



Article

Application of a Convolutional Neural Network for the Detection of Sea Ice Leads

Jay P. Hoffman ^{1,*} , Steven A. Ackerman ¹ , Yinghui Liu ² , Jeffrey R. Key ² and Iain L. McConnell ³

- ¹ Cooperative Institute for Meteorological Satellite Studies (CIMSS), University of Wisconsin-Madison, Madison, WI 53706, USA; stevea@ssec.wisc.edu
- ² National Oceanic and Atmospheric Administration (NOAA), Madison, WI 53706, USA; yinghui.liu@noaa.gov (Y.L.); jeff.key@noaa.gov (J.R.K.)
- ³ Space Science and Engineering Center (SSEC), University of Wisconsin-Madison, Madison, WI 53706, USA; iain.mcconnell@wisc.edu
- * Correspondence: jay.hoffman@ssec.wisc.edu; Tel.: +1-608-890-1690

Abstract: Despite accounting for a small fraction of the surface area in the Arctic, long and narrow sea ice fractures, known as “leads”, play a critical role in the energy flux between the ocean and atmosphere. As the volume of sea ice in the Arctic has declined over the past few decades, it is increasingly important to monitor the corresponding changes in sea ice leads. A novel approach has been developed using artificial intelligence (AI) to detect sea ice leads using satellite thermal infrared window data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Visible Infrared Imaging Radiometer Suite (VIIRS). In this new approach, a particular type of convolutional neural network, a U-Net, replaces a series of conventional image processing tests from our legacy algorithm. Results show the new approach has a high detection accuracy with F1 Scores on the order of 0.7. Compared to the legacy algorithm, the new algorithm shows improvement, with more true positives, fewer false positives, fewer false negatives, and better agreement between satellite instruments.

Keywords: leads; sea ice; MODIS; VIIRS; convolutional neural network; U-Net



Citation: Hoffman, J.P.; Ackerman, S.A.; Liu, Y.; Key, J.R.; McConnell, I.L. Application of a Convolutional Neural Network for the Detection of Sea Ice Leads. *Remote Sens.* **2021**, *13*, 4571. <https://doi.org/10.3390/rs13224571>

Academic Editors: Sang-Hoon Hong, Hyun-Cheol Kim, Sang-Wan Kim and Batuhan Osmanoglu

Received: 6 October 2021
Accepted: 11 November 2021
Published: 13 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Leads are narrow, quasi-linear fractures in sea ice that form as a result of divergence or shear within the ice pack. Leads can extend from less than a kilometer to several hundred kilometers in length. Sea ice leads are an important factor in the energy flux between the ocean and atmosphere in the Arctic, particularly in winter. In daylight, leads absorb more solar energy than the surrounding ice due to their lower albedo, warming the water and accelerating melt. At night, without ice cover, the relatively warm sea water radiates heat and moisture into the atmosphere. The turbulent heat flux over sea ice leads can be two orders of magnitude larger than that over the ice surface in winter due to the air-water temperature difference [1]. Though sea ice leads cover a small percentage of the total surface area of the Arctic Ocean [2], the heat fluxes dominate the wintertime Arctic boundary layer heat budget [1,3]. In the winter, spring, and fall, leads impact the local boundary layer structure [4], boundary layer cloud properties [5], thus affecting the surface energy balance. Wind and stresses within the sea ice are the primary factors in the formation of leads [6]. For climate scale studies (on the order of 20 years), trends in lead characteristics [7] help advance our understanding of both thermodynamic and dynamic [8,9] processes in the Arctic.

In recent years, there has been a focus on lead detection using moderate resolution thermal infrared (IR) satellite imagers [10–12]. Sea ice lead characteristics can be derived from satellite passive and active microwave data [8,13], with the primary advantage being that clouds are transparent at microwave wavelengths. However, these platforms either

lack wide spatial coverage or are at such coarse resolution that sea ice fractures are often too narrow to resolve.

Hoffman et al. [11] uses a series of conventional contextual tests (Sobel, Hough, etc.) to detect sea ice leads as areas of high thermal contrast using infrared satellite imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS), which was later adapted to the Visible Infrared Imaging Radiometer Suite (VIIRS). This methodology is referred to as the “legacy” sea ice leads detection algorithm. In recent years, many projects have investigated the application of artificial intelligence (AI) in sea ice detection and classification [14–16]. This study presents a novel approach for the detection of sea ice leads by applying existing AI software [17] toward a new application. In this new approach, spatial information and trainable AI replace the empirically derived detection algorithm tests in the legacy technique [11]. The primary objectives are to improve sea ice leads detection performance and provide continuity between the MODIS and VIIRS sea ice lead detections.

2. Materials and Methods

For lead detection, the relative temperature contrast between water (warm), ice (cold), or clouds (warm or cold), is more important than quantifying the brightness temperature. For this study, thermal imagery is actually a scaled image of radiance counts rather than brightness temperatures (BT). A contrast enhancement filter is used, each image tile is scaled from local minimum to local maximum, and pixel values are converted to an array of bytes (0–255). For computational expediency, and because of the image enhancements, it is not necessary to convert the raw radiance counts into temperature units. Since temperature has diurnal, seasonal, and latitudinal variability, normalization removes the potential for bias based on date or location. The lead detection technique could be applied to the imagery in the native projection (or another applicable projection), but for this study imagery is projected into a standard 1 km resolution EASE-Grid 2.0 projection [18]. The imagery projection also helps reduce the instrument specificity, reduce the possibility for detection bias as a function of scan angle, and allows the method to be more portable to other imagery sources.

A key to the detection method is the development of a “derived truth” mask and training data set. Because of the remoteness of the arctic, ground truth is not readily available. Hand analysis of satellite imagery serves as the best available option for validation. An example of the hand analysis workflow is depicted in Figure 1. A leads mask (red) is an image layer drawn on top of a 11 μm brightness temperature image where leads are identified using a stylus and touch screen in image editing software. The leads mask layer is exported as a standalone image to be used for training. The “truth” mask and training dataset are developed with an iterative approach. In the first iteration, hand analysis was used to draw a leads mask that was readily apparent in a brightness temperature image. Using this hand analysis, a model was trained and tested. For the subsequent iterations, detection results from the previous model iteration are used as a starting point. Image editing software again is used; an eraser tool manually deletes obvious commission errors and drawing tools add obvious omission errors. The masks do not need to be perfect; the detection model is functional even if the training dataset included omission and commission errors. Results improve with each iteration; fewer errors are apparent and less hand analysis is necessary to edit the truth masks as the detection skill increases. Ultimately, a test case dataset with “truth” masks was developed for four days, 1 January, 1 February, 1 March, and 1 April all from 2020. The lead masks are used as a proxy for “ground truth”, representative of leads that are large enough to be readily apparent in both MODIS and VIIRS.

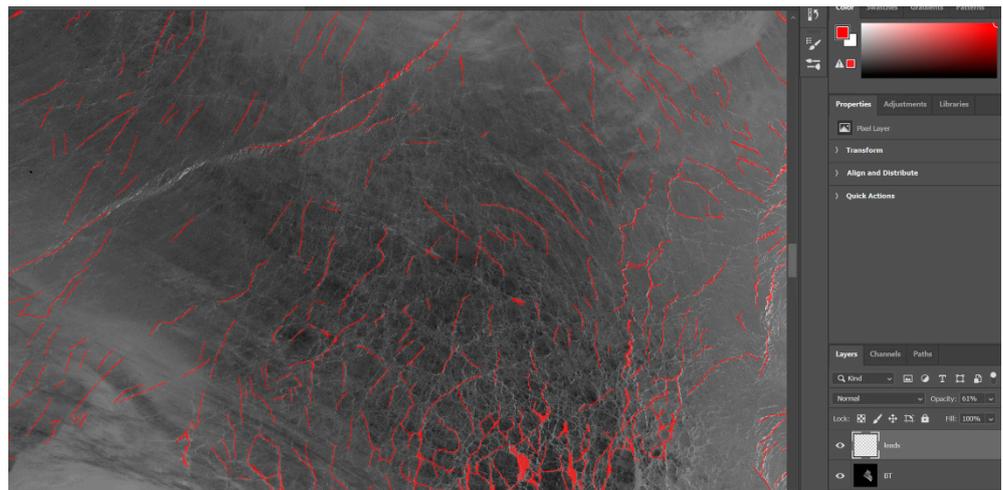


Figure 1. Sample screen capture showing a portion of the VIIRS 11 μm brightness temperature granule from 9 January 2019 at 0342 UTC as a background layer with a leads layer drawn by hand (red) where the brightness temperature pattern indicates a lead is likely.

