

Article

Automatic Mapping of Thermokarst Landforms from Remote Sensing Images Using Deep Learning: A Case Study in the Northeastern Tibetan Plateau

Lingcao Huang ^{1,*} , Lin Liu ¹ , Liming Jiang ^{2,3} and Tingjun Zhang ^{4,5}

¹ Earth System Science Programme, Faculty of Science, The Chinese University of Hong Kong, Hong Kong SAR, China; liulin@cuhk.edu.hk

² State Key Laboratory of Geodesy and Earth's Dynamics, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China; jlm@whigg.ac.cn

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Key Laboratory of West China's Environment (DOE), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China; tjzhang@lzu.edu.cn

⁵ University Cooperation for Polar Research (UCPR), Beijing 100875, China

* Correspondence: huanglingcao@link.cuhk.edu.hk; Tel.: +852-39439324

Received: 18 November 2018; Accepted: 15 December 2018; Published: 19 December 2018



Abstract: Thawing of ice-rich permafrost causes thermokarst landforms on the ground surface. Obtaining the distribution of thermokarst landforms is a prerequisite for understanding permafrost degradation and carbon exchange at local and regional scales. However, because of their diverse types and characteristics, it is challenging to map thermokarst landforms from remote sensing images. We conducted a case study towards automatically mapping a type of thermokarst landforms (i.e., thermo-erosion gullies) in a local area in the northeastern Tibetan Plateau from high-resolution images by the use of deep learning. In particular, we applied the DeepLab algorithm (based on Convolutional Neural Networks) to a 0.15-m-resolution Digital Orthophoto Map (created using aerial photographs taken by an Unmanned Aerial Vehicle). Here, we document the detailed processing flow with key steps including preparing training data, fine-tuning, inference, and post-processing. Validating against the field measurements and manual digitizing results, we obtained an *F1* score of 0.74 (precision is 0.59 and recall is 1.0), showing that the proposed method can effectively map small and irregular thermokarst landforms. It is potentially viable to apply the designed method to mapping diverse thermokarst landforms in a larger area where high-resolution images and training data are available.

Keywords: DeepLab; permafrost degradation; semantic segmentation; thermokarst landforms; thermo-erosion gullies; Tibetan Plateau; Unmanned Aerial Vehicle Images

1. Introduction

Permafrost is the ground at or below 0 °C for two or more consecutive years. It underlies about 24% of the land surface in the Northern Hemisphere [1]. According to the borehole temperature records, permafrost is undergoing unprecedented warming in both the Arctic and on the Tibetan Plateau [2–7]. Ground warming can cause permafrost degradation mainly through widespread active layer thickening and development of thermokarst landforms [7–10]. Thermokarst landforms and processes (due to thawing of ice-rich permafrost) have complex interactions with hydrology, vegetation, slope aspects, and soil texture in local environment over time, which characterizes the distinctive landforms on the ground [11,12]. Development of thermokarst landforms may accelerate the decay of permafrost carbon and release of greenhouse gases, affecting the global climate [13–16].

The distribution and evolution of thermokarst landforms are essential for understanding the influence of thermokarst on surface hydrological processes, geomorphological processes, carbon exchange, vegetation, and ecosystem as a whole at local, regional, and global scales.

Obtaining the distribution of thermokarst landforms is challenging because they are widespread in permafrost areas and significantly vary in characteristics. According to an estimation based on an empirical framework, the regions with thermokarst landforms occupy approximately 20% of the northern circumpolar permafrost area [14]. More than 20 types of thermokarst landforms, such as thermo-erosion gullies, retrogressive thaw slumps, and thermokarst lakes, spread throughout remote and isolated permafrost areas [11]. Both their size and characteristics vary significantly due to differences in the local environment [17]. Usually, it is necessary to conduct fieldwork, such as using Global Positioning System (GPS) to map thermokarst landforms (e.g., thermo-erosion gullies) in the field [18], or manually identifying them (e.g., retrogressive thaw slumps) on high-resolution satellite images [19–21]. However, field measurements and manual digitizing from remote sensing images are time-consuming and limit the extent of investigation to a small and local area.

Automatic mapping of thermokarst landforms using remote sensing images is a potentially efficient way of obtaining their distribution, especially when targeting a large area such as the Tibetan Plateau. Automatic or semi-automatic methods based on remote sensing images have been developed to monitor the landform changes in permafrost areas. For example, an automated processing chain for the analysis of Landsat images detected permafrost disturbances over large areas in the Arctic [22,23]. Tasseled Cap trend analysis on a stack of Landsat images identified distribution and growth of retrogressive thaw slumps in northwestern Canada [24]. However, it is still challenging to automatically map non-lake thermokarst landforms due to their small sizes and irregular shapes. For instance, thermo-erosion gullies whose size can range from several decimeters to tens of meters in width and from a few meters to several kilometers in length [25,26]. Moreover, these landforms appear subtle or similar to other land cover types such as water tracks on remote sensing images. A few studies successfully utilize high-resolution remote sensing images to map or qualify these landforms [27,28]. However, these studies are limited to small areas due to the data availability or only detected specific landforms.

In this study, we proposed a method based on deep learning to automatically map small, irregular thermokarst landforms from high-resolution remote sensing images. Our method innovatively utilizes a supervised deep learning algorithm building on the Convolutional Neural Networks (CNN) [29]. Deep learning is widely used and has achieved unprecedented and promising results in many fields (such as computer vision and artificial intelligence) in the past few years [30–32]. Moreover, many studies apply deep learning to map land cover, detect objects, and estimate poverty from remote sensing images [33–36]. However, none of them targets thermokarst landforms. The objectives of this study were: (1) to assess the capability of CNN in mapping non-lake thermokarst landforms from high-resolution images; and (2) to analyze the distribution of thermokarst landforms in the study area in the Northeastern Tibetan Plateau.

Due to the small size of thermokarst landforms, we needed high-resolution remote sensing images (whose spatial resolution is higher than one meter). Previous methods use multispectral images for mapping landscape dynamics and thermokarst landforms [22,23,27,28]. However, the ones with high spatial resolution are typically only available from commercial satellites and expensive. Instead of integrating various remote sensing datasets, we only utilized three-bands (i.e., red, green, and blue) images, which are relatively easy to obtain from satellites or airborne platforms. In this paper, we demonstrate the performance of the proposed method and discuss its advantages as well as limitations.

2. Study Area

The study area (100.88°E to 100.92°E, 37.99°N to 38.02°N) is located on the Eboiling Mountain in the northeastern edge of the Tibetan Plateau (Figure 1). The mean elevation is around 3600 m a.s.l.

The area is approximately 6 km². The near-surface soils are covered by a peat layer whose thickness ranges from 0.4 to 4 m [37–39]. Permