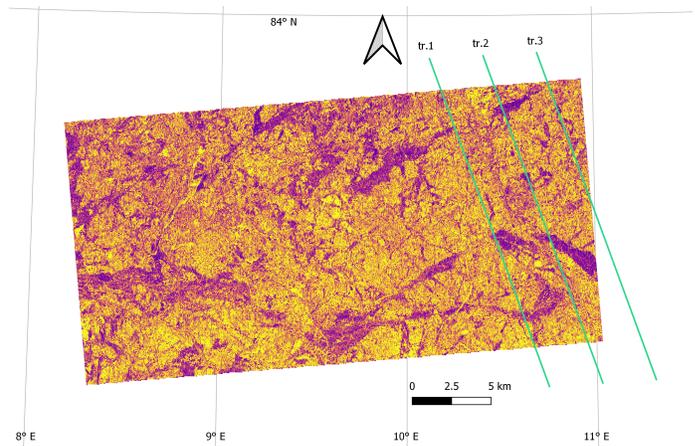


Fig. 14. An overview of the TS-X VV channel image and ICESat-2 collocated strong beam tracks (width not to scale) acquired on 26th April 2020.

Although the spatial resolution for the TS-X images is the same order of magnitude as the ICESat-2 segment lengths, the sensitivity to small-scale surface features is driven by the radar wavelength and radiometric resolution. Resolving this point is worth further investigation.

VI. CONCLUSION

Semi-supervised learning opens up the potential to classify or otherwise analyse remote sensing datasets for which very limited ground truth is available. This is particularly useful as such datasets are likely to include the largest or most up to date. These circumstances clearly apply to sea-ice classification.

The results presented here suggest that AAE based methods can perform credible image segmentation, extrapolating from very limited labelled data. Using this method, new datasets could be more rapidly incorporated into the set of analysed data and provided to users. A source of uncertainty in applying this method for pixel classification is the performance of edge cases, either at physical boundaries in pixel space or at class boundaries in the encoded space. Small changes in the network's perception of these pixels can have a significant impact, particularly on the detection of small scale features. However, the point of transition from intra-class variation to class boundary is a question all classifiers have to address. Compared to earlier studies using similar amounts of labelled data [21], the model is better able to handle novel input images. The reduction in the spatial resolution during the encoding and the remaining classification uncertainty lead to a non-negligible degradation in the quality of the output. There are many developments from the work presented here that could lead to substantial improvements in the future. Using a broader range of input images including seasonal variation would improve the robustness of the classification. More detailed statistical comparisons with data from *in situ* measurements or other independent data sources will better characterise the performance of the network. The "ground truth" used for training could also benefit from real surface measurements instead of relying solely on visual interpretation

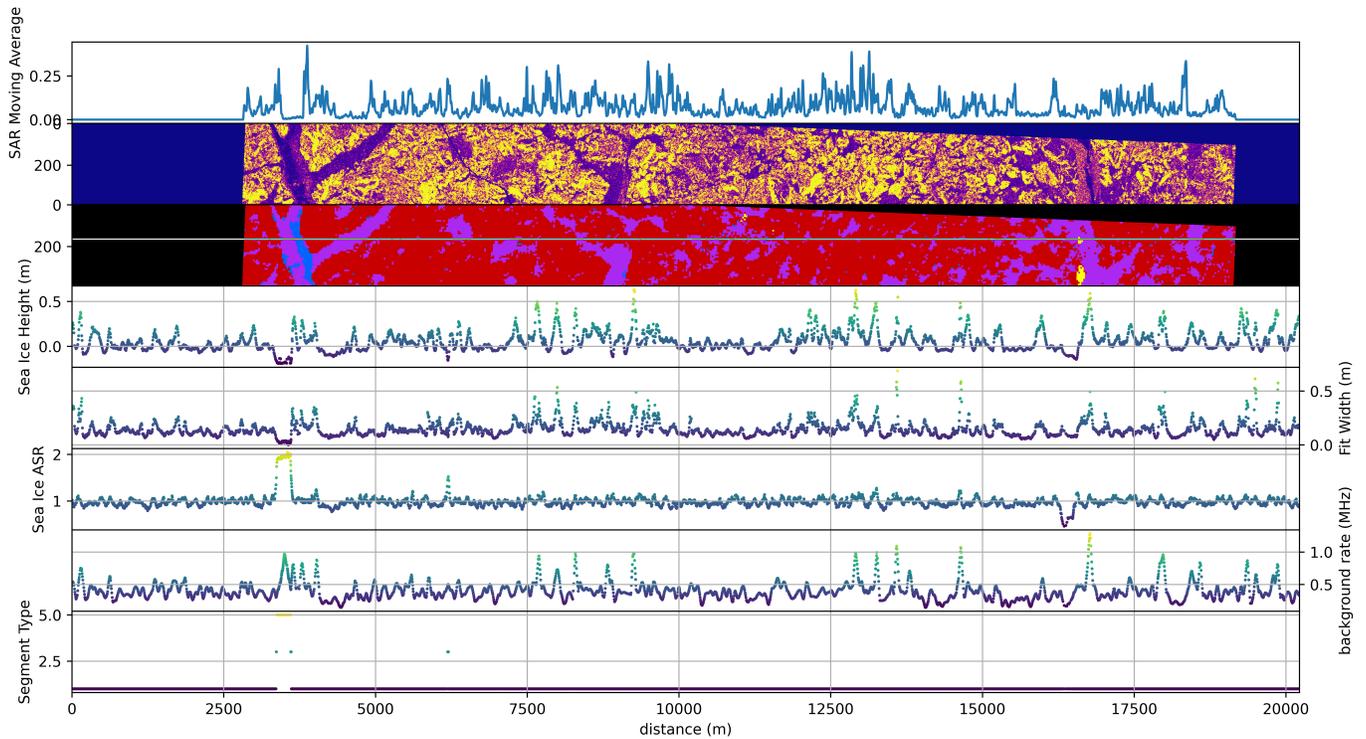


Fig. 15. Collocation of TS-X SM and ICESat-2 track 1 strong beam from 14th Nov 2019. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

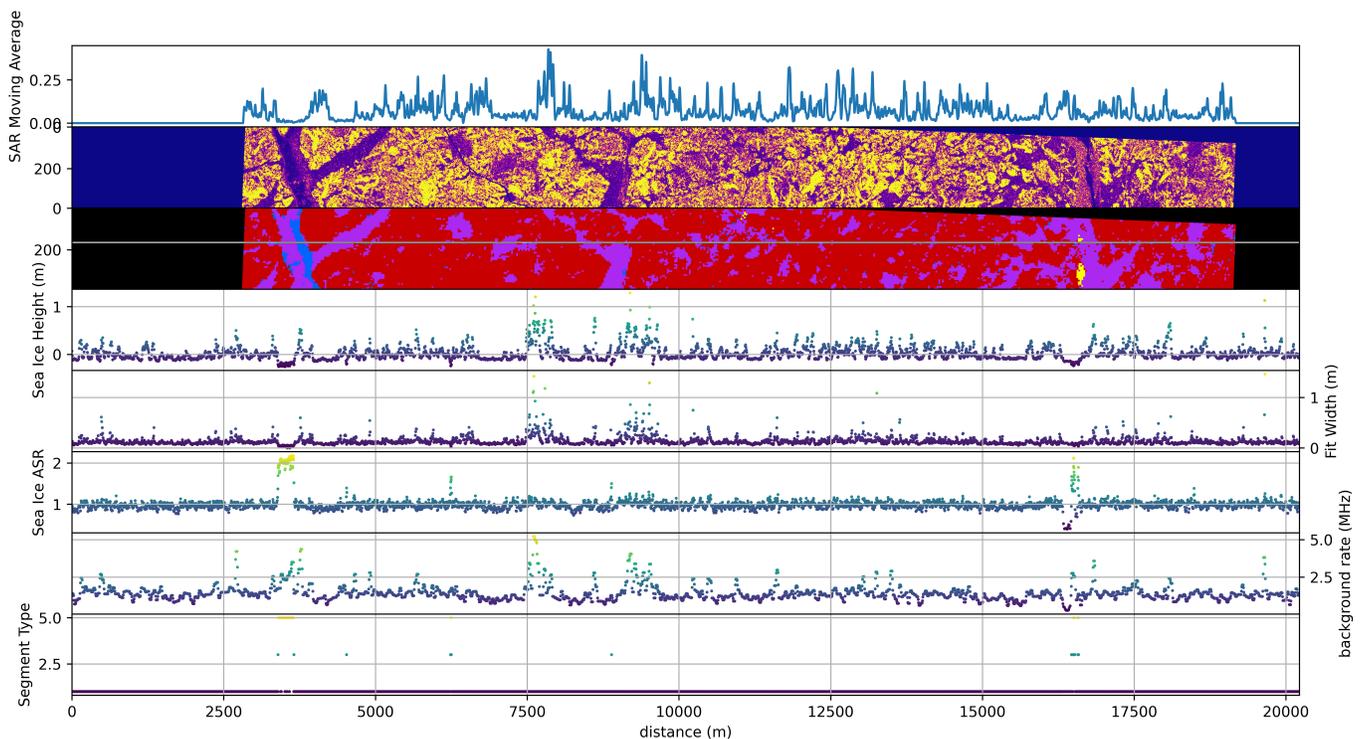


Fig. 16. Collocation of TS-X SM and ICESat-2 track 1 weak beam from 14th Nov 2019. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