A particular type of convolutional neural network, a U-Net, was originally described in the literature [19] for the field of microbiology as a way to detect cell walls in microscopic imagery. Despite the vastly different physical properties and scales, the U-Net application in microbiology example in Figure 2 of Ronneberger et al. [19] is visually similar to the U-Net model detection of sea ice leads using satellite imagery example illustrated in Figure 2 of this study. A version of U-Net code, as available in an online code repository [17], has been applied to detect sea ice leads using the 11 μm brightness temperature imagery from MODIS (band 31, AQUA and TERRA imagery, with corresponding geolocation data [20–23] and VIIRS (I-5, SNPP and NOAA-20 imagery and navigation [24–27] north of 60° latitude. A description of the parameters used in the sea ice leads detection U-Net model is described in Table 1; a more complete description of the model is available in Ronneberger et al. [19] and on GitHub [17].

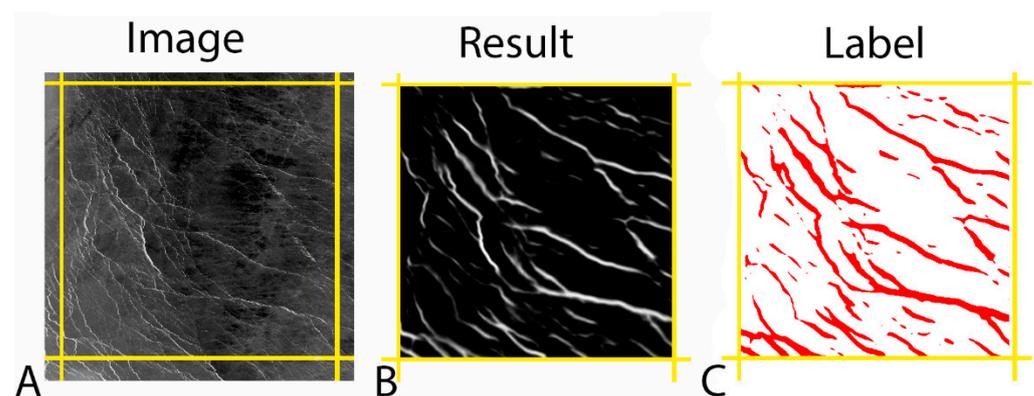


Figure 2. A small 512 km \times 512 km sample of the VIIRS 11 μm brightness temperature image from 0024 UTC on 1 January 2020 (warm features are bright) in panel (A). The corresponding U-Net greyscale image of segmentation results in panel (B). Where the segmentation results are above the detection threshold, the binary lead labels are shown in red in panel (C).

Table 1. Parameters used in U-Net.

Parameter	Value
Activation—inner layers	Relu
Activation—last layer	Sigmoid
Loss function	Binary cross-entropy
Optimizer	Adam
Learning rate	0.0001
Number of epochs	200
Steps per epoch	30
Batch size	2

The U-Net [17] was developed to work for 512×512 pixel imagery. For the application of lead detection, satellite imagery is remapped to an equal area, 1-km resolution EASE-Grid 2.0 [18] and then divided into 512×512 pixel tiles. For testing, the projected imagery from each satellite granule is divided into tiles as shown in Figure 3. The tiles are spaced with 50 pixels of overlap, and the results from the outermost 25 pixel boundary are ignored to avoid detection artifacts around the tile edges. Sea ice lead detection is attempted for each tile containing imagery over ocean water, and the resulting tiles are then reassembled to form a detection mask for each satellite granule.

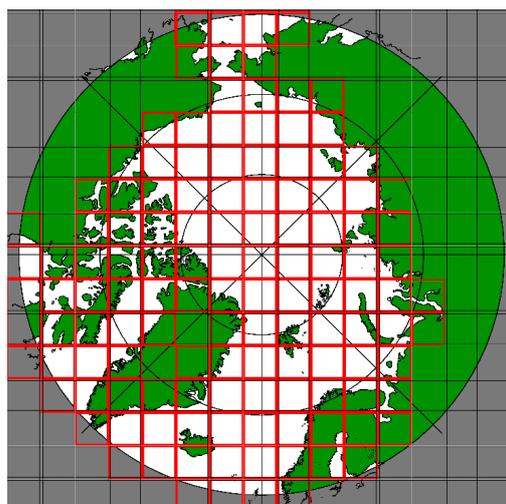


Figure 3. The Arctic processing domain. The squares represent overlapping tiles. Lead detection is attempted north of 60° latitude in tiles that contain at least a portion of ocean water—highlighted in red. The tiles overlap by 50 pixels in all directions; results from the outer 25 pixels are ignored.

For training, 6000 lead mask (label) tiles are generated from the derived-truth masks along with tiles of the corresponding satellite imagery from MODIS and VIIRS on three of the four day case study. To separate testing and training, separate input imagery is used. The models train using brightness imagery averaged over 3-h time-series (multiple satellite overpasses occur over each 3-h period and temperature imagery is averaged at these locations), while the test dataset includes individual granule imagery (no time-averaging). The training imagery spans the polar domain (the domain is on the order of 7000×7000 km, but individual imagery granules only occupy a small proportion of the domain) and is much larger than the 512×512 km/pixel U-Net model domain. To augment the training data, rather than following the tile pattern shown in Figure 3, (used when testing the model), training image tiles are randomly positioned within the much larger polar domain. In this way, an order of magnitude more samples (512×512 km

images) can be generated for training than testing. The same area may appear in multiple training images, but the amount of overlap is random such that lead features do not appear at the exact same position in the training dataset, and the samples positions will be different from those used in testing. To further separate training and testing data—and to avoid overfitting—previously established data augmentation methods are applied to the training data [17]: rotation, width shift, height shift, shear, zoom, and horizontal flip. To avoid edge artifacts, a small buffer region of 50 pixels is used so that there is some overlap in the coverage tiles but the 25 pixels along the edge are discarded (as depicted in yellow in Figure 2).

The above method was used to train two models in TensorFlow 2.2.0. One model trained exclusively with MODIS data, the other with VIIRS data. Each model was trained with 6000 tiles (512×512 pixels): 30 images per step times 200 epochs. As previously developed [19], loss was calculated with the TensorFlow binary crossentropy loss function. The U-Net technique used was previously developed [19] and coded in Python [17]; the intention of this project is only to apply largely off-the-shelf software for a new application—sea ice leads detection.

Each tile is processed by the U-Net, creating a per-pixel lead prediction array of values in the range 0–1. That lead prediction array is scaled and interpreted as an 8-bit image, read as a byte array of U-Net detection score values with pixel brightness ranging from 0 to 255 (as shown in Figure 2B). Each pixel in the scaled leads prediction image is classified as lead or non-lead based on a detection threshold; lead prediction pixels brighter than the threshold are classified as leads (as shown in Figure 2C). Illustrated in Figure 4, the distribution of detection scores overall is bimodal in Figure 4A. In the “Leads” panel, Figure 4B, the majority of lead scores are well above the detection thresholds. Conversely, lead-free locations are depicted in the “Non-Leads” panel, Figure 4C, to illustrate that non-lead detection scores tend to be near zero. The detection thresholds are based on the receiver operating characteristics (ROC) plot in Figure 5. The detection thresholds are defined for each satellite as the ROC curve intersection with the 10% false positive rate line, which is illustrated on the left of Figure 5. On the right of Figure 4, the corresponding false positive and true positive rates are shown as a function of the maximum daily detection score. Here the false positive rate is defined as the false positives divided by the total number of negatives and the true positive rate is the number of true positives divided by the total number of positives. With the detection thresholds at the intersection of the false positive curve with the 10% line, the corresponding MODIS detection threshold is 32/255 and the VIIRS detection threshold is 45/255 (both with a 97% true positive rate).

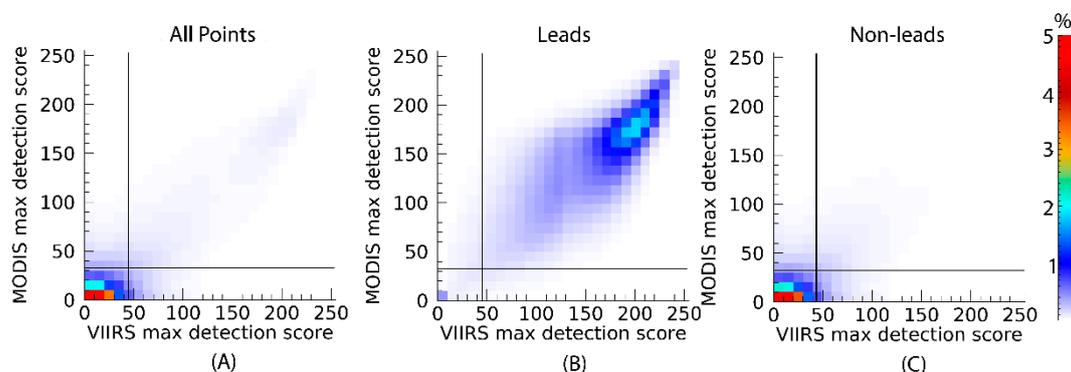


Figure 4. The relative frequency of detection score for all points (A), for leads in the validation mask (B), and for non-lead locations (C) in the validation masks. This is for the aggregate of 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. The detection thresholds are shown as black lines at 45 for VIIRS and 32 for MODIS.

