



Article

Understanding the Effects of Optimal Combination of Spectral Bands on Deep Learning Model Predictions: A Case Study Based on Permafrost Tundra Landform Mapping Using High Resolution Multispectral Satellite Imagery

Md Abul Ehsan Bhuiyan ^{1,*} , Chandi Witharana ¹, Anna K. Liljedahl ^{2,3}, Benjamin M. Jones ³, Ronald Daanen ⁴, Howard E. Epstein ⁵, Kelcy Kent ⁵, Claire G. Griffin ⁵ and Amber Agnew ¹

¹ Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, USA; Chandi.witharana@uconn.edu (C.W.); amber.agnew@uconn.edu (A.A.)

² Woodwell Climate Research Center, Falmouth, MA 02540, USA; aliljedahl@whrc.org

³ Institute of Northern Engineering, University of Alaska Fairbanks, Fairbanks, AK 99775, USA; bmjones3@alaska.edu

⁴ Alaska Department of Natural Resources, Division of Geological & Geophysical Surveys, Fairbanks, AK 99775, USA; ronald.daanen@alaska.gov

⁵ Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA; hee2b@virginia.edu (H.E.E.); kck7bw@virginia.edu (K.K.); cg4pm@virginia.edu (C.G.G.)

* Correspondence: md.bhuiyan@uconn.edu

Received: 29 July 2020; Accepted: 15 September 2020; Published: 17 September 2020



Abstract: Deep learning (DL) convolutional neural networks (CNNs) have been rapidly adapted in very high spatial resolution (VHSR) satellite image analysis. DLCNN-based computer visions (CV) applications primarily aim for everyday object detection from standard red, green, blue (RGB) imagery, while earth science remote sensing applications focus on geo object detection and classification from multispectral (MS) imagery. MS imagery includes RGB and narrow spectral channels from near-and/or middle-infrared regions of reflectance spectra. The central objective of this exploratory study is to understand to what degree MS band statistics govern DLCNN model predictions. We scaffold our analysis on a case study that uses Arctic tundra permafrost landform features called ice-wedge polygons (IWPs) as candidate geo objects. We choose Mask RCNN as the DLCNN architecture to detect IWPs from eight-band Worldview-02 VHSR satellite imagery. A systematic experiment was designed to understand the impact on choosing the optimal three-band combination in model prediction. We tasked five cohorts of three-band combinations coupled with statistical measures to gauge the spectral variability of input MS bands. The candidate scenes produced high model detection accuracies for the F1 score, ranging between 0.89 to 0.95, for two different band combinations (coastal blue, blue, green (1,2,3) and green, yellow, red (3,4,5)). The mapping workflow discerned the IWPs by exhibiting low random and systematic error in the order of 0.17–0.19 and 0.20–0.21, respectively, for band combinations (1,2,3). Results suggest that the prediction accuracy of the Mask-RCNN model is significantly influenced by the input MS bands. Overall, our findings accentuate the importance of considering the image statistics of input MS bands and careful selection of optimal bands for DLCNN predictions when DLCNN architectures are restricted to three spectral channels.

Keywords: deep learning; tundra; ice-wedge polygons; Mask R-CNN; satellite imagery; permafrost; Arctic

1. Introduction

Automated image analysis has long been a challenging problem in multiple domains. Over the last decade, deep learning (DL) convolutional neural nets (CNNs) have reshaped the boundaries of computer vision applications (CV), enabling unparalleled opportunities for automated image analysis. Applications span from everyday image understanding through industrial inspections to medical image analysis [1]. Conspicuous shortfalls of traditional per-pixel based approaches when confronted with sub-meter scale remote sensing imagery (satellite and aerial) have shifted the momentum towards novel paradigms, such as object-based image analysis (OBIA) [2], in which homogenous assemblages of pixels are considered in classification process. Recently, OBIA has been flanked by the challenges of big data [3] and scalability [4]. The success of DLCNNs in CV applications has received great interest from the remote sensing community [5]. There has been an explosion of studies integrating DLCNN to address remote sensing classification problems spanning from general land use and land cover mapping [6,7] to targeted feature extraction [8–10]. Deep learning CNNs excel at object detection [10–12], semantic segmentation (multiple objects of the same class indicate a single object) [7,13], and semantic object instance segmentation (multiple objects of the same class indicates distinct individual objects) [14]. Over the years, a plethora of DLCNN architectures have been proposed, developed, and tested. The influx of new DLCNNs continues to grow. Each has its own merits and disadvantages with respect to the detection and/or classification problem at hand. The appreciation for DLCNNs in the remote sensing domain is increasing, while some of the facets that are unique to remote sensing image analysis have been overlooked along the way.

Remote sensing scene understanding deviates from everyday image analysis in multiple ways, such as imaging sensors and their characteristics, coverage and viewpoints, and the objects and their behaviors in question. From the standpoint of Earth imaging, the image can be perceived as a reduced representation of the scene [2]. The image modality departs from the scene modality depending on the sensor characteristics. Scene objects are real-world objects, where image objects are assemblages of spatially arranged samples that model the scene. Images are only snapshots, and their size and shape are dependent on the sensor type and spatial sampling. For instance, certain land cover types, such as vegetation, are well-pronounced, exhibiting greater discriminative capacity in the near infrared (NIR) region than in the visible range. If the imaging sensor is constrained to the visible range, we limit ourselves from getting the advantage of the NIR wavelengths in classification algorithms. Similar to spectral strengths, the spatial resolution of the imaging sensor can either prohibit or permit our ability to construct the shape of the geo objects and their spatial patterning [15]. There is no single spatial scale that explains all the objects, but the semantics we pursue are organized into a continuum of scales [16,17]. In essence, an image represents the sensor's view of the reality and not the explicit representation of scene objects.

These are practical challenges that apply to very high spatial resolution (VHSR) multispectral (MS) commercial satellite imagery. The luxury of VSHR satellite imagery is that the wavelengths are not confined to traditional panchromatic, standard red-green-blue (RGB) channels, or NIR. VSHR satellite imagery includes both visible and NIR regions and, therefore, produce an array of multiple spectral channels. For instance, the WorldView02 sensor captures eight MS channels at less than 2 m resolution, and data fusion techniques allow resolution-enhanced MS products at sub-meter spatial resolutions. Besides spatial details, discriminating one geo object from another could be straightforward or difficult depending on their spectral responses recorded in the MS channels. Selection of optimal spectral bands is a function of the type of environment and the kind of information pursued in the classification process. In remote sensing mapping applications, land cover types and their constituent geo objects exhibit unique reflectance behaviors in different wavelengths, or spectral channels, enabling opportunities to discriminate them from each other and characterize them into semantic classes. This leaves the question for the user to select the optimal spectral channels from the MS satellite imagery for DLCNN model applications. The decision is difficult when candidate DLCNN architectures restrain the input to only three spectral channels. An intriguing question is should one adhere to RGB channels while

ruling out the criticality of other spectral bands, or is it necessary to mine all MS bands to choose the optimal bands for model predictions? To the best of our knowledge, this is a poorly explored problem despite its validity in remote sensing applications. Here, we make an exploratory attempt to understand this problem based on a case study that branches out from our on-going project on Arctic permafrost thaw mapping from commercial satellite imagery [18–21].