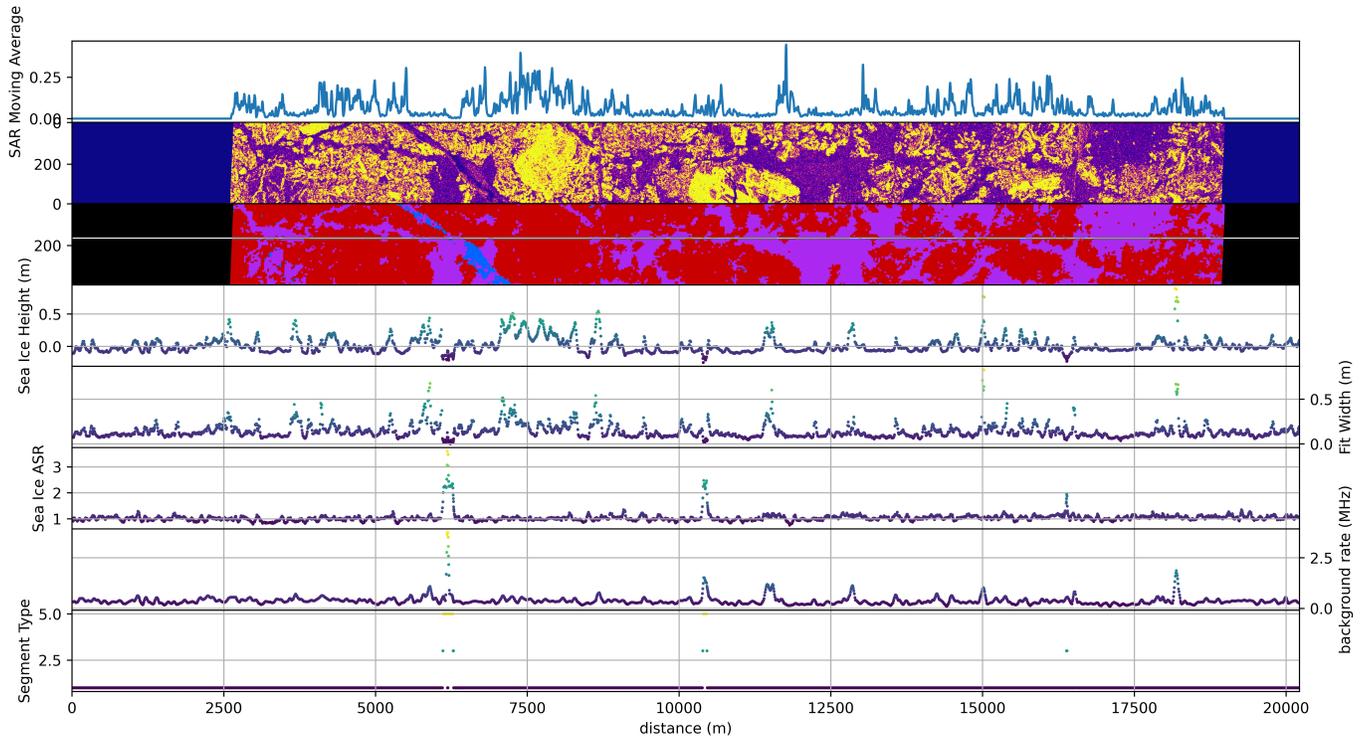


Fig. 17. Collocation of TS-X SM and ICESat-2 track 2 strong beam from 14th Nov 2019. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