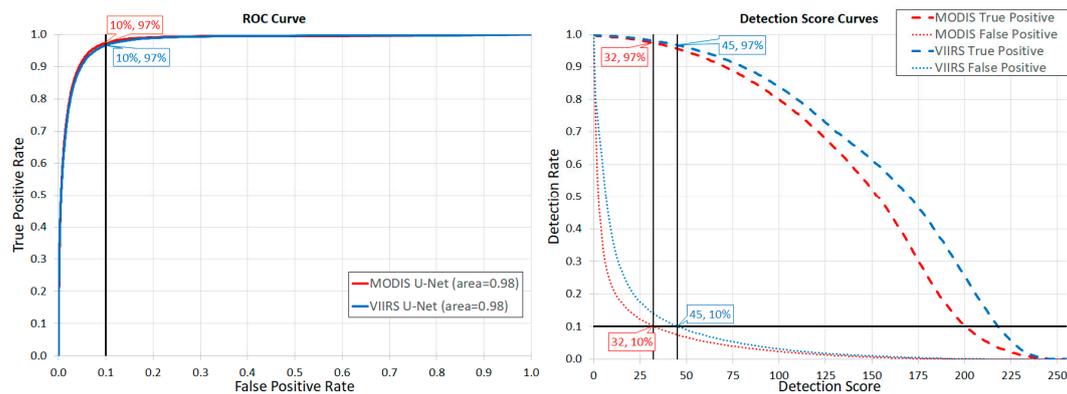


Figure 5. On the left, the receiver operating characteristic (ROC) curve for MODIS (red) and VIIRS (blue) shows the detection threshold at the intersection of the 10% false positive rate (with a true positive rate of 97% for VIIRS and MODIS) is selected as the detection threshold. On the right, the same true positive and false positives curves are shown, but this time as a function of detection score.

The detection process is repeated for each overpass. After all granules are processed, daily composite results are recorded in the corresponding MODIS or VIIRS product file that contains an array with the number of total overpasses, the number of overpasses with a potential lead detection score above the detection threshold, and the maximum detection score at each location. A pixel is considered a positively identified sea ice lead if the U-Net lead detection score is above the threshold in three or more overpasses. While the detection threshold for a potential lead was defined as having a false positive rate of 10% (as illustrated in Figure 4), by using repeat detection criteria, the false positive rate for positively identified leads is reduced by approximately 50%. Validation is performed by comparison of the daily composite results against the hand-derived validation masks.

3. Results

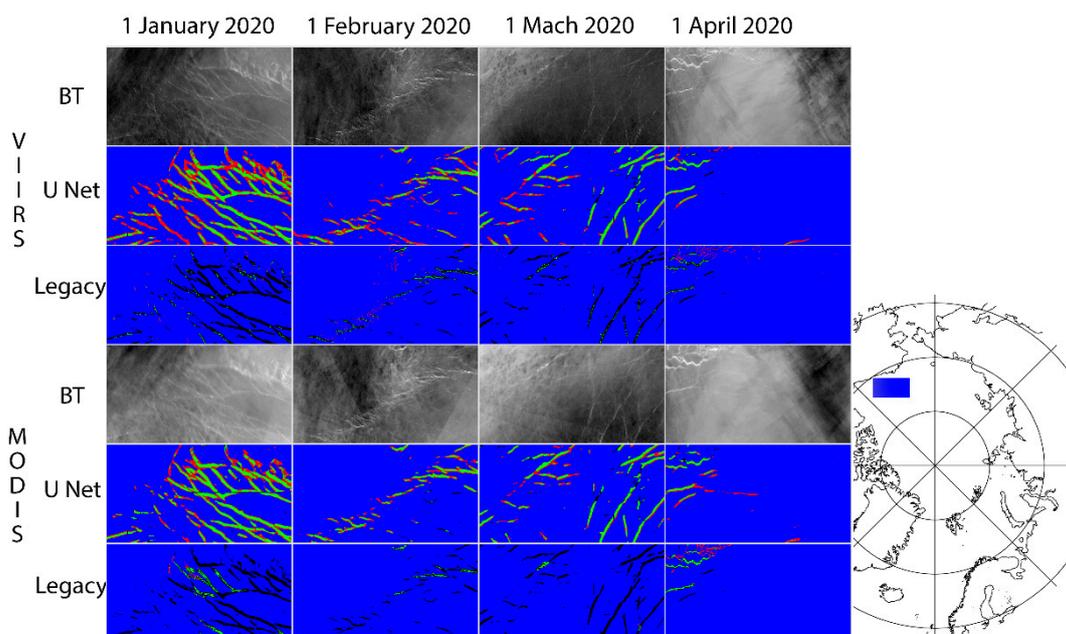
Results from the Beaufort Sea on the first day of January, February, March, and April, 2020 are shown in the multi-panel image in Figure 6 (results from the entire domain are available in the Appendix A). In this region and for the Arctic in general, the overwhelming majority of cases are true negatives or “correct negative” (blue), where validation and detection both identify as not a lead. True positives or “hits” (green) are where the validation mask and detection technique identify a lead; a false positive or “false alarm” (red) is where a lead is detected despite the absence of a lead in the validation mask. A false negative or “miss” (black) is where the validation mask contains a lead that is undetected by the given detection technique. Again, in the absence of ground truth in the Arctic, the masks derived from hand analysis are used as a proxy for validation. There are more true positives (or hits, green) with the U-Net than the legacy product. The legacy product is prone to false negatives (or misses, in black). Through visual inspection of the region shown in Figure 6 and the rest of the Arctic shown in Appendix A, there is significantly more agreement in terms of co-located detections between MODIS and VIIRS when comparing the U-Net products, in contrast to the legacy technique, which does not show strong agreement across satellite platforms. For the April case (far right), persistent cloud cover significantly reduces the amount of detectable leads in the Beaufort Sea. In the other cases, there appears to be some evidence of intermittent cloud contamination, but the U-Net is largely able to detect leads in partly cloudy conditions. In contrast, the legacy lead detection algorithm has a cloud mask dependency, which prevents lead detection in regions flagged as cloudy, and this is a factor in the under-detection of leads in the legacy product.

Table 2. Contingency table, color coded legend for Figure 6 and variable codes for Table 3.

	Validation Mask Lead	Validation Mask Non-Lead (Sea Ice)
Product lead	A (True Positive, “hit”)	B (False Positive, “false alarm”)
Product non-lead (sea ice)	C (False Negative, “miss”)	D (True Negative, “correct negative”)

Table 3. Lead detection statistics for the aggregate results from 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. The percent correct (PC), probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), Hanssen-Kiuper skill score (KSS), and F1 Score are provided with defining equations for variables defined in Table 2. Results are reported for positively identified leads with potential leads in parentheses.

	PC $\frac{A+D}{A+B+C+D}$	POD $\frac{A}{A+C}$	FAR $\frac{B}{A+B}$	CSI $\frac{A}{A+B+C}$	KSS $\frac{AD-BD}{(A+C)(B+D)}$	F1 Score $\frac{A}{A+0.5(B+C)}$
VIIRS: Positive U-Net (Potential)	94% (88%)	93% (97%)	0.50 (0.68)	0.48 (0.32)	0.88 (0.85)	0.65 (0.48)
MODIS: Positive U-Net (Potential)	95% (88%)	92% (97%)	0.46 (0.68)	0.51 (0.32)	0.88 (0.85)	0.68 (0.48)
VIIRS: Legacy (Potential)	94% (93%)	15% (18%)	0.63 (0.67)	0.12 (0.13)	0.14 (0.15)	0.22 (0.23)
MODIS: Legacy (Potential)	94% (88%)	16% (30%)	0.66 (0.82)	0.12 (0.13)	0.14 (0.22)	0.22 (0.23)

**Figure 6.** The VIIRS (top) and MODIS 11 μm composite brightness temperature imagery and leads product results from (left to right) 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. The color table is detailed in Table 2, blue is true negative, green is true positive, red is false positive, and black is false negative in comparison with the validation mask.