Permafrost thaw has been observed across the Arctic tundra [22]. Ice-rich permafrost landscapes commonly include ice wedges, for which growth and degradation is responsible for creating polygonized land surface features termed ice-wedge polygons (IWP). The lack of knowledge on fine-scale morphodynamics of polygonized landscapes introduces uncertainties to regional and pan-Arctic estimates of carbon, water, and energy fluxes [23]. Logistical challenges and high costs hamper field-based mapping of permafrost-related features over large spatial extents. In this regard, VHSR commercial satellite imagery enables transformational opportunities to observe, map, and document the micro-topographic transitions occurring in polygonal tundra at multiple spatial and temporal frequencies.

The entire Arctic has been imaged at 0.5 m resolution by commercial satellite sensors (DigitalGlobe, Inc., Westminster, CO, USA). However, imagery is still largely underutilized, and derived Arctic science products are rare. A considerable number of local-scale studies have analyzed ice wedge degradation processes using satellite imagery and manned-/unmanned aerial imagery/LiDAR data [24–27]. Most of the studies to date have relied on manual image interpretation and/or semi-automated approaches [25,26,28]. Therefore, there is a need and an opportunity for utilization of VHSR commercial imagery in regional scale mapping efforts to spatio-temporally document microtopographic changes due to thawing ice-rich permafrost. The bulk of remote sensing image analysis methods suffer from scalability and image complexities, but DLCNNs hold great promise in high throughput image analysis. Several pilot efforts [18,20,29–31] have demonstrated the potential adaptability of pre-trained DLCNN architectures in ice-wedge polygon mapping via the transfer learning strategy. However, the potential impacts of MS band statistics on DLCNN model predictions have been overlooked.

Owing to the increasing access of MS imagery and growing demand for a suite of pan-Arctic scale permafrost map products, there is a timely need to understand how spectral statistics of input imagery influence DLCNN model performances. Despite the design goals of DLCNN architectures to learn higher order abstractions of imagery without pivoting to the variations of low-level motifs, studies have documented the potential impacts of image quality, spectral/spatial artifacts of image compression, and other pre-processing factors on DLCNNs. Dodge et al. [32] described image quality impacts on multiple deep neural network models for image classification and showed that DL networks were sensitive to image quality distortions. Consequently, Dodge and Karam [33] carried out a comparison between human and deep learning recognition performance considering quality distortions and demonstrated that DL performance is still much lower than human performance on distorted images. Vasiljevic et al. [34] also investigated the effect of image quality on recognition by convolutional networks, which suffered a significant degradation in performance due to blurring along with a mismatch between training and input image statistics. Moreover, Karahan et al. [35] presented the influence of image degradations on the performance of deep CNN-based face recognition approaches, and their results indicated that blur, noise, and occlusion cause a significant decrease in performance. All these findings from previous studies provided useful insights towards developing computer visions applications, considering image quality that can perform reliably on image datasets.

Benchmark image datasets in CV applications are largely confined to RGB imagery, and trained DLCNNs are typically used for those data. Examples include imageNet, COCO, VisionData, MobileNet, etc. [36–39]. Priority for three spectral channels has become the de facto standard in everyday image analysis. As discussed before, this is a limiting factor in remote sensing applications. Training DLCNN architecture from scratch requires enormous amounts of training data to curtail overfitting. Because of this, transfer learning is becoming the standard practice to work with a limited amount of training data. In such circumstances, input channels are confined to three spectral channels,

despite the original image containing more than three channels. In remote sensing, multispectral bands could significantly affect the capacity to be invariant to quality distortions. Selection of the optimal spectral band combination from all the available multispectral channels for model training and prediction is important because the dominant land cover types (heterogeneity) control the global image statistics as well as local spectral variance. Our contention is that improper band selection or reliance solely on RGB channels can potentially hamper mapping accuracies. The information content in MS channels should be prudently capitalized on DLCNN model predictions; otherwise, we are discarding valuable cues that are advantageous in automated detection and classification processes. The central objective of this exploratory study is to understand to what degree MS band statistics govern the DLCNN model predictions. We scaffold our analysis on a case study that includes ice-wedge polygons in two common tundra vegetation types (tussock and non-tussock sedge) as candidate geo objects. We choose Mask RCNN as the candidate DLCNN architecture to detect ice-wedge polygons from eight-band Worldview-02 commercial satellite imagery. A systematic experiment was designed to understand the impact of choosing the optimal three-band combination on model prediction. We tasked five cohorts of three-bands combinations coupled with statistical measures to gauge the spectral variability of input MS bands.

2. Study Area and Image Data

Our study area covers coastal and upland tundra near Nuiqsut on the North Slope of Alaska (Figure 1). We obtained two summer-time WorldView-02 (WV2) commercial satellite image scenes from tussock sedge and non-tussock sedge tundra regions from the Polar Geospatial Center (PGC) at the University of Minnesota. The WV2 sensor records spectral reflectance at eight discrete wavelengths representing coastal blue (band 1), blue (band 2), green (band 3), yellow (band 4), red (band 5), red edge (band 6), NIR1 (band 7), and NIR2 (band 8). The spatial resolution of the data product is ~0.5 m with 16-bit radiometric resolution. Scenes were chosen from tussock and non-tussock sedge tundra regions based on the Circumpolar Arctic Vegetation Map (CAVM) [40], which presents important baseline reference data for pan-Arctic vegetation monitoring in tundra ecosystems [41]. Alaska has heterogeneous tundra types such as tussock sedge, dwarf shrub, and moss tundra in the foothills of northern Alaska [41]. The region generally covers (Figure 1, details in [40]): (i) Non-tussock sedge, dwarf-shrub, moss tundra: moist tundra dominated by sedges and dwarf shrubs < 40 cm tall, with a well-developed moss layer; (ii) Tussock sedge, dwarf-shrub, moss tundra: moist tundra, dominated by tussock cottongrass and dwarf shrubs <40 cm tall; (iii) Sedge: wetland complexes in the colder/warmer areas of the Arctic, dominated by sedges, grasses, and mosses. We chose two WV2 satellite image scenes focusing only two candidate vegetation types: (1) tussock sedge and (2) non-tussock sedge tundra for our analysis.

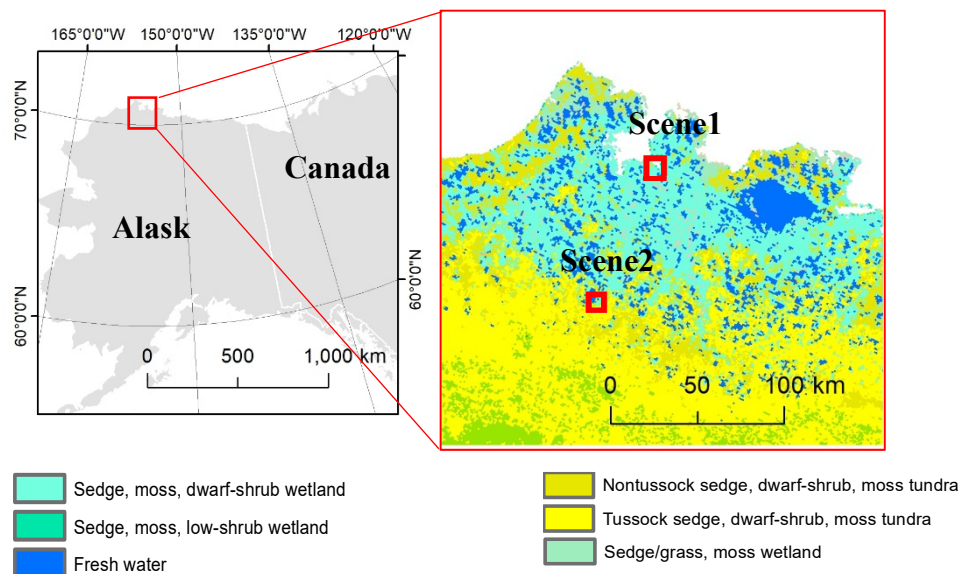


Figure 1. Location of candidate image scenes overlain on the Circumpolar Arctic Vegetation Map (CAVM, [40]).