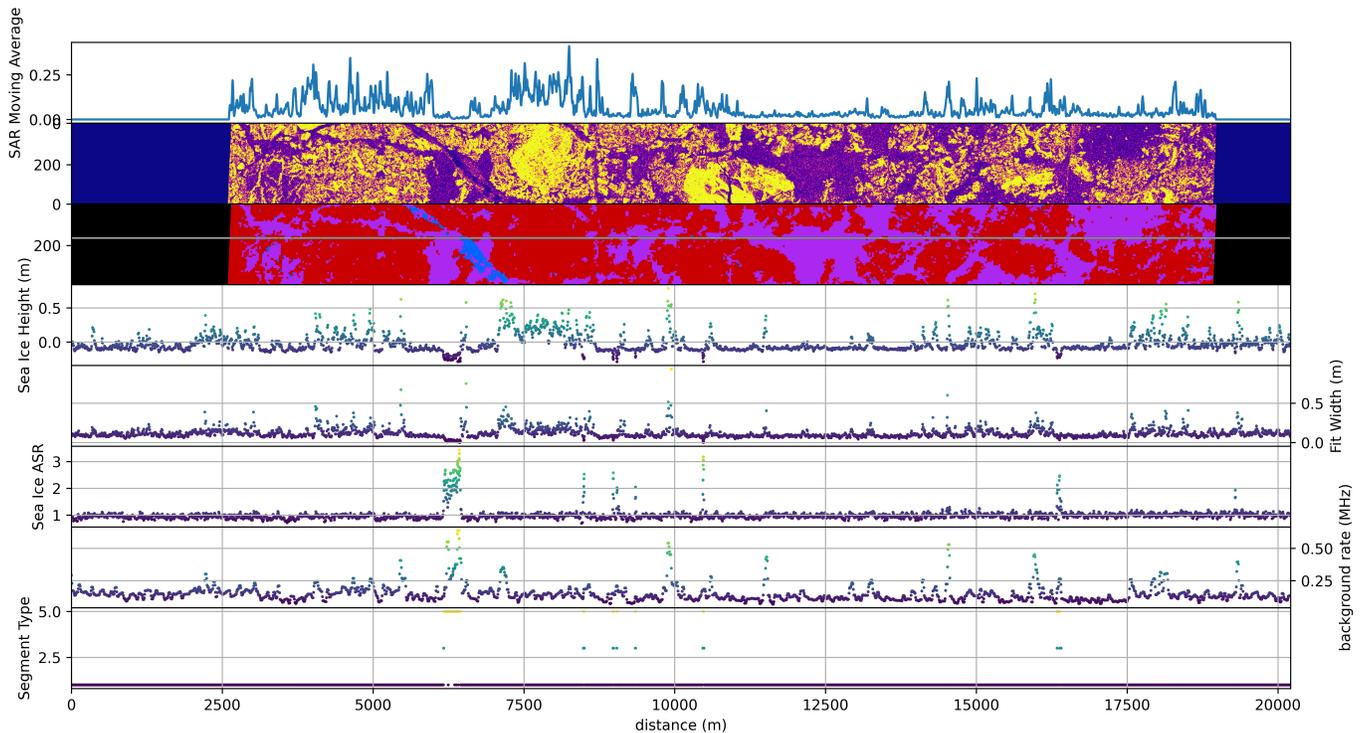


Fig. 18. Collocation of TS-X SM and ICESat-2 track 2 weak beam from 14th Nov 2019. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