To avoid errant lead detections in open water, a post-processing step is used to filter the results by a mask of sea ice concentration [28] so that sea ice leads are only considered within the largest expanse of continuous sea ice within the Arctic Ocean. The resulting statistics are given in Table 3. Because of the imbalanced nature of lead distributions—the proportion of correct predictions (PC) is largely an artifact of the prevalence of true negatives, and not necessarily an indication of detection skill. The more pertinent detection metrics are reported and defined in Table 3, the probability of detection (POD, which is also known as Recall), False alarm ratio (FAR), critical success index (CSI, which also known as Intersection over Union or IoU), Hanssen-Kiuper skill score (KSS), and F1 Score. Table 2 defines these metrics and reports the results for positively identified leads and potential leads. The F1 Score is calculated as $2 \times ((\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}))$. It is also

known as the F Score or F Measure. Precision is the number of true positives divided by the sum of true positives plus false positives, and recall or POD is the number of true positives divided by the sum of true positives plus false negatives. For completeness, a full suite of statistics is provided, however the F1 Score and CSI (or IoU) are the most significant statistics for this application because both give weight to true positives, false positives, and false negatives, without giving weight to the most common true negative category.

The detection criteria are different between the legacy and U-Net approach, but the general concept between a potential lead and positively identified lead is similar. In the legacy algorithm, a potential lead is defined as any area with thermal characteristics of a lead (primarily high thermal contrast relative to surrounding pixels), and a positively identified lead is the subset of potential leads that pass a series of conventional tests (Sobel, Hough, etc.) [11]. For the U-Net leads, a potential lead is any location with a U-Net score above the detection threshold, while a positively identified lead is in the subset of potential lead where the detection occurred in three or more overpasses within the same day. Depending on the end user's application—sensitivity to omission or commission errors, potential leads generally maximize the true positives while positively identified leads minimize the false positives. As a result, potential leads tend to have higher POD and KSS scores, but positively identified leads have higher PC, CSI (IoU), FAR, and F1 scores. In comparison to the legacy technique, the U-Net technique has on the order of six times the POD and four times the skill scores (CSI, KSS, and F1), while also showing a small reduction in false detections (FAR). It is also significant to highlight the skill improvement attained through the repeat detection criteria used to elevate a potential lead to a positively identified lead. The F1 Score for a U-Net positively identified lead shows more skill compared against a potential lead, whereas there is no F1 Score improvement between potential and positively identified leads in the legacy technique. Another measure of detection skill is illustrated in the precision recall curves in Figure 7 where both MODIS and VIIRS demonstrate significant detection skill with nearly identical curves and area under the curve equal to 0.76.

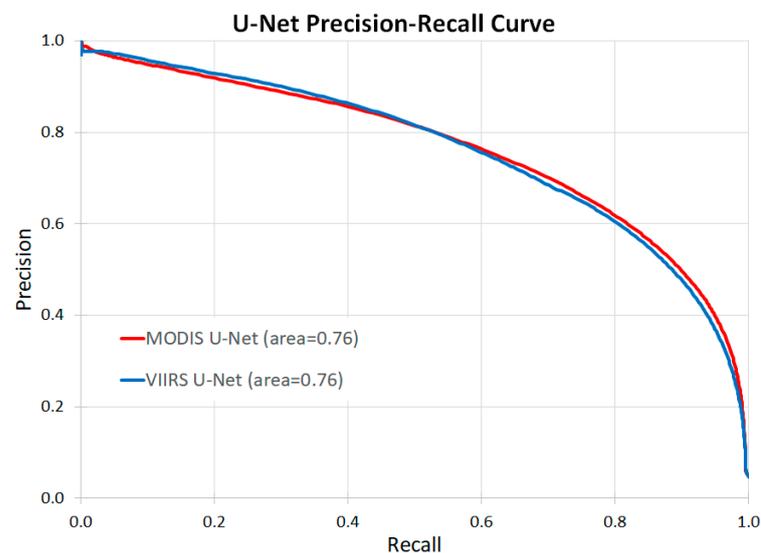


Figure 7. Precision-Recall curves for MODIS (red) and VIIRS (blue) aggregated from 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. Recall, known as POD, is defined as $A/(A + C)$ and precision is $A/(A + B)$ from Table 2.

Not only does the U-Net technique show significant improvement in detection skill, one of the primary motivations to develop a new sea ice leads detection technique was to achieve better cross-satellite product agreement than was being achieved with the legacy product. There should be wide agreement between both satellites; only minor differences between satellites would be attributed to differences in overpass time, spatial resolution,

and to a lesser degree, scan angle and spectral differences. An analysis of the co-location of detections is shown in Figure 8. Again, the aggregate of lead detections by satellite and detection method from 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020, are grouped based on if the lead detection corresponds with another lead detection and/or with a lead in the validation mask. The majority of the U-Net detections are co-located with a valid lead (green and blue). In contrast, a larger proportion of the leads in the legacy product are not co-located with another product detection (black). Among the detections that are not co-located with another product detection, we would suspect that leads detected by multiple products but not the validation mask (blue) are most likely an artifact of an imperfect validation mask (omissions in the validation mask), which is consistent with a visual inspection of the masks. In contrast, the false positives without a co-located lead detection (black) are more likely a mixture of false detections and omissions from the validation mask. The significance is that the false detections that are in only one product (red) make up a much smaller proportion of the U-Net lead detections than the legacy lead detections. Moreover, visual inspection of the result masks is consistent with this interpretation, where many of the “false” leads detected by U-Net appear to be omission errors in the validation masks, while cloud edge artifacts appear to be the likely cause of the false detections visually apparent in the legacy product. Overall, the U-Net technique is a significant improvement over the legacy technique with more true positives, false positives that tend to look like true leads (likely validation mask omissions), and fewer false negatives.

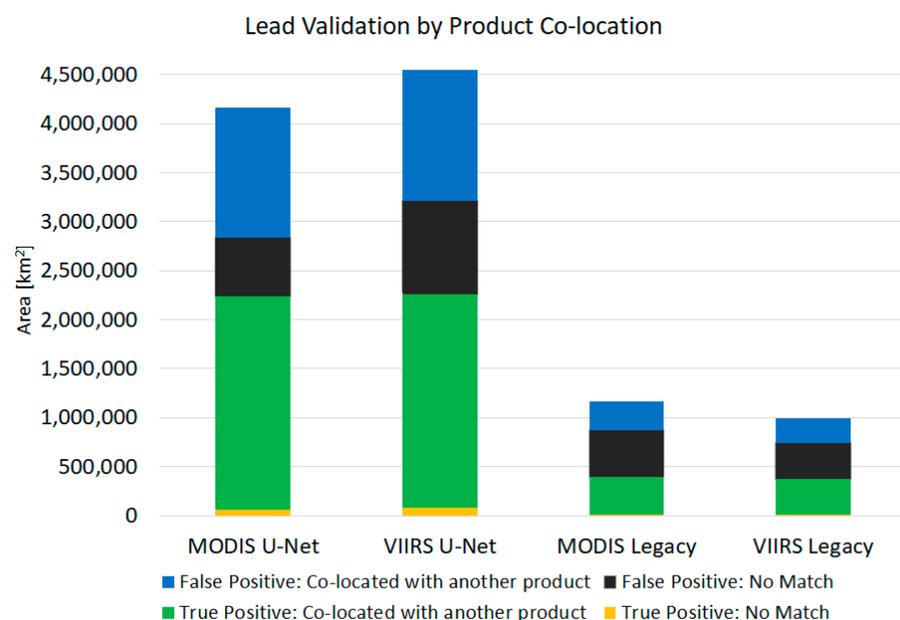


Figure 8. The aggregate co-location of true positive and false positive lead detection by satellite and detection method from 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. True positives are shown where a lead detection corresponds with (green) and without (yellow) co-location of a lead detection from another product. False positives are also shown where a lead detection corresponds with (blue) and without (black) co-location of a lead detection.

Using a fairly limited dataset with only four days of hand analyzed validation masks, three of the four day validation sets were used to train the model; the remaining day was withheld for testing and validation. The detection performance from the three days within the training dataset are indistinguishable from the fourth day that was withheld from the training dataset. To further ensure results are not overfit to the truth masks, a test case, spanning January through April of 2020 and composite results are illustrated in comparison with the legacy product in Figure 9. Beyond the four-day case study, validation data are limited, so it is not possible to perform the same rigorous detection skill analysis as

presented for the cases where hand-analyzed validation masks exist. However, results from the winter of 2020 confirm that the U-Net does provide more consistency between MODIS and VIIRS than in the legacy product. Given the spatial patterns of the aggregate results, we infer that the U-Net appears to have many more true positives (detection features that look like leads) and fewer false positives (fewer small features that look like cloud edge artifacts) than the legacy product. A more rigorous study of lead detection over a long time series will be the focus of upcoming work.