3. Methodology

3.1. Mapping Workflow

In this study, we used a deep learning convolutional neural net (CNN) architecture called Mask-RCNN [42]. It is a semantic segmentation algorithm, which has been successfully applied in ice-wedge polygon mapping in Arctic regions [18–21]. The DL algorithm performs the object instance segmentation with outputs as predicted binary masks with classification information [18–21]. Our workflow has three stages. The first stage involves image pre-processing and tiling, in which the input image scene is partitioned into smaller subsets with dimensions of 200 pxl × 200 pxl. The second stage involves the application of the Mask RCNN model over the input tiles. In the last stage, the model predictions for each tile get post-processed and generate a seamless shapefiles corresponding to the input image scene. Figure 2 exhibits a generalized schematic diagram of the workflow.

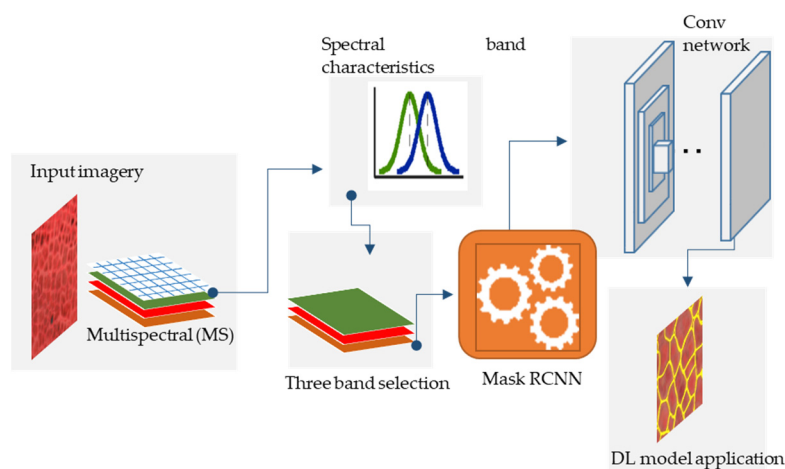


Figure 2. Simplified schematic of the automated ice wedge polygon mapping workflow.

We utilized ResNet-101 as the backbone of the Mask R-CNN model. The model was trained with a mini-batch size of two image tiles, 350 steps per epoch, learning rate of 0.001, learning momentum

of 0.9, and weight decay of 0.0001. To minimize overfitting, random horizontal flips augmentation was applied to introduce variability in the training data for acceptable estimation accuracy. During calibration, the weights and biases of each neuron were estimated iteratively by minimizing a mean squared error cost function using a gradient descent algorithm with back propagation [43]. We exercised a transfer learning strategy. Pre-trained Mask RCNN was retrained using hand-annotated ice-wedge polygon samples generated from commercial satellite imagery. Approximately 40,000 ice-wedge polygon samples were incorporated. In the training schedule, the samples were divided into three categories of training (80%), validation (10%), and testing (10%) [18].

Like other commonly used DLNNs, the pre-trained Mask RCNN is based on the training data repositories of everyday image information, such as coco data, which limits the use of more than three channels. In contrast, commercial satellite imagery captures substantial landscape heterogeneity based on an array of wavelengths, which may bias the model predictions. In this study, we chose two multispectral image scenes from varying tundra types: tussock and non-tussock sedge in our modeling approach. There is a unique opportunity to improve IWP mapping utilizing multispectral bands along with landscape information at regional scales. High spatial resolution is essential to accurately describe feature shapes and textural patterns, while high spectral resolution is needed to classify thematically detailed land-use and land-cover types [4,44]. Specifically, to ensure the spectral quality of the original image scene, the magnitude and variability of the band-wise measures in each pixel should be examined. This stage also helps to decrease the spectral heterogeneity, which identifies the best spectral band combination to produce a robust image classification model. In particular, to obtain the best combination of spectral bands from input multispectral imagery, we present three statistical measures: (1) probability distribution function (PDF), (2) cumulative distribution function (CDF), and (3) coefficient of variation (CV) [45–53]. To ensure the spectral quality of the original image scene, the magnitude and variability of the band-wise measure in each pixel were examined using these three statistical measures, which helps to decrease the spectral heterogeneity, leading to the best spectral band combination to produce a robust image detection model.

As the first step in the pipeline, the most effective combination of bands is obtained by examining the shape of the cumulative density function (CDF) to control the number of pixel-based features for multispectral images and automatically determined a fixed number of features using CDF. Linag et al. [53] applied CDF along with a kernel density estimation method, which is used to obtain the detection threshold for the High-Resolution SAR Image application. Based on previous studies, we used CDF in our analysis to explain the distribution of the reflectance values among multiple spectral bands, and we choose the most similar three bands. Inamdar et al. [50] conducted a PDF matching technique for preprocessing of multitemporal remote sensing images for various applications, such as supervised classification, partially supervised classification, and change detection and successfully evaluated these techniques across scenarios. This study showed the necessity of matching distributions of two (or more) multispectral bands of remote sensing images for enhancing image modeling. Furthermore, Pitié et al. [51] also proposed a PDF-based algorithm to determine the possible change of content in image data to avoid excessive stretching of the mapping functions and successfully reduced the magnitude of the stretching. All these investigations motivate advancing the IWP mapping application by prudently mining multispectral bands using the PDF function.

To examine the magnitude and variability of the band-wise measure, the coefficient of variation (CV—standard deviation divided by the mean) was examined to ensure the spectral quality of the original image scene [45,46,48]. To compare the degree of variation from one data series to another, we used the CV where we considered distributions with $CV < 1$ to be low variance, while we considered those with $CV > 1$ to be high variance. Previous studies used this metric to measure spectral variance to explore multispectral band influences in image processing applications. For example, Bovolo et al. [48] evaluated the effectiveness of their image processing approach in homogeneous and border areas using the CV in order to improve the geometric fidelity of mapping applications. Wang et al. [45] carried out an experiment using CVs for visible and near-infrared spectral regions across pixel sizes,

and they found significant changes in the relative contribution to spectral diversity from leaf traits (detectable in the visible region) to canopy structure with increasing pixel size. To apply the CV in a preprocessing dataset, it helps to understand the spectral heterogeneity, which leads to a more robust image classification model than traditional object-detection models. Based on visual inspections, we confined ourselves to five cohorts of three-band combinations. Optimization of multispectral image bands based on CV, PDF, and CDF reduces the ambiguity, uncertainty, and errors in IWP mapping. The selected band combination had the maximum amount of information and the correlation between the bands was as small as possible. We also conducted a Kruskal-Wallis [54] test, where p -values among band combinations were calculated to see the significance of the accuracy assessment for each band combination. The Kruskal-Wallis test is a rank-based nonparametric test, which is used to determine significant differences among datasets. The Kruskal-Wallis test does not assume that the data are normally distributed, and this test allows us to compare more than two independent datasets [55]. The purpose of this statistical analysis is to systematically evaluate how the spectral character of the input data influences the prediction accuracies of DLCNN models.