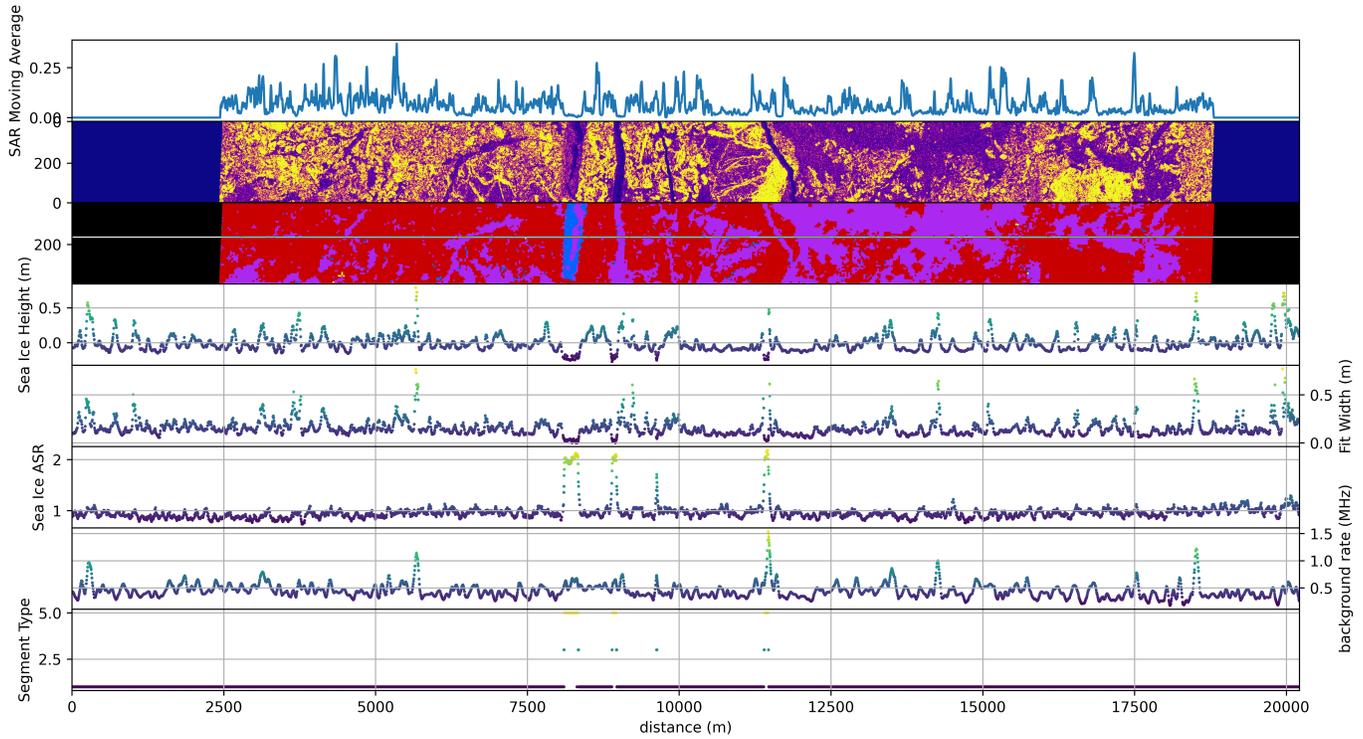


Fig. 19. Collocation of TS-X SM and ICESat-2 track 3 strong beam from 14th Nov 2019. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

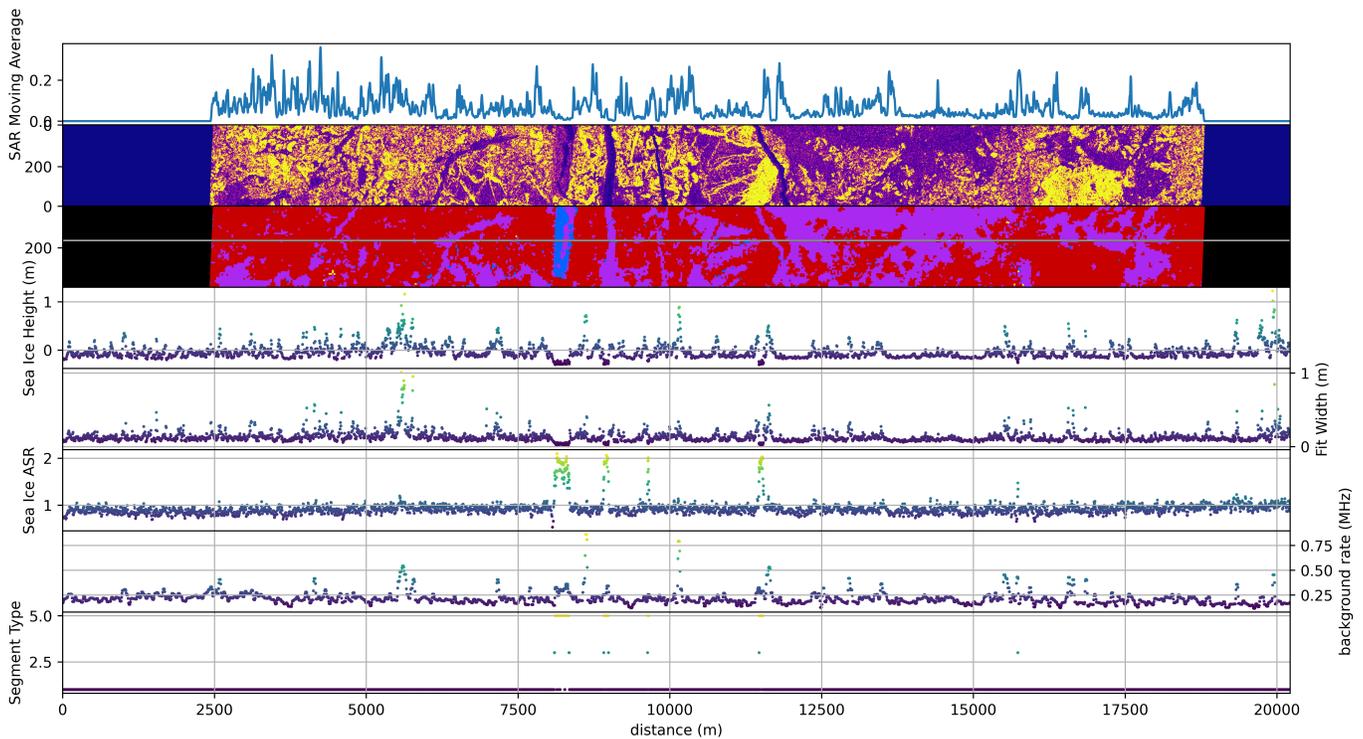


Fig. 20. Collocation of TS-X SM and ICESat-2 track 3 weak beam from 14th Nov 2019. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