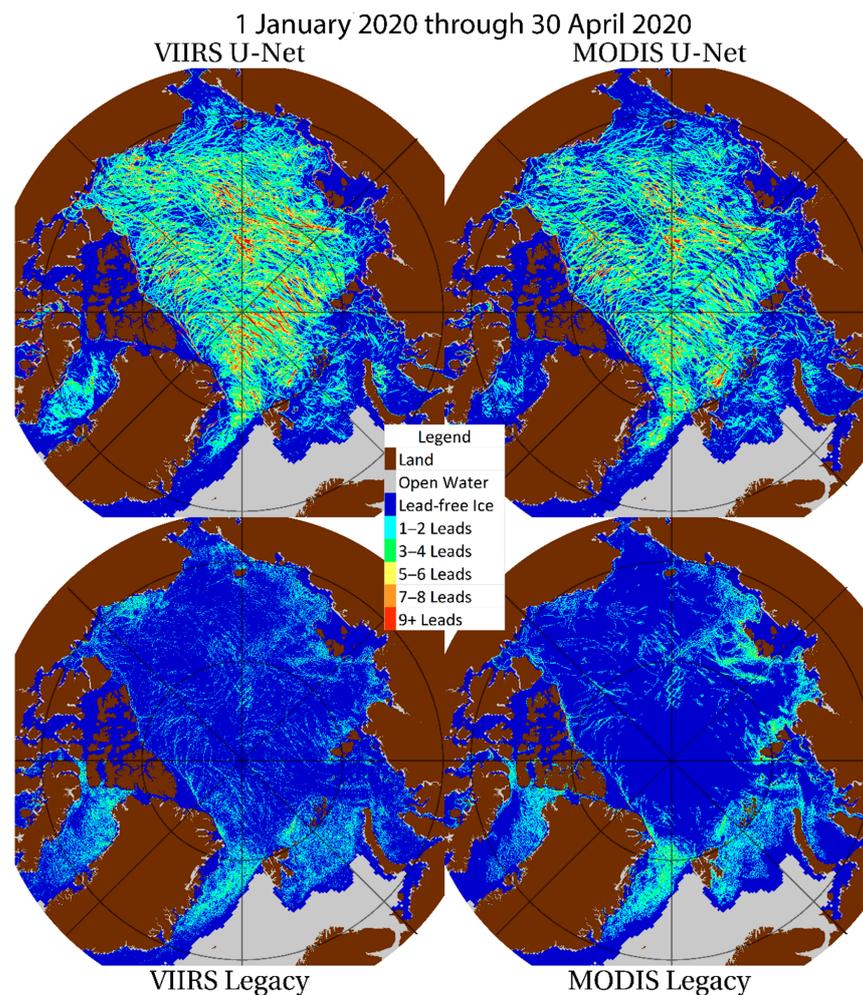


Figure 9. The aggregate of lead detections from 1 January 2020 through 30 April 2020. The U-Net detects more leads, shows better agreement across satellites, and appears to perform well for cases beyond the training dataset.

4. Discussion

The premise is that the more often a potential lead occurs during the course of a day, the more likely it is that the potential lead is a true lead, whereas fewer repeat observations tend to be related to false positives, for example, short lived features such as a cloud edge. Figure 10 illustrates detection skill inferred by the number of repeat observations of a potential lead. The majority of leads are observed in more than four overpasses, and in the right, non-leads tend to have a lead detection score above the detection threshold in less than three overpasses.

We anticipate that expanding the validation set and retraining the model in future work will likely continue to improve detection metrics. Similarly, future work could retrain the U-Net based on validation datasets from higher resolution instruments such as Synthetic Aperture Radar [29] or Sentinel-2 [30], compared against other products,

and/or applied to other historical, current, or future satellite imagery while also exploring other visible, thermal, and microwave bands. As an example, a comparison has been performed of the satellite altimeter based ICESat-2 sea ice leads product [31] against the U-Net and legacy thermal IR-based lead products. Given the very narrow spatial coverage (approximately 10-m beam width) and 91 day coverage cycle from ICESat-2, it would be difficult to derive validation imagery that could be used to train IR-based lead detections. We would expect to see significant differences in the amount of leads that can be detected on different spatial scales, for moderate resolution IR imagers, we would not expect to detect leads narrower than approximately 250 m [32–34]. However, it is possible to identify the frequency of an IR-based lead co-located with a detection from the altimeter, and these results are summarized in Table 4. Again, given the spatial resolution of the ICESat-2 ATLAS laser measurements are two orders of magnitude smaller than the 1-km resolution of the IR leads product, we would expect a significant proportion of sea ice leads detected by ICESat-2 will be too small for detection with moderate resolution IR. However, what can be observed in the comparisons against ICESat-2 is that the U-Net technique detects more leads than the legacy technique. Moreover, with the U-Net there is significantly better agreement between MODIS and VIIRS; only a small proportion of U-Net lead detections are detected by one but not both satellites, whereas the legacy product shows relatively poor agreement between satellites. Overall, the comparisons with ICESat-2 are consistent with the findings from the IR-based validation masks. At present the U-Net performs well detecting moderate to large-scale leads; further research would be necessary to detect small scale, sub-resolution leads. For example, sub-resolution leads would be more readily detected using a 4–11 μm brightness temperature difference.

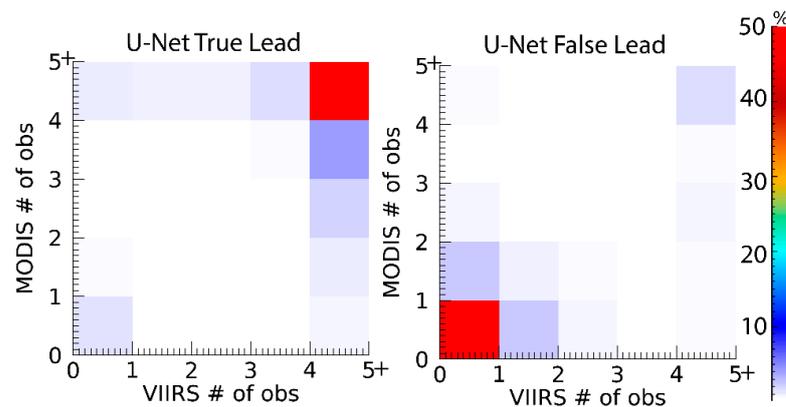


Figure 10. The number of observations of a potential lead by satellite and detection method compared against the aggregate of validation data from 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. Notice the U-Net true detections tend to have a high rate of observations. Similarly, for false detections, the U-Net has a low rate of repeat observations (and would therefore fail to meet the conditions to be classified as a lead).

Table 4. ICESat-2 co-location with moderate resolution lead detection; aggregate results from 1 January 2020, 1 February 2020, 1 March 2020, and 1 April 2020. Results are reported as the percentage of ICESat-2 leads with a lead detection within 1 km of a MODIS and/or VIIRS lead detection. Positively identified leads are reported first, with potential leads in parentheses.

	U-Net	Legacy
MODIS	19% (27%)	8% (17%)
VIIRS	21% (28%)	7% (8%)
MODIS or VIIRS	23% (32%)	13% (21%)
MODIS and VIIRS	17% (23%)	2% (5%)

Some differences will be expected across satellite platforms due to differences in orbital patterns (overpass times) where leads may move or change, clouds may obscure the leads, etc. Further, due to spatial differences in the instruments, we would expect the coarser resolution MODIS data to have a bias toward detecting wider leads while VIIRS, with finer resolution, would detect narrower leads that may be too small to have a detectable signal in MODIS. Comparing the MODIS and VIIRS detection score densities illustrated in Figure 4, there is a slight bias where the VIIRS score tends to be slightly higher than MODIS, and this is also reflected in the slightly different detection thresholds.

5. Conclusions

The nature of AI is that the quality of the results are limited by the quality of the training datasets. In the application presented here — detecting sea ice leads (fractures) in satellite imagery — developing quality validation may be the limiting factor due to the laborious and imperfect hand analysis. However, in some ways the qualitative detection skill may exceed quantitative detection skill; i.e., visual inspection of the results shows that valid leads are detected in areas that the validation mask flags as lead-free. As work progresses, adding more cases to the validation set should help minimize omission and commission errors. Going forward, comparisons against time-matched imagery from higher resolution imagery or lead masks from other sources could also be a good way to validate results. However, using the existing validation dataset has shown that the U-Net architecture can successfully be applied for the detection of sea ice leads from moderate resolution satellite-based thermal imagery. Results demonstrate a high level of detection skill, and improvement of the legacy technique, with more true positives, few false positives, fewer false negatives, and better agreement between satellite instruments.

Ultimately, work will continue toward the application of this U-Net technique to the entire MODIS and VIIRS time-series. From this archive, in excess of 20 years, a rigorous analysis of the characteristics of sea ice leads will be studied. By establishing separate models for MODIS and VIIRS, the intention is to be able to provide a stable leads product over the entire satellite archive. As MODIS nears end of life, the VIIRS product will be able to continue to provide leads detection products for years to come. Furthermore, the initial work has been focused on the detection of leads in the winter season; it may be necessary to expand the training dataset to include warm season cases (that include warm season clouds and melt ponds) before the technique could be applied to the summer season.