3.2. Accuracy Estimates for Model Prediction

The DLCNN model assessment was performed using different error metrics: random error quantification (root mean square error, RMSE), systematic error quantification (absolute mean relative error, AMRE), and uncertainty quantification (F1 Score). The error metrics used in study are summarized below.

Absolute mean relative error (AMRE) is the mean of the relative percentage error, calculated by the normalized average:

$$AMRE = \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right) \right| \quad (1)$$

where the number of predicted polygons is \hat{y} , and the number of actual polygons is y_i , and n is the quantity of samples used in the calculation. For an unbiased model, the AMRE would be 0 [56,57].

The root mean square error (RMSE) is a statistical metric used to measure the magnitude of the random error and defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[\hat{y}_i - y_i - \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \right]^2} \quad (2)$$

Here, RMSE ranges from 0 (an optimal value) to positive infinity [58,59].

Correctness indicates how many of the predicted positives were truly positive. Completeness determines what percentage of actual positives were detected. F1 Score derives a balance between Correctness and Completeness into one value [60]:

$$\text{Correctness} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Completeness} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1 Score} = \frac{2 * \text{Correctness} * \text{Completeness}}{\text{Correctness} + \text{Completeness}} \quad (5)$$

Here, true positive (TP) is the number of polygons correctly identified, false positive (FP) is the number of polygons identified by model but not true, and false negative (FN) is undetected polygons. Higher values of Correctness indicate that there are fewer false positives in the classification. If the model always predicts positive magnitudes, Completeness will be 1, which indicates that ice-wedge polygons are properly detected by the model. Moreover, an F1 score of 1 indicates perfect prediction of ice-wedge polygons.

4. Results and Discussion

4.1. Statistical Measures for Input Image

The qualitative evaluation of the cumulative probability matching results is performed by visually comparing the shapes of the eight band distributions for tussock and non-tussock sedge tundra regions, which is shown in Figure 3. The CDF of spectral bands showed that band 7 and band 8 deviated significantly from the other bands for the non-tussock tundra sedge type (Figure 3). The band combination of 3,4,5 approximately presented similar spread and variability.

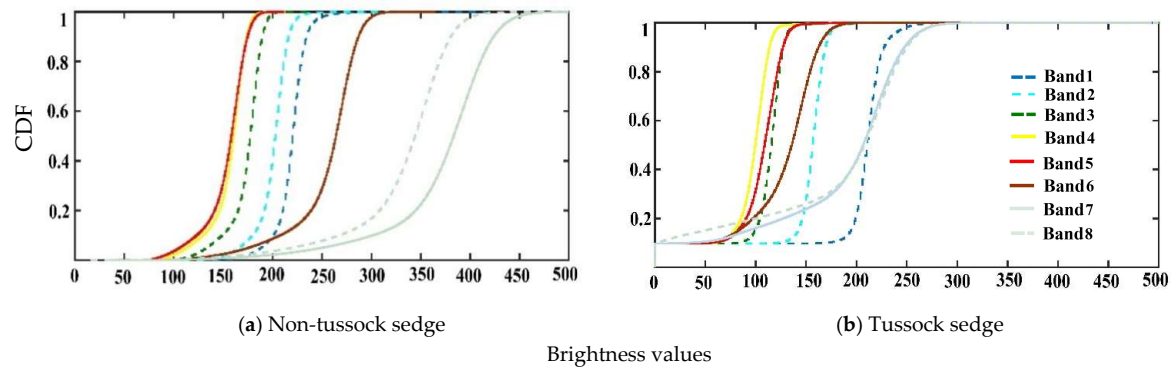


Figure 3. Cumulative distribution function for the multispectral band datasets.

Figure 4 shows the comparison of PDFs of the brightness values among all the spectral bands. Band 7 and band 8 displayed a clear separation of their PDFs from the other spectral bands for the non-tussock sedge tundra type. Specifically, it is interesting to note that bands 3,4,5 respond almost similarly compared to other spectral bands for both tundra types (tussock sedge and non-tussock sedge), highlighting the sensitivity of choosing band combination for the model predictions and its dependence on the brightness values.

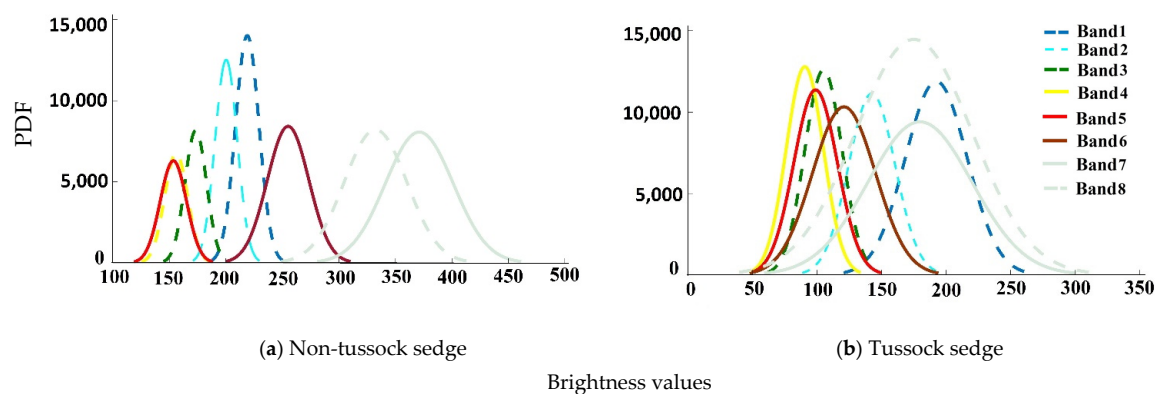


Figure 4. Probability distribution function for the multispectral band datasets.

We produced box plots of the brightness values for spectral bands for tussock sedge and non-tussock sedge tundra types, as shown in Figure 5. There were no significant differences in terms of variability for the band combination of 1,2,3 and the band combination of 3,4,5, indicating that both combinations could be applied in model predictions. We used the coefficient of variation (CV) to evaluate the degree of variation of various spectral bands. From the analysis of CV from spectral bands, the distributions of brightness values showed low variability ($CV < 1$), which is shown in Figure 6. In terms of CV, the band combination of 1,2,3 and the band combination of 3,4,5 were consistent for tussock and non-tussock sedge tundra regions. These results helped us to understand the feasibility and reliability of the remote sensing information extraction for large-scale application research.

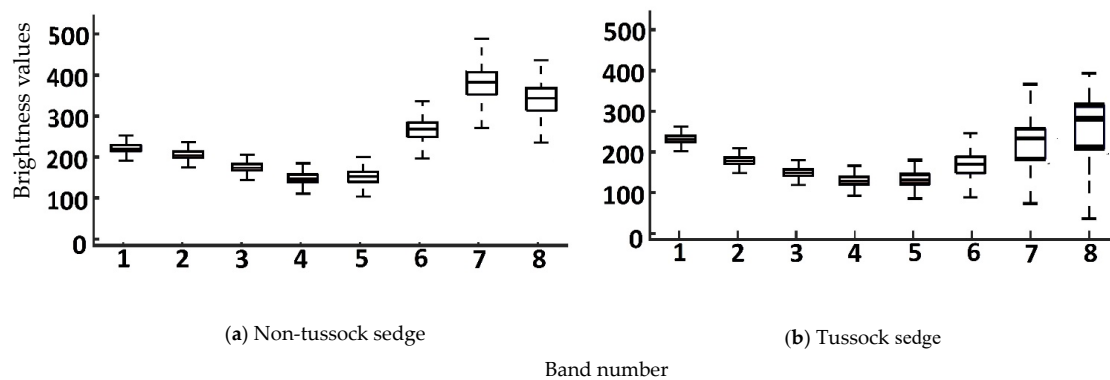


Figure 5. Box plot for the multispectral band datasets. In each box, the central mark is the median, and the edges are the first and third quartiles.