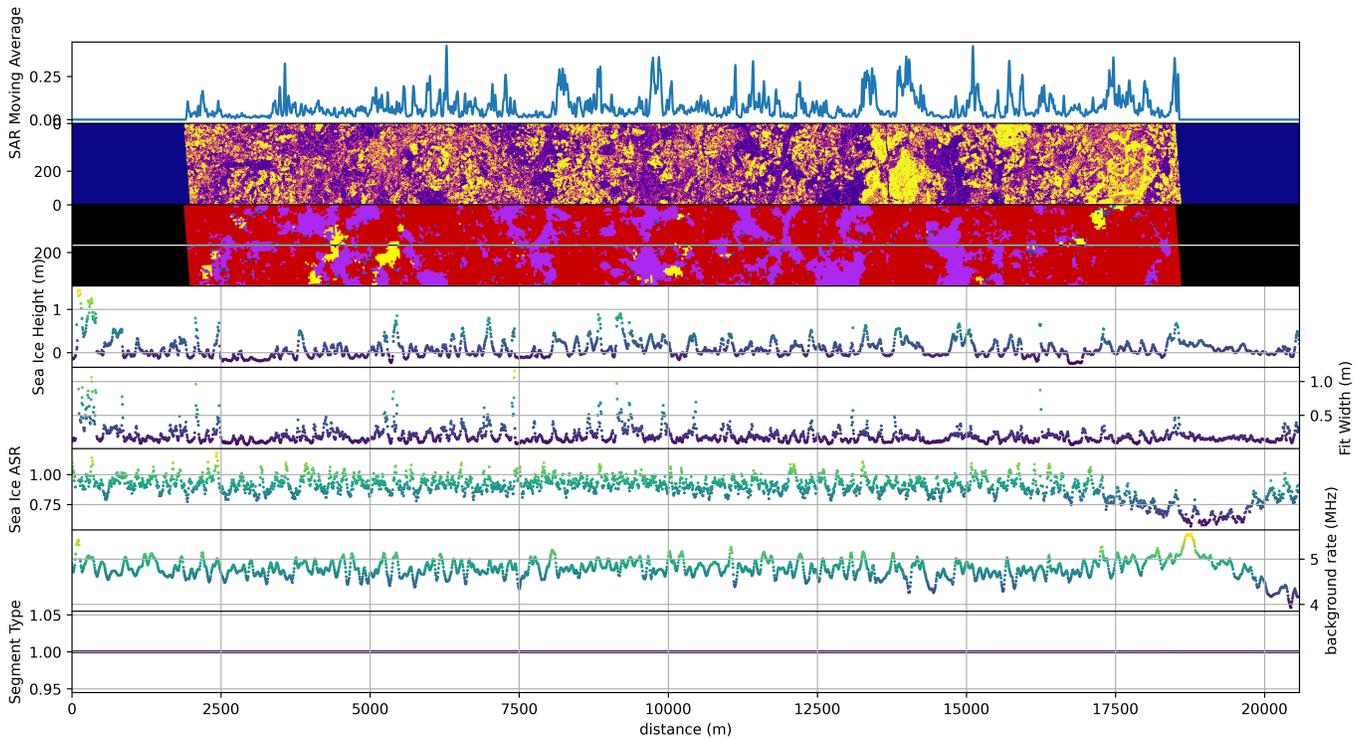


Fig. 21. Collocation of TS-X SM and ICESat-2 track 1 strong beam from 12th April 2020. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