Author Contributions: Conceptualization, J.P.H., S.A.A., Y.L. and J.R.K.; methodology, J.P.H. and I.L.M.; software, J.P.H., Y.L. and I.L.M.; validation, J.P.H. formal analysis, J.P.H.; investigation, J.P.H.; resources, J.P.H., Y.L. and I.L.M.; data curation, J.P.H. and Y.L.; writing—original draft preparation, J.P.H.; writing—review and editing, J.P.H., S.A.A., Y.L., J.R.K. and I.L.M.; visualization, J.P.H.; supervision, S.A.A.; project administration, S.A.A.; funding acquisition, S.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA, grant number 80NSSC18K0786.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The U-Net model files, example code, validation masks, and product results used in this paper are available via anonymous ftp at <ftp://frostbite.ssec.wisc.edu/unet> (accessed on 10 November 2021).

Acknowledgments: The Aqua and Terra MODIS, and SNPP and NOAA-20 VIIRS datasets were acquired from the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC), located in the Goddard Space Flight Center in Greenbelt, Maryland (<https://ladsweb.nascom.nasa.gov/> (accessed on 10 November 2021)) and the Atmosphere SIPS located at the University of Wisconsin–Madison (<https://sips.ssec.wisc.edu> (accessed on 10 November 2021)). The code for this project is based off the repository at [19] <https://github.com/zhixuhao/unet> (accessed on 10 November 2021). A portion of this work was supported by the SSEC2022 project

at the Space Science and Engineering Center at UW-Madison. The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. Government position, policy, or decision.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The results from the full masks from the case study dates are shown in Figures A1–A4. True negatives are in blue, true positives are green, false positives are red, and false negatives are black. Land is shown as brown, and open water, determined by being outside of the largest continuous region of sea ice [24] is shown as gray.

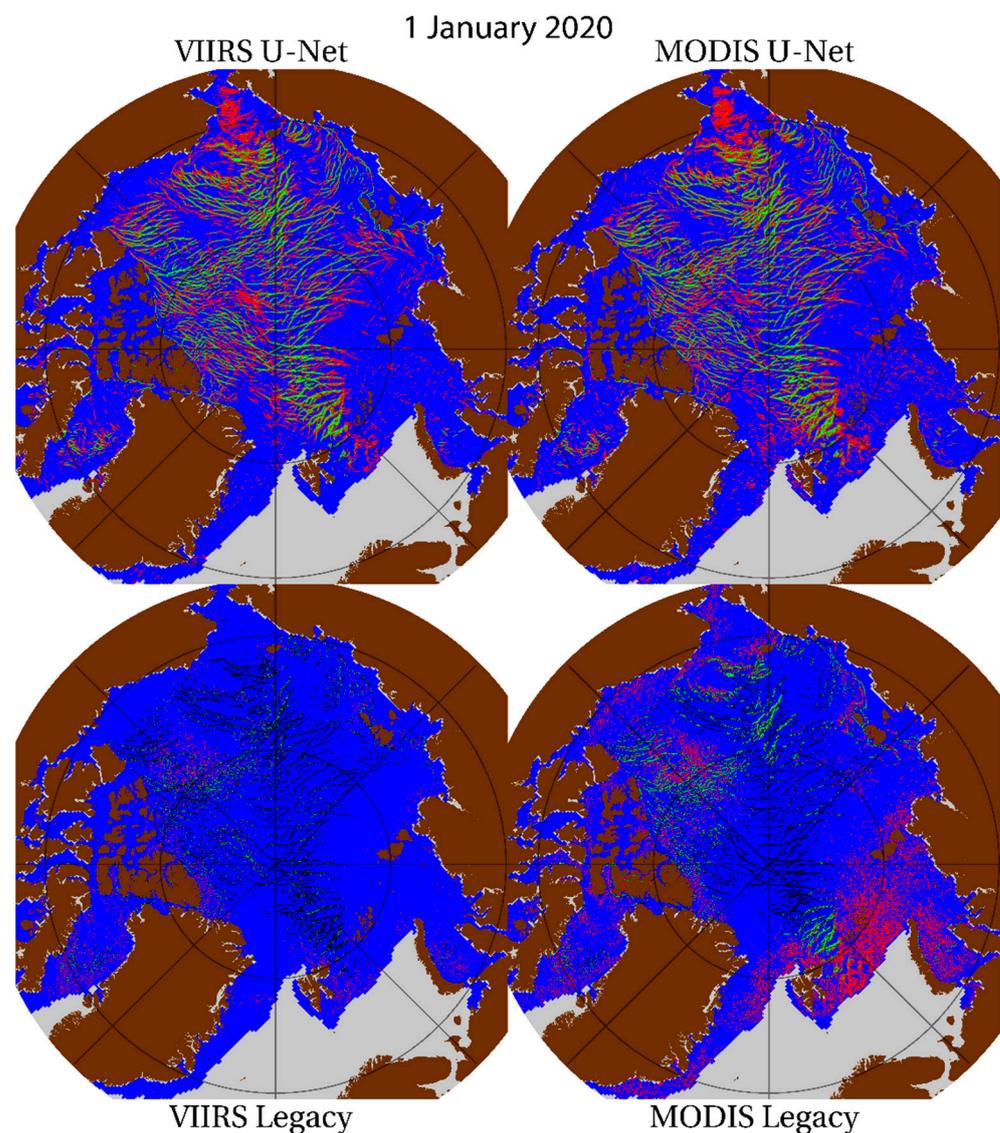


Figure A1. The full results from which a subset was shown in Figure 6. VIIRS (left) and MODIS (right) leads product results from U-Net (top) and Legacy product (bottom) on 1 January 2020. The color table is detailed in Table 2, blue is true negative, true positive is green true positive, red is false positive, and black false negative in comparison with the validation mask. Land is shown in brown and gray depicts open water.

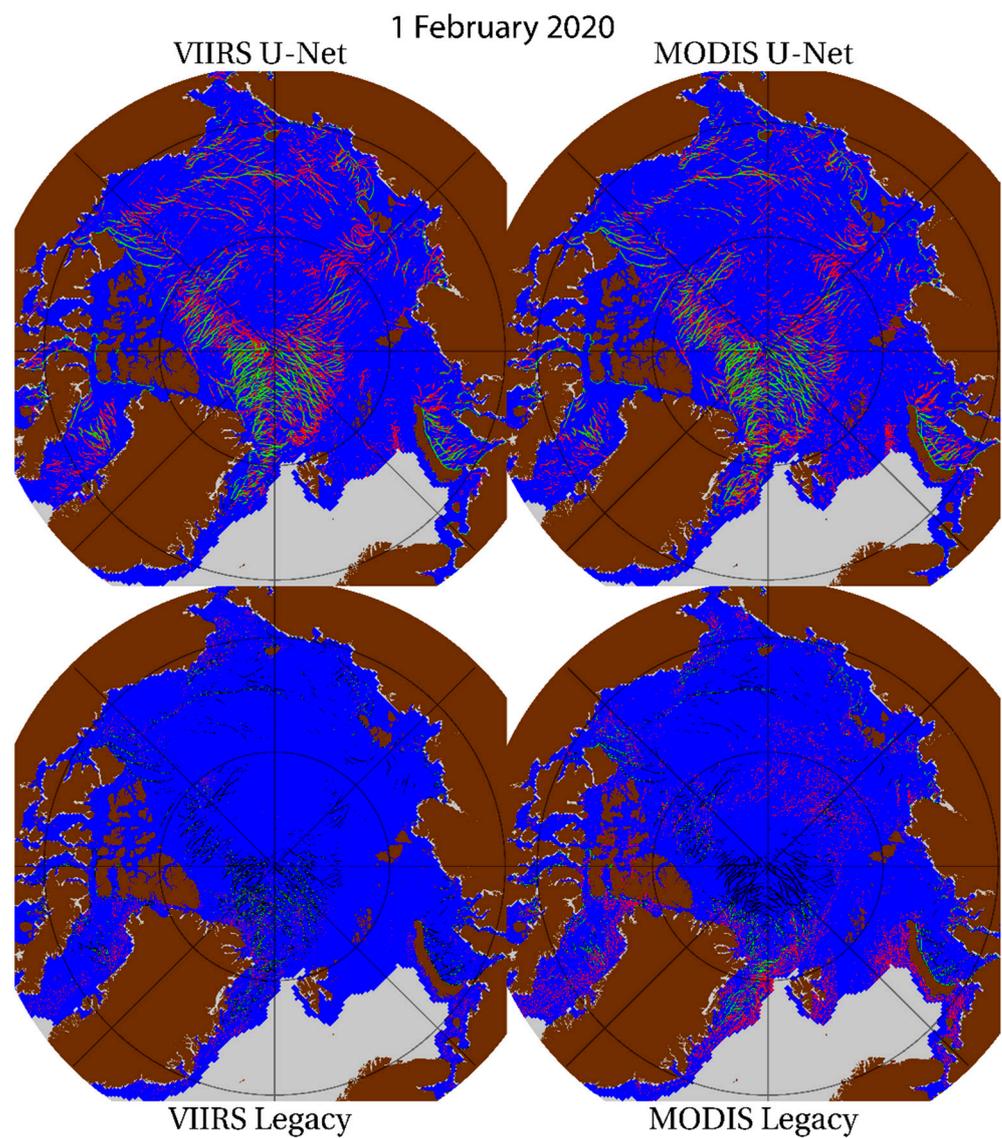


Figure A2. The full results from which a subset was shown in Figure 6. VIIRS (**left**) and MODIS (**right**) leads product results from U-Net (**top**) and Legacy product (**bottom**) on 1 February 2020. The color table is detailed in Table 2, blue is true negative, true positive is green true positive, red is false positive, and black false negative in comparison with the validation mask. Land is shown in brown and gray depicts open water.