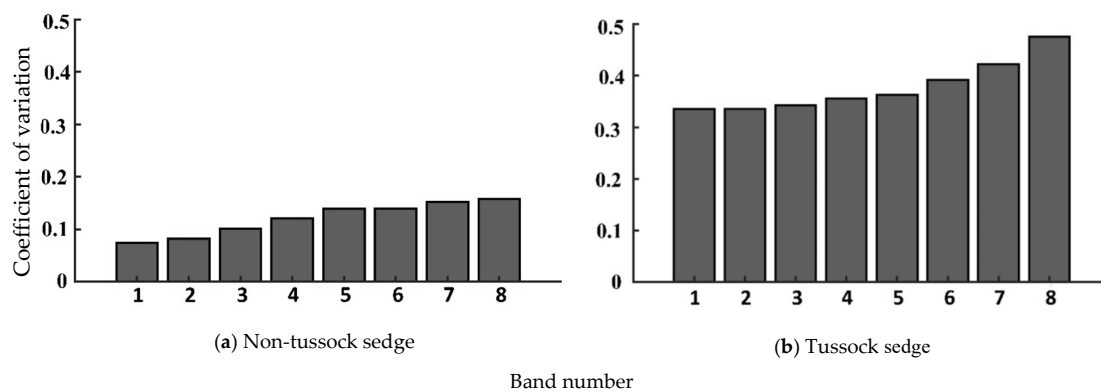


Figure 6. Coefficient of variation for the multispectral band datasets.

4.2. Statistical Measures for Model Prediction

We statistically evaluated the performances of the DL model in detecting IWPs. For the quantitative assessments, from each image scene, we randomly selected 15 subsets to manually delineate polygons as a reference (ground-truth polygons). Systematic error was exhibited for all image scenes, and results showed significantly low AMRE values (0.16–0.19) for three-band combinations (1,2,3 and 3,4,5) for the different tundra types (tussock and non-tussock sedge) (Table 1). Other combinations produced comparatively high systematic error (0.33–0.38). The results provide insights on how to select a proper band combination for remote sensing satellite images for deep learning modeling. Similarly, the random error for two band combinations (1,2,3 and 3,4,5) varied from 0.17 to 0.21 for the two candidate scenes (Table 2), indicating a robust performances of the new version of the ice-wedge polygon mapping algorithm in varying tundra types (tussock and non-tussock sedge). Therefore, this methodology adopted the best combination scheme to perform the image analysis for IWP mapping.

Table 1. Summary statistics of Absolute Mean relative Error (AMRE) for candidate scenes.

Image Scene	Three Band Combination				
	1,2,3	2,3,5	2,3,7	3,4,5	3,5,7
Non-tussock sedge	0.17	0.35	0.33	0.16	0.38
Tussock sedge	0.19	0.38	0.35	0.19	0.38

Table 2. Summary statistics of Root Mean Square Error (RMSE) for candidate scenes.

Image Scene	Three Band Combination				
	1,2,3	2,3,5	2,3,7	3,4,5	3,5,7
Non-tussock sedge	0.20	0.38	0.34	0.19	0.42
Tussock sedge	0.21	0.39	0.35	0.17	0.47

We used three quantitative error statistics (correctness, completeness, and F1 score) to show the performances of the framework. The candidate scenes 1 and 2 produced high model detection accuracies for the F1 score, ranging between 0.89 to 0.95, for two band combinations (1,2,3 and 3,4,5) (Figure 7, Table 3). Similarly, scenes 1 and 2 scored high values for completeness (85–91%). In the two cases, the correctness metric scored ~1, allowing less freedom for false alarms. Optimization of the multispectral bands improved the detection accuracy of the model.

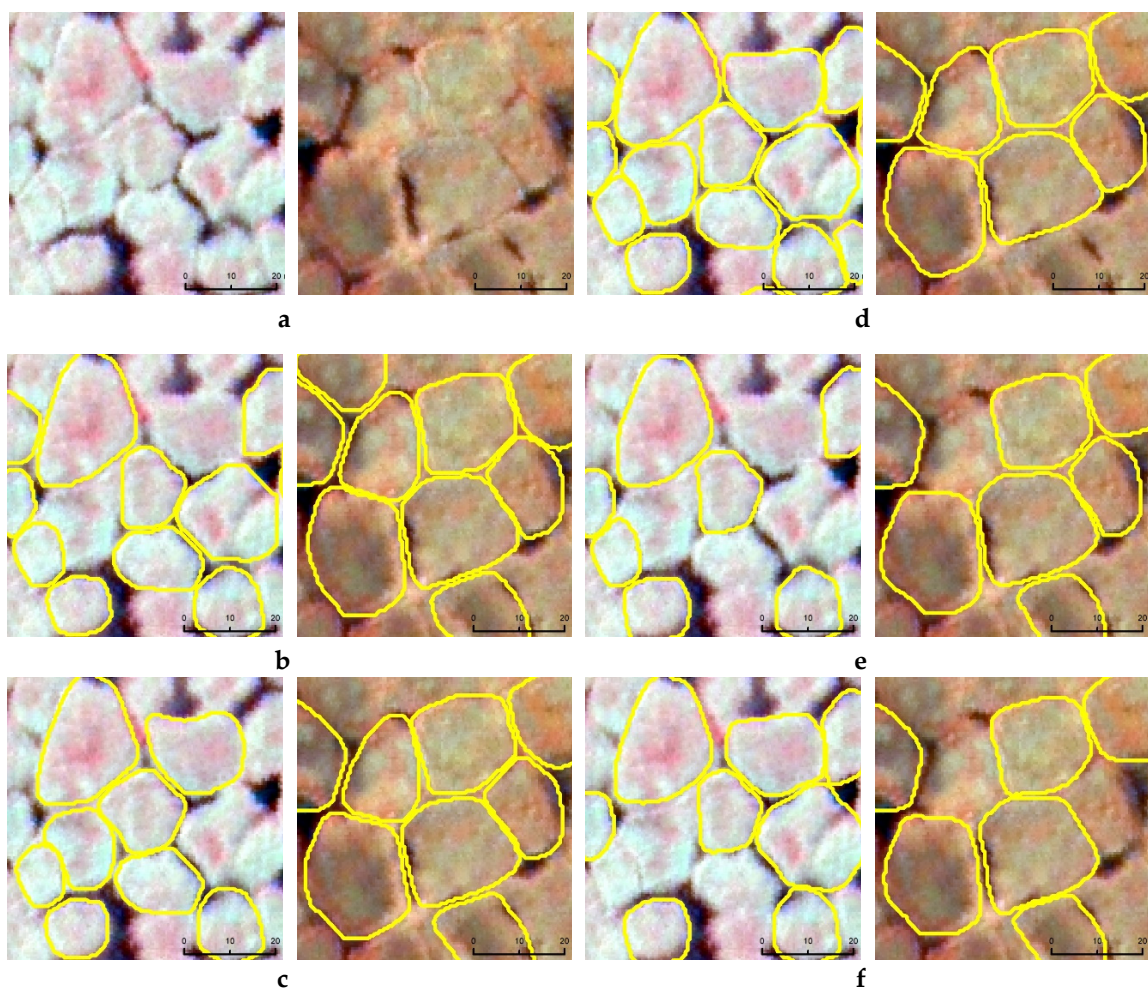
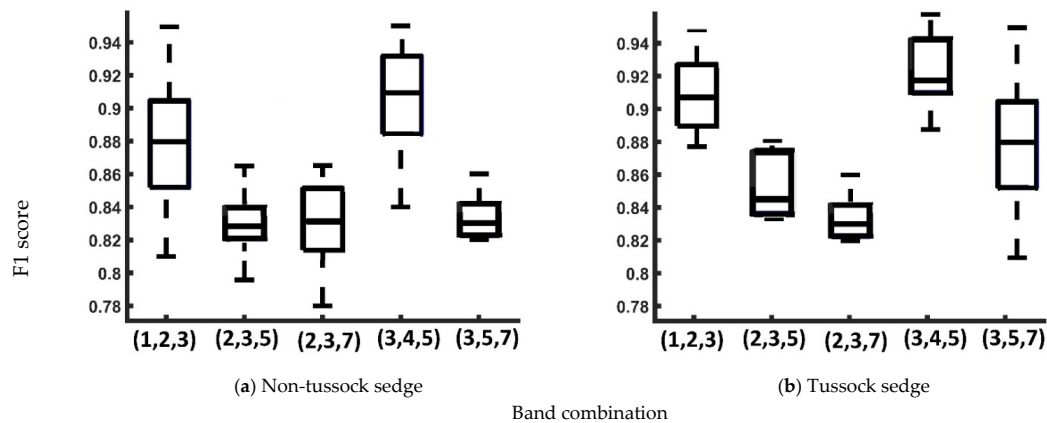