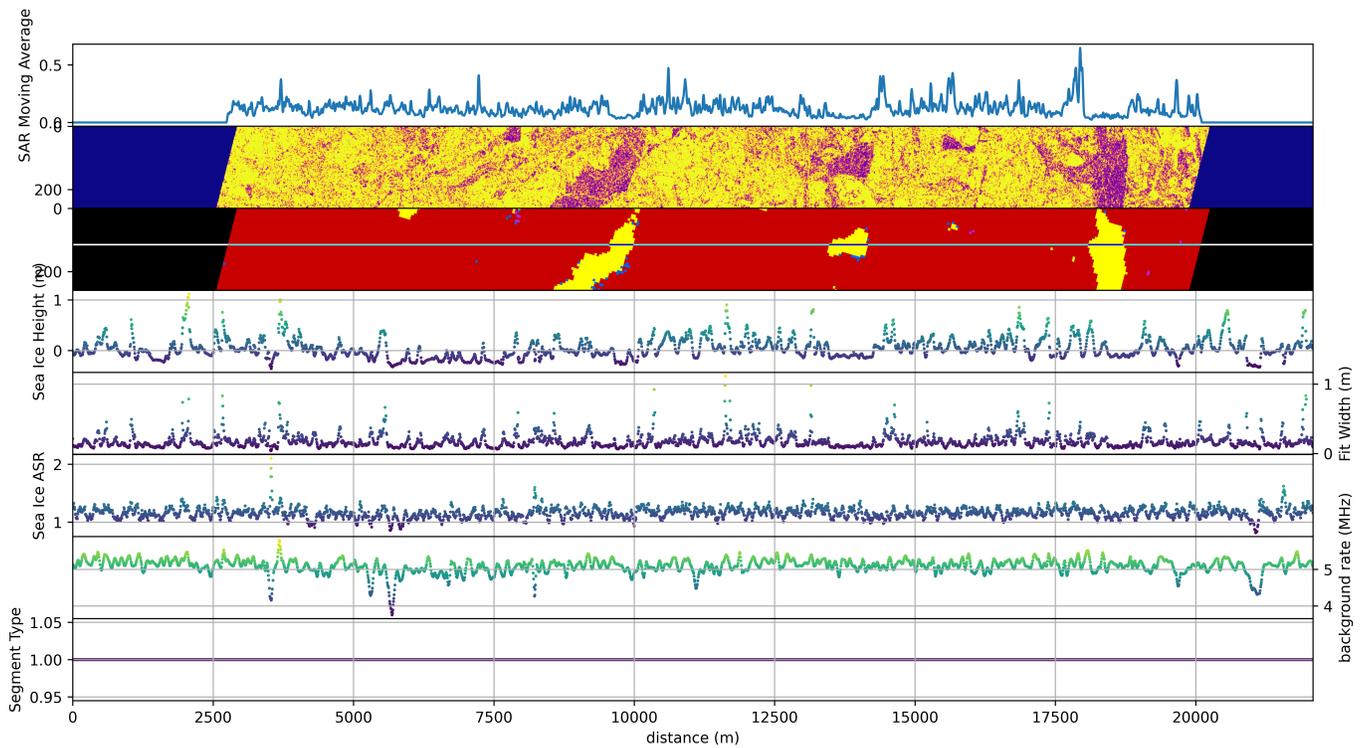


Fig. 22. Collocation of TS-X SM and ICESat-2 track 2 strong beam from 26th April 2020. From top to bottom: TS-X 5by5 pixel moving average along ICESat-2 track, TS-X VV channel sigma0, Classified scene, ICESat-2 sea ice height, ICESat-2 width of the Gaussian fit to the photon scattering surface height distribution, ICESat-2 apparent surface reflectance, ICESat-2 background rate, ICESat-2 surface type.

of the SAR images and will contain more class transition boundaries. The two-dimensional templates used to train the discriminators to guide the form of the latent space will incorporate features or constraints derived from physical models of ice types and their spatial distribution.

APPENDIX A LOSS COEFFICIENTS

Table II presents a summary of the coefficients used in the loss calculation.

TABLE II
A SUMMARY OF THE LOSS CALCULATION COEFFICIENTS FOR THE
AUTO-ENCODING AND SUPERVISED COMPONENTS.

Loss Component	a	b	c
Pixelwise	40	1.25	250
Mean	25	1.6	150
Std dev	20	1.5	250
Skew	10	2.0	250
Encoded elementwise	20	1.5	250

ACKNOWLEDGMENT

The author wishes to acknowledge Dr. Suman Singha for a number of profitable discussions and additional feedback during the writing process. The TS-X images shown in figures 7-9 were acquired using TerraSAR-X AO ressel_OCE2082 and those in figures 10, 13 and 14 were acquired using the TerraSAR-X AO suman_OCE3562_4.