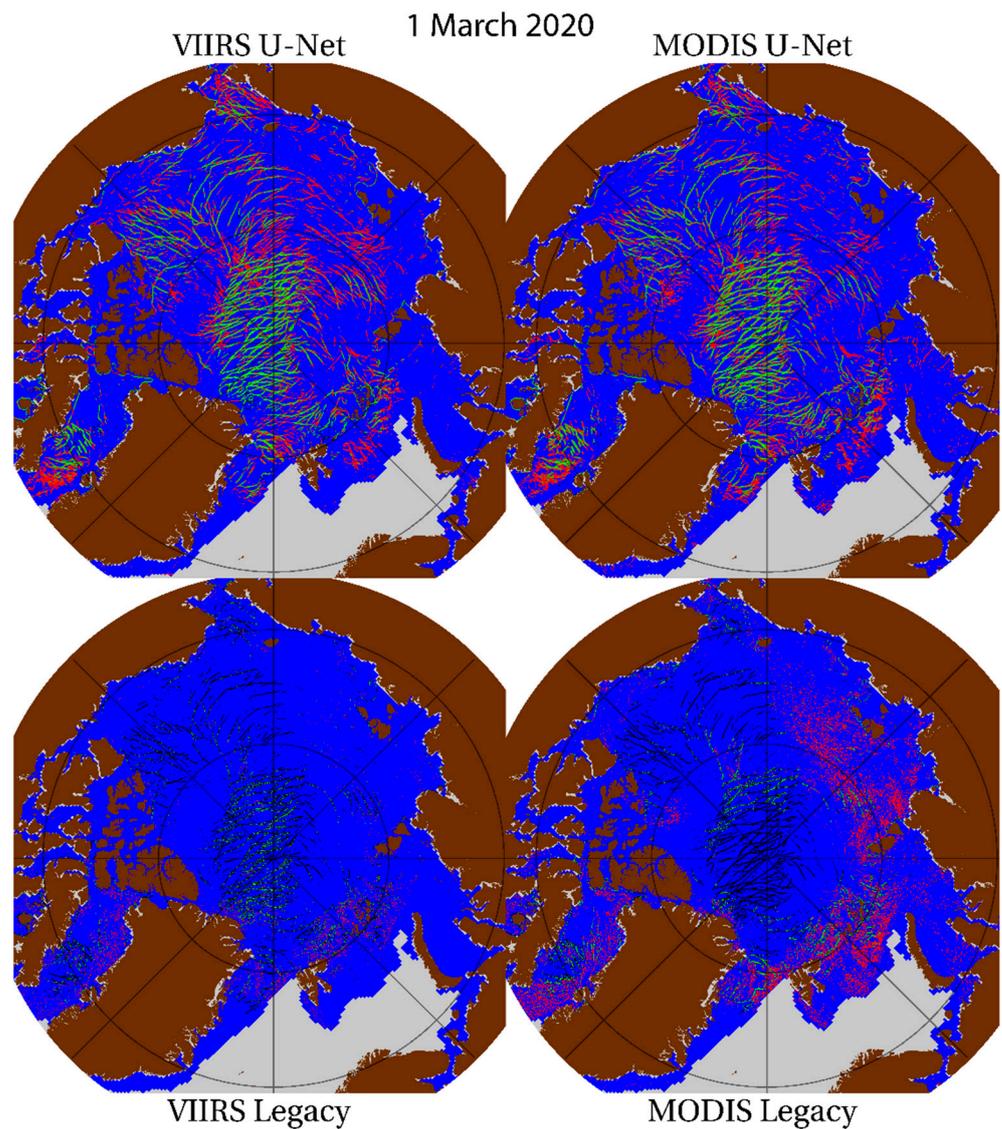


Figure A3. The full results from which a subset was shown in Figure 6. VIIRS (left) and MODIS (right) leads product results from U-Net (top) and Legacy product (bottom) on 1 March 2020. The color table is detailed in Table 2, blue is true negative, true positive is green true positive, red is false positive, and black false negative in comparison with the validation mask. Land is shown in brown and gray depicts open water.

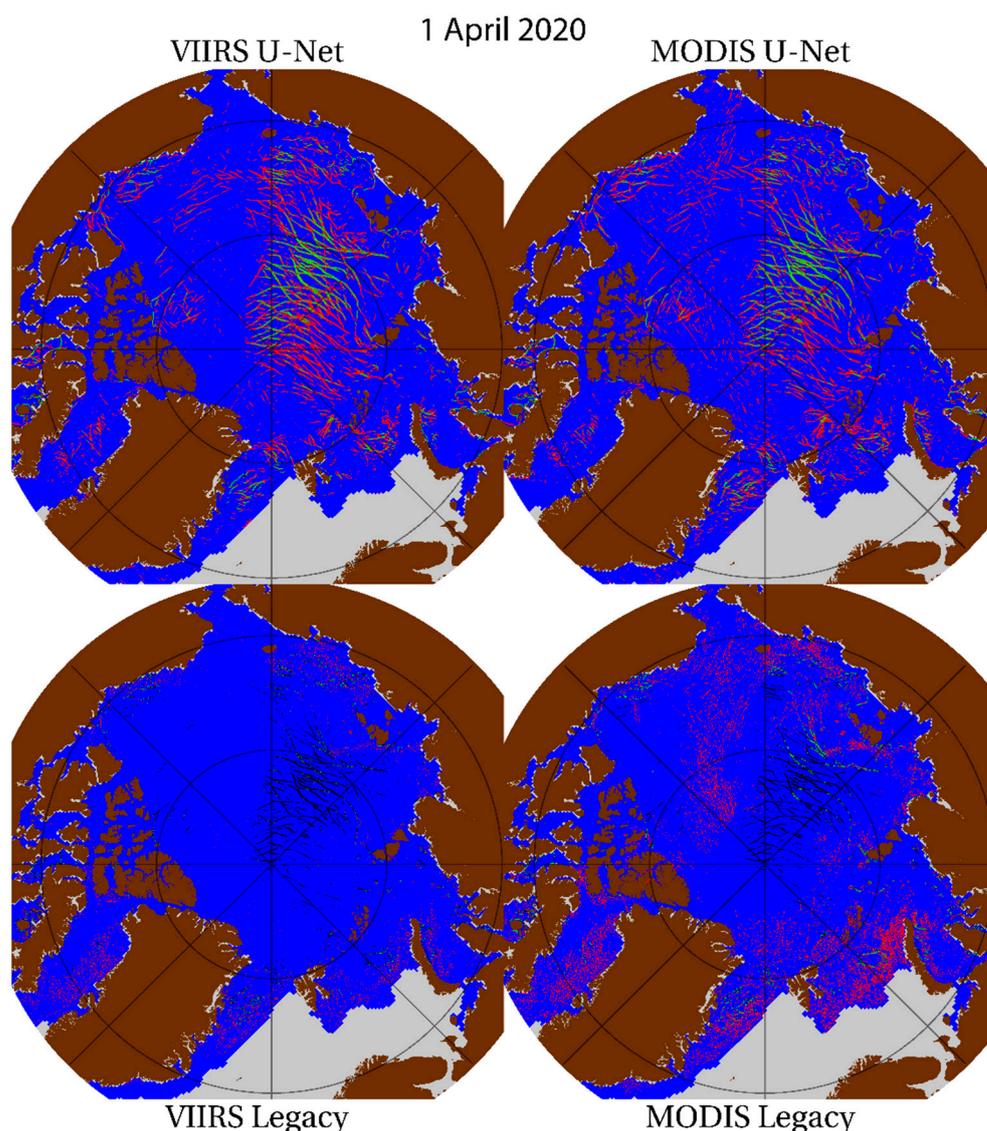


Figure A4. The full results from which a subset was shown in Figure 6. VIIRS (left) and MODIS (right) leads product results from U-Net (top) and Legacy product (bottom) on 1 April 2020. The color table is detailed in Table 2, blue is true negative, true positive is green true positive, red is false positive, and black false negative in comparison with the validation mask. Land is shown in brown and gray depicts open water.