Figure 7. Simplified zoomed-in views of original imagery and model predictions. (a) original imagery for non-tussock sedge (from left column 1) and original imagery for tussock sedge (from left column 2); band combination (b) (1,2,3); (c) (2,3,5); (d) (2,3,7); (e) (3,4,5); (f) (3,5,7). Yellow outlines denote automatically detected IWPs. Imagery © [2012,2015] DigitalGlobe, Inc. (Westminster, CO, USA).

Table 3. Accuracy assessment of detection for candidate image scenes.

Band Combination	Non-Tussock Sedge			Tussock Sedge		
	Correctness	Completeness	F1 Score	Correctness	Completeness	F1 Score
1,2,3	1	85%	0.89	1	89%	0.92
2,3,5	1	81%	0.84	1	82%	0.85
2,3,7	1	82%	0.84	1	82%	0.83
3,4,5	1	86%	0.91	1	91%	0.95
3,5,7	1	82%	0.83	1	82%	0.88

Box plots for the F1 score for five three-band cohorts (1,2,3; 2,3,5; 2,3,7; 3,4,5; and 3,5,7) are presented in Figure 8. The distribution of F1 scores showed substantial discrepancies in terms of variability for the band combinations for (2,3,5), (2,3,7), and (3,5,7) compared to other band combinations ((1,2,3) and (3,4,5)). Therefore, the two band combinations (1,2,3) and (3,4,5) are treated as the best combinations of spectral bands to create robust models for ice-wedge polygon mapping.

**Figure 8.** Box plot of F1 score for the different band combinations. In each box, the central mark is the median, and the edges are the first and third quartiles.

The Kruskal–Wallis test was conducted to determine whether the medians from various datasets are different or not. We calculated p -values for two band combinations (1,2,3 and 3,4,5) compared with other combinations to test for significant differences in F1 scores (Table 4). The p -values were found to be $< \alpha$ ($\alpha = 0.05$) for the combinations of (2,3,5), (2,3,7), and (3,5,7). On the other hand, the two band combinations (1,2,3) and (3,4,5) were not statistically different with p -values $> \alpha = 0.05$. Therefore, the two band combinations (1,2,3) and (3,4,5) are used in deep learning-based modeling for this study area.

Table 4. p -values of the Kruskal–Wallis test for F1 scores.

Non-Tussock Sedge			Tussock Sedge		
Band Combination	1,2,3	3,4,5	Band Combination	1,2,3	3,4,5
1,2,3	1	0.0824	1,2,3	1	0.0189
2,3,5	4.22×10^{-4}	2.33×10^{-5}	2,3,5	0.0	0.0
2,3,7	0.0011	4.01×10^{-5}	2,3,7	0.0	0.0
3,4,5	0.0824	1	3,4,5	0.0189	1
3,5,7	3.77×10^{-4}	7.83×10^{-5}	3,5,7	1.40×10^{-3}	4.11×10^{-5}

In this exploratory study, we use two tundra vegetation types, tussock sedge and non-tussock sedge. However, arctic tundra includes additional vegetation types. Therefore, the model can be biased

when it is applied to other tundra vegetation types such as prostrate dwarf-shrub, herb, lichen tundra; rush/grass, forb, cryptogam tundra; graminoid, prostrate dwarf-shrub, forb tundra, etc. In the future, this experiment can be extended considering more diverse tundra landscapes, such as graminoid and shrub dominated vegetation cover types, to systemically gauge the improvement of the DL model prediction accuracies.

The mapping workflow discerned the IWPs by exhibiting significantly low random and systematic error for band combinations (1,2,3). Results suggest that the prediction accuracy of the Mask-RCNN model is significantly influenced by the input MS bands. The DL model exhibited high detection accuracies (89% to 95%) compared to previous studies [20,21]. According to the results, it is also suggested that the removal of the unnecessary bands information is crucial for enhancing the model accuracy. Most of the DLCNN architectures are tailored to operate in the context of three-band or single-band images. Therefore, when practicing transfer-learning strategies, there is a high chance of overlooking the wealth of information contained in multispectral image channels and their correlation among bands. This additional information could prudently leverage model performances and prediction accuracy. Overall, transfer-learning strategies optimized the model with limited data utilizing processing time and computation resources instead of original DL based hardcore programming competencies. Moreover, the use of all spectral bands does not necessarily improve classification accuracy, but there is a high tendency to degrade to prediction accuracies. The sensible approach is to methodologically select optimal three bands cohort with respect to the classification problem in hand. Therefore, to create an automated ice-wedge polygon mapping framework, optimal combination of spectral bands was obtained from the very high resolution satellite imagery. In summary, we conducted our exploratory analysis based on transfer learning strategy to get the most benefit out from existing CNN weights that are formulated based on three-band imagery from everyday image analysis. While the current exploratory study scaffolds on a single candidate DL model, a candidate geo-object, and a satellite sensor model, in future research, we will aim to expand the experimental design incorporating multiple DL models, a suite of geo-object types, and several satellite sensor model to investigate the effect of band statistics on DL model predictions.