REFERENCES

- [1] A. J. Schweiger, K. R. Wood, and J. Zhang, "Arctic Sea Ice Volume Variability over 1901-2010: A Model-Based Reconstruction," *Journal of Climate*, vol. 32, no. 15, pp. 4731 – 4752, 2019. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/32/15/jcli-d-19-0008.1.xml>
- [2] R. D. C. Mallett, J. C. Stroeve, M. Tsamados, J. C. Landy, R. Willatt, V. Nandan, and G. E. Liston, "Faster decline and higher variability in the sea ice thickness of the marginal arctic seas when accounting for dynamic snow cover," *The Cryosphere*, vol. 15, no. 5, pp. 2429–2450, 2021. [Online]. Available: <https://tc.copernicus.org/articles/15/2429/2021/>
- [3] N. Zakhvatkina, V. Smirnov, and I. Bychkova, "Satellite SAR Data-based Sea Ice Classification: An Overview," *Geosciences*, vol. 9, no. 4, 2019. [Online]. Available: <https://www.mdpi.com/2076-3263/9/4/152>
- [4] R. Ressel, S. Singha, S. Lehner, A. Rösel, and G. Spreen, "Investigation into Different Polarimetric Features for Sea Ice Classification Using X-Band Synthetic Aperture Radar," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 3131–3143, 2016.
- [5] S. Singha, M. Johansson, N. Hughes, S. M. Hvidegaard, and H. Skourup, "Arctic Sea Ice Characterization Using Spaceborne Fully Polarimetric L-, C-, and X-Band SAR With Validation by Airborne Measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3715–3734, 2018.
- [6] A. Bogdanov, S. Sandven, O. Johannessen, V. Alexandrov, and L. Bobylev, "Multisensor Approach to Automated Classification of Sea Ice Image Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 7, pp. 1648–1664, 2005.
- [7] Canadian Ice Service, <https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions.html>, Meteorological Service of Canada.
- [8] Norwegian Ice Service, <https://cryo.met.no/en/ice-service>, Norwegian Meteorological Institute.
- [9] R. Kruk, M. C. Fuller, A. S. Komarov, D. Isleifson, and I. Jeffrey, "Proof of Concept for Sea Ice Stage of Development Classification Using Deep Learning," *Remote Sensing*, vol. 12, no. 15, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/15/2486>

- [10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *arXiv e-prints*, p. arXiv:1406.2661, Jun. 2014.
- [11] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial Autoencoders," *arXiv e-prints*, p. arXiv:1511.05644, Nov. 2015.
- [12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 448–456. [Online]. Available: <http://proceedings.mlr.press/v37/ioffe15.html>
- [13] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, p. arXiv:1412.6980, Dec. 2014.
- [14] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>
- [15] H. Breit, T. Fritz, U. Bals, M. Lachaise, A. Niedermeier, and M. Vonavka, "TerraSAR-X SAR Processing and Products," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 727–740, Feb 2010.
- [16] R. Kwok, A. A. Petty, G. Cunningham, T. Markus, D. Hancock, A. Ivanoff, J. Wimert, M. Bagnardi, N. Kurtz, and the ICESat-2 Science Team, "ATLAS/ICESat-2 L3A Sea Ice Height, Version 4. ATL07." Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2021.
- [17] M. Studinger *et al.*, "IceBridge Narrow Swath ATM L1B Elevation and Return Strength, Version 2," <https://doi.org/10.5067/CXEQS8KVIXEI>, Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center, 2014.
- [18] R. Kwok, S. Kacimi, T. Markus, N. T. Kurtz, M. Studinger, J. G. Sonntag, S. S. Manizade, L. N. Boisvert, and J. P. Harbeck, "ICESat-2 Surface Height and Sea Ice Freeboard Assessed With ATM Lidar Acquisitions From Operation IceBridge," *Geophysical Research Letters*, vol. 46, no. 20, pp. 11 228–11 236, 2019. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084976>
- [19] A. A. Petty, M. Bagnardi, N. T. Kurtz, R. Tilling, S. Fons, T. Armitage, C. Horvat, and R. Kwok, "Assessment of ICESat-2 Sea Ice Surface Classification with Sentinel-2 Imagery: Implications for Freeboard and New Estimates of Lead and Floe Geometry," *Earth and Space Science*, vol. 8, no. 3, p. e2020EA001491, 2021, e2020EA001491 2020EA001491. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EA001491>
- [20] J. Scheick *et al.*, "icepyx: Python tools for obtaining and working with ICESat-2 data," <https://github.com/icesat2py/icepyx>, 2019–.
- [21] S. Singha and J. Imber, "Synthetic Aperture Radar based Sea Ice Type Retrieval using Deep Neural Network," May 2019, ESA Living Planet Symposium.



James Imber received the MSci. degree in Natural Sciences from the University of Cambridge, Cambridge, UK in 2005 and the DPhil. degree in Particle Physics from the University of Oxford, Oxford, UK in 2010. He has held post-doctoral research associate positions at Stony Brook University, NY, USA and Laboratoire Leprince Ringuet, Palaiseau, France. He is currently a researcher at the Remote Sensing Technology Institute, German Space Center (DLR) in Bremen, Germany, focusing on the application of advanced machine learning to extract information from synthetic aperture radar images of the ocean surface.