References

1. Andreas, E.L.; Persson, P.O.G.; Grachev, A.A.; Jordan, R.E.; Horst, T.W.; Guest, P.S.; Fairall, C.W. Parameterizing Turbulent Exchange over Sea Ice in Winter. *J. Hydrometeorol.* **2010**, *11*, 87–104. [[CrossRef](#)]
2. Miles, M.W.; Barry, R. A 5-year satellite climatology of winter sea ice leads in the western Arctic. *J. Geophys. Res. Space Phys.* **1998**, *103*, 21723–21734. [[CrossRef](#)]
3. Maykut, G.A. Energy exchange over young sea ice in the central Arctic. *J. Geophys. Res. Space Phys.* **1978**, *83*, 3646–3658. [[CrossRef](#)]
4. Lüpkes, C.; Vihma, T.; Birnbaum, G.; Wacker, U. Influence of leads in sea ice on the temperature of the atmospheric boundary layer during polar night. *Geophys. Res. Lett.* **2008**, *35*, 35. [[CrossRef](#)]
5. Liu, Y.; Key, J.R.; Liu, Z.; Wang, X.; Vavrus, S.J. A cloudier Arctic expected with diminishing sea ice. *Geophys. Res. Lett.* **2012**, *39*, 39. [[CrossRef](#)]
6. Wang, Q.; Ilicak, M.; Gerdes, R.; Drange, H.; Aksenov, Y.; Bailey, D.A.; Bentsen, M.; Biastoch, A.; Bozec, A.; Boening, C.; et al. An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part I: Sea ice and solid freshwater. *Ocean Model.* **2016**, *99*, 110–132. [[CrossRef](#)]

7. Qu, M.; Pang, X.; Zhao, X.; Lei, R.; Ji, Q.; Liu, Y.; Chen, Y. Spring leads in the Beaufort Sea and its interannual trend using Terra/MODIS thermal imagery. *Remote Sens. Environ.* **2021**, *256*, 112342. [[CrossRef](#)]
8. Petty, A.A.; Bagnardi, M.; Kurtz, N.T.; Tilling, R.; Fons, S.; Armitage, T.; Horvat, C.; Kwok, R. Assessment of ICESat-2 Sea Ice Surface Classification with Sentinel-2 Imagery: Implications for Freeboard and New Estimates of Lead and Floe Geometry. *Earth Space Sci.* **2021**, *8*, 2020ea001491. [[CrossRef](#)]
9. Nguyen, A.T.; Heimbach, P.; Garg, V.V.; Ocaña, V.; Lee, C.; Rainville, L. Impact of Synthetic Arctic Argo-Type Floats in a Coupled Ocean–Sea Ice State Estimation Framework. *J. Atmos. Ocean. Technol.* **2020**, *37*, 1477–1495. [[CrossRef](#)]
10. Willmes, S.; Heinemann, G. Pan-Arctic lead detection from MODIS thermal infrared imagery. *Ann. Glaciol.* **2015**, *56*, 29–37. [[CrossRef](#)]
11. Hoffman, J.P.; Ackerman, S.A.; Liu, Y.; Key, J.R. The Detection and Characterization of Arctic Sea Ice Leads with Satellite Imagers. *Remote Sens.* **2019**, *11*, 521. [[CrossRef](#)]
12. Reiser, F.; Willmes, S.; Heinemann, G. A New Algorithm for Daily Sea Ice Lead Identification in the Arctic and Antarctic Winter from Thermal-Infrared Satellite Imagery. *Remote Sens.* **2020**, *12*, 1957. [[CrossRef](#)]
13. Röhrs, J.; Kaleschke, L. An algorithm to detect sea ice leads by using AMSR-E passive microwave imagery. *Cryosphere* **2012**, *6*, 343–352. [[CrossRef](#)]
14. Asadi, N.; Scott, K.A.; Komarov, A.S.; Buehner, M.; Clausi, D.A. Evaluation of a Neural Network With Uncertainty for Detection of Ice and Water in SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 247–259. [[CrossRef](#)]
15. Han, Y.; Liu, Y.; Hong, Z.; Zhang, Y.; Yang, S.; Wang, J. Sea Ice Image Classification Based on Heterogeneous Data Fusion and Deep Learning. *Remote Sens.* **2021**, *13*, 592. [[CrossRef](#)]
16. Khaleghian, S.; Ullah, H.; Krmer, T.; Eltoft, T.; Marinoni, A. Deep semi-supervised teacher-student model based on label propagation for sea ice classification. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10761–10772. [[CrossRef](#)]
17. Zhi, X. Unet.GitHub Repository. 2019. Available online: <https://github.com/zhixuhao/unet.git> (accessed on 10 November 2021).
18. Brodzik, M.J. Ease-grid: A versatile set of equal-area projections and grids. In *Discrete Global Grids*; Knowles, K.W., Ed.; National Center for Geographic Information & Analysis: Santa Barbara, CA, USA, 2002.
19. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 5–9 October 2015; Springer: New York, NY, USA, 2015; pp. 234–241.
20. MODIS Science Data Support Team. MODIS/AQUA Geolocation Fields 5-Min 11a Swath 1 km. Available online: <https://dx.doi.org/10.5067/MODIS/MYD03.006> (accessed on 10 November 2021).
21. MODIS Science Data Support Team. MODIS/TERRA Calibrated Radiances 5-Min 11b Swath 1 km. Available online: <http://dx.doi.org/10.5067/MODIS/MOD03.006> (accessed on 10 November 2021).
22. MODIS Science Data Support Team. MODIS/TERRA Geolocation Fields 5-Min 11a Swath 1 km. Available online: <http://dx.doi.org/10.5067/MODIS/MYD021KM.006> (accessed on 10 November 2021).
23. MODIS Science Data Support Team. MODIS/TERRA Calibrated Radiances 5-Min 11b Swath 1 km. Available online: <http://dx.doi.org/10.5067/MODIS/MOD021KM.006> (accessed on 10 November 2021).
24. VIIRS Calibration Support Team. VIIRS/JPSS1 Imagery Resolution Bands 6-Min 11b 6-Min Swath 375 m. Available online: <https://dx.doi.org/10.5067/VIIRS/VJ102IMG.021> (accessed on 10 November 2021).
25. VIIRS Calibration Support Team. VIIRS/JPSS1 Imagery Resolution Terrain-Corrected Geolocation I1 6-min Swath 375 m. Available online: <https://dx.doi.org/10.5067/VIIRS/VJ103IMG.021> (accessed on 10 November 2021).
26. VIIRS Calibration Support Team. VIIRS/NPP Imagery Resolution Bands 6-Min 11b 6-Min Swath 375 m. Available online: <https://dx.doi.org/10.5067/VIIRS/VNP02IMG.002> (accessed on 10 November 2021).
27. VIIRS Calibration Support Team. VIIRS/NPP Imagery Resolution Terrain-Corrected Geolocation I1 6-min Swath 375 m. Available online: <https://dx.doi.org/10.5067/VIIRS/VNP03IMG.002> (accessed on 10 November 2021).
28. Meier, W.; Fetterer, F.; Windnagel, A.; Stewart, S. *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration*, 4th ed.; NSIDC: National Snow and Ice Data Center: Boulder, CO, USA. [[CrossRef](#)]
29. Murashkin, D.; Spreen, G.; Huntemann, M.; Dierking, W. Method for detection of leads from Sentinel-1 SAR images. *Ann. Glaciol.* **2018**, *59*, 124–136. [[CrossRef](#)]
30. König, M.; Hieronymi, M.; Oppelt, N. Application of Sentinel-2 MSI in arctic research: Evaluating the performance of atmospheric correction approaches over arctic sea ice. *Front. Earth Sci.* **2019**, *7*, 22. [[CrossRef](#)]
31. Kwok, R.; Petty, A.A.; Cunningham, G.; Markus, T.; Hancock, D.; Ivanoff, A.; Wimert, J.; Bagnardi, M.; Kurtz, N. *The ICESat-2 Science Team. Atlas/Icesat-2 13a Sea Ice Freeboard*, 4th ed.; National Snow and Ice Data Center: Boulder, CO, USA, 2001.
32. Key, J.; Stone, R.; Maslanik, J.; Ellefsen, E. The detectability of sea-ice leads in satellite data as a function of atmospheric conditions and measurement scale. *Ann. Glaciol.* **1993**, *17*, 227–232. [[CrossRef](#)]
33. Key, J.; Maslanik, J.; Ellefsen, E. The effects of sensor field-of-view on the geometrical characteristics of sea ice leads and implications for large-area heat flux estimates. *Remote Sens. Environ.* **1994**, *48*, 347–357. [[CrossRef](#)]
34. Key, J.R. The area coverage of geophysical fields as a function of sensor field-of-view. *Remote Sens. Environ.* **1994**, *48*, 339–346. [[CrossRef](#)]