5. Conclusions

Our study highlights the importance of the band combinations in the use of multispectral datasets on deep learning convolutional neural net (DLCNN) model prediction accuracy. Here, in the context of ice-wedge polygon mapping, we applied the DLCNN architecture, to detect IWPs from eight-band Worldview-02 VHSR satellite imagery in the North slope of Alaska across two common tundra vegetation types (tussock sedge and non-tussock sedge tundra). We found that the prediction accuracy of the Mask-RCNN model is significantly influenced by the input of MS bands. Optimization of multispectral image bands combination reduced the ambiguity, uncertainty, and errors in ice-wedge polygon mapping. Results suggest that mapping applications depend on the careful selection of the best band combinations. The selected band combination had the maximum amount of information with the least amount of correlation among the bands and the correlation between the bands were as small as possible. Overall, our findings emphasize the importance of considering the image statistics of input MS bands and careful selection of optimal bands for DLCNN predictions, when the architectures are restricted with three spectral channels. Thinking beyond yet another target detection based on transfer learning, we have materialized an-image-to-assessment pipeline, which is capable of parallel processing of sheer volumes of image scenes on HPC resources. Our framework is not fundamentally limited to IWP mapping but rather an extensible workflow to perform high throughput, regional-scale mapping tasks such as other permafrost features; thaw slumps, capillaries, shrubs, etc. Our future study is to apply the mapping algorithm over large land areas of the heterogeneous Arctic tundra.

Author Contributions: M.A.E.B. designed the study, and performed the experiments, and led the manuscript writing and revision. C.W. led the overall project, co-designed the study, and contributed to the development of the paper. A.K.L. contributed to the interpretation of results and the writing of the paper, B.M.J. contributed

to interpret results and, together with R.D., H.E.E., K.K., and C.G.G. contributed to the writing of the paper. A.A. provided training data and interpret results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the U.S. National Science Foundation Grant Nos.: 1720875, 1722572, and 1721030. Supercomputing resources were provided by the eXtreme Science and Engineering Discovery Environment (Award No. DPP190001). The authors would like to thank Polar Geospatial Center, University of Minnesota for imagery support. BMJ was supported by the U.S National Science Foundation Grant No. OIA-1929170.

Acknowledgments: We would like to thank Torre Jorgenson and Mikhail Kanevskiy for their domain expertise on permafrost feature classification.

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

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
2. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
3. Lang, S.; Baraldi, A.; Tiede, D.; Hay, G.; Blaschke, T. Towards a (GE) OBIA 2.0 manifesto—Achievements and open challenges in information & knowledge extraction from big Earth data. In Proceedings of the GEOBIA, Montpellier, France, 18–22 June 2018.
4. Witharana, C.; Lynch, H. An object-based image analysis approach for detecting penguin guano in very high spatial resolution satellite images. *Remote Sens.* **2016**, *8*, 375. [CrossRef]
5. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
6. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
7. Li, Y.; Qi, H.; Dai, J.; Ji, X.; Wei, Y. Fully convolutional instance-aware semantic segmentation. *arXiv* **2016**, arXiv:1611.07709.
8. Yang, B.; Kim, M.; Madden, M. Assessing optimal image fusion methods for very high spatial resolution satellite images to support coastal monitoring. *Gisci Remote Sens.* **2012**, *49*, 687–710. [CrossRef]
9. Wei, X.; Fu, K.; Gao, X.; Yan, M.; Sun, X.; Chen, K.; Sun, H. Semantic pixel labelling in remote sensing images using a deep convolutional encoder-decoder model. *Remote Sens. Lett.* **2018**, *9*, 199–208. [CrossRef]
10. Abdulla, W. Mask R-Cnn for Object Detection and Instance Segmentation on Keras and Tensorflow. Available online: https://github.com/matterport/Mask_RCNN (accessed on 25 August 2020).
11. Dai, J.; He, K.; Sun, J. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: Piscataway, NJ, USA, 2016.
12. Ren, Z.; Sudderth, E.B. Three-dimensional object detection and layout prediction using clouds of oriented gradients. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1525–1533.
13. Vuola, A.O.; Akram, S.U.; Kannala, J. Mask-RCNN and U-net ensembled for nuclei segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019; pp. 208–212.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the MICCAI 2015, Munich, Germany, 5–9 October 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; LNCS: New York, NY, USA, 2015; Volume 9351, pp. 234–241. [CrossRef]
15. Woodcock, C.E.; Strahler, A.H. The factor of scale in remote sensing. *Remote Sens. Environ.* **1987**, *21*, 311–332. [CrossRef]
16. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; Van der Meer, F.; Van der Werff, H.; Van Coillie, F.; et al. Geographic object-based image analysis—Towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191. [CrossRef]
17. Hay, G.J. Visualizing ScaleDomain Manifolds: A multiscale geobased approach. *Scale Issues Remote Sens.* **2014**, 139–169. [CrossRef]

18. Bhuiyan, M.A.E.; Witharana, C.; Liljedahl, A.K. Big Imagery as a Resource to Understand Patterns, Dynamics, and Vulnerability of Arctic Polygonal Tundra. In Proceedings of the AGUFM 2019, San Francisco, CA, USA, 9–13 December 2019; p. C13E-1374.
19. Witharana, C.; Bhuiyan, M.A.E.; Liljedahl, A.K. Towards First pan-Arctic Ice-wedge Polygon Map: Understanding the Synergies of Data Fusion and Deep Learning in Automated Ice-wedge Polygon Detection from High Resolution Commercial Satellite Imagery. In Proceedings of the AGUFM 2019, San Francisco, CA, USA, 9–13 December 2019; p. C22C-07.
20. Zhang, W.; Witharana, C.; Liljedahl, A.; Kanevskiy, M. Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery. *Remote Sens.* **2018**, *10*, 1487. [[CrossRef](#)]
21. Zhang, W.; Liljedahl, A.K.; Kanevskiy, M.; Epstein, H.E.; Jones, B.M.; Jorgenson, M.T.; Kent, K. Transferability of the deep learning mask R-CNN model for automated mapping of ice-wedge polygons in high-resolution satellite and UAV images. *Remote Sens.* **2020**, *12*, 1085. [[CrossRef](#)]
22. Liljedahl, A.K.; Boike, J.; Daanen, R.P.; Fedorov, A.N.; Frost, G.V.; Grosse, G.; Hinzman, L.D.; Iijma, Y.; Jorgenson, J.C.; Matveyeva, N.; et al. Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology. *Nat. Geosci.* **2016**, *9*, 312–318. [[CrossRef](#)]
23. Turetsky, M.R.; Abbott, B.W.; Jones, M.C.; Anthony, K.W.; Olefeldt, D.; Schuur, E.A.G.; Koven, C.; McGuire, A.D.; Grosse, G.; Kuhry, P. *Permafrost Collapse is Accelerating Carbon Release*; Nature Publishing Group: London, UK, 2019.
24. Jones, B.M.; Grosse, G.; Arp, C.D.; Miller, E.; Liu, L.; Hayes, D.J.; Larsen, C.F. Recent Arctic tundra fire initiates widespread thermokarst development. *Sci. Rep.* **2015**, *5*, 15865. [[CrossRef](#)]
25. Ulrich, M.; Grosse, G.; Strauss, J.; Schirmer, L. Quantifying Wedge-Ice Volumes in Yedoma and Thermokarst Basin Deposits. *Permafr. Periglac. Process.* **2014**, *25*, 151–161. [[CrossRef](#)]
26. Muster, S.; Heim, B.; Abnizova, A.; Boike, J. Water body distributions across scales: A remote sensing based comparison of three arctic tundra wetlands. *Remote Sens.* **2013**, *5*, 1498–1523. [[CrossRef](#)]
27. Lousada, M.; Pina, P.; Vieira, G.; Bandeira, L.; Mora, C. Evaluation of the use of very high resolution aerial imagery for accurate ice-wedge polygon mapping (Adventdalen, Svalbard). *Sci. Total Environ.* **2018**, *615*, 1574–1583. [[CrossRef](#)]
28. Skurikhin, A.N.; Wilson, C.J.; Liljedahl, A.; Rowland, J.C. Recursive active contours for hierarchical segmentation of wetlands in high-resolution satellite imagery of arctic landscapes. In Proceedings of the Southwest Symposium on Image Analysis and Interpretation 2014, San Diego, CA, USA, 6–8 April 2014; pp. 137–140.
29. Abolt, C.J.; Young, M.H.; Atchley, A.L.; Wilson, C.J. Brief communication: Rapid machine-learning-based extraction and measurement of ice wedge polygons in high-resolution digital elevation models. *Cryosphere* **2019**, *13*, 237–245. [[CrossRef](#)]
30. Huang, F.; Zhang, J.; Zhou, C.; Wang, Y.; Huang, J.; Zhu, L. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* **2020**, *17*, 217–229. [[CrossRef](#)]
31. Jiang, P.; Chen, Y.; Liu, B.; He, D.; Liang, C. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access* **2019**, *7*, 59069–59080. [[CrossRef](#)]
32. Samuel, D.; Karam, L. Understanding how image quality affects deep neural networks. In Proceedings of the 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016.
33. Dodge, S.; Karam, L. A study and comparison of human and deep learning recognition performance under visual distortions. In Proceedings of the 2017 26th International Conference on Computer Communication and Networks (ICCCN), Vancouver, BC, Canada, 31 July–3 August 2017; pp. 1–7.
34. Vasiljevic, I.; Chakrabarti, A.; Shakhnarovich, G. Examining the impact of blur on recognition by convolutional networks. *arXiv* **2016**, arXiv:1611.05760.
35. Karahan, S.; Yildirim, M.K.; Kirtac, K.; Rende, F.S.; Butun, G.; Ekenel, H.K. September. How image degradations affect deep cnn-based face recognition? In Proceedings of the 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 21–23 September 2016; pp. 1–5.
36. Gu, Y.; Wang, Y.; Li, Y. A Survey on Deep Learning-Driven Remote Sensing Image Scene Understanding: Scene Classification, Scene Retrieval and Scene-Guided Object Detection. *Appl. Sci.* **2019**, *9*, 2110. [[CrossRef](#)]

37. Russakovsky, O.J.; Deng, H.; Su, J.; Krause, S.; Satheesh, S.; Ma, Z.; Huang, A.; Karpathy, A.; Khosla, M.; Bernstein, A.C.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [\[CrossRef\]](#)
38. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 2012 Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
39. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
40. Raynolds, M.K.; Walker, D.A.; Balser, A.; Bay, C.; Campbell, M.; Cherosov, M.M.; Daniëls, F.J.; Eidesen, P.B.; Ermokhina, K.A.; Frost, G.V.; et al. A raster version of the Circumpolar Arctic Vegetation Map (CAVM). *Remote Sens. Environ.* **2019**, *232*, 111297. [\[CrossRef\]](#)
41. Walker, D.A.F.J.A.; Daniëls, N.V.; Matveyeva, J.; Šibík, M.D.; Walker, A.L.; Breen, L.A.; Druckenmiller, M.K.; Raynolds, H.; Bültmann, S.; Hennekens, M.; et al. Wirth Circumpolar arctic vegetation classification Phytocoenologia. *Phytocoenologia* **2018**, *48*, 181–201. [\[CrossRef\]](#)
42. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask RCNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
43. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Internal Representations by Error Propagation. Available online: https://web.stanford.edu/class/psych209a/ReadingsByDate/02_06/PDPVollChapter8.pdf (accessed on 15 August 2020).
44. Ehlers, M.; Klonus, S.; Åstrand, P.J.; Rosso, P. Multi-sensor image fusion for pansharpening in remote sensing. *Int. J. Image Data Fusion* **2010**, *1*, 25–45. [\[CrossRef\]](#)
45. Wang, S.; Quan, D.; Liang, X.; Ning, M.; Guo, Y.; Jiao, L. A deep learning framework for remote sensing image registration. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 48–164. [\[CrossRef\]](#)
46. Arnall, D.B. Relationship between Coefficient of Variation Measured by Spectral Reflectance and Plant Density at Early Growth Stages. Doctoral Dissertation, Oklahoma State University, Stillwater, OK, USA, 2004.
47. Chuvieco, E. *Fundamentals of Satellite Remote Sensing: An Environmental Approach*; CRC Press: Boca Raton, FL, USA, 2016.
48. Bovolo, F.; Bruzzone, L. A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2963–2972. [\[CrossRef\]](#)
49. Jhan, J.P.; Rau, J.Y. A normalized surf for multispectral image matching and band Co-Registration. International Archives of the Photogrammetry. *Remote Sens. Spat. Inf. Sci.* **2019**. [\[CrossRef\]](#)
50. Inamdar, S.; Bovolo, F.; Bruzzone, L.; Chaudhuri, S. Multidimensional probability density function matching for preprocessing of multitemporal remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1243–1252. [\[CrossRef\]](#)
51. Pitié, F.; Kokaram, A.; Dahyot, R. N-dimensional probability function transfer and its application to color transfer. In Proceedings of the IEEE International Conference Comput Vision, Santiago, Chile, 7–13 December 2005; Volume 2, pp. 1434–1439.
52. Pitié, F.; Kokaram, A.; Dahyot, R. Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.* **2007**, *107*, 123–137. [\[CrossRef\]](#)
53. Liang, Y.; Sun, K.; Zeng, Y.; Li, G.; Xing, M. An Adaptive hierarchical detection method for ship targets in high-resolution SAR images. *Remote Sens.* **2020**, *12*, 303. [\[CrossRef\]](#)
54. McKight, P.E.; Najab, J. Kruskal-wallis test. In *The Corsini Encyclopedia of Psychology*; Wiley: Hoboken, NJ, USA, 2010; p. 1. [\[CrossRef\]](#)
55. Fagerland, M.W.; Sandvik, L. The Wilcoxon-Mann-Whitney test under scrutiny. *Stat. Med.* **2009**, *28*, 1487–1497. [\[CrossRef\]](#)
56. Bhuiyan, M.A.; Nikolopoulos, E.I.; Anagnostou, E.N. Machine learning-based blending of satellite and reanalysis precipitation datasets: A multiregional tropical complex terrain evaluation. *J. Hydrometeor.* **2019**, *20*, 2147–2161. [\[CrossRef\]](#)
57. Bhuiyan, M.A.; Begum, F.; Ilham, S.J.; Khan, R.S. Advanced wind speed prediction using convective weather variables through machine learning application. *Appl. Comput. Geosci.* **2019**, *1*, 100002.

58. Bhuiyan, M.A.E.; Nikolopoulos, E.I.; Anagnostou, E.N.; Quintana-Seguí, P.; Barella-Ortiz, A. A nonparametric statistical technique for combining global precipitation datasets: Development and hydrological evaluation over the Iberian Peninsula. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 1371–1389. [[CrossRef](#)]
59. Bhuiyan, M.A.E.; Yang, F.; Biswas, N.K.; Rahat, S.H.; Neelam, T.J. Machine learning-based error modeling to improve GPM IMERG precipitation product over the brahmaputra river basin. *Forecasting* **2020**, *2*, 14. [[CrossRef](#)]
60. Powers, D. Evaluation: From precision, recall and f-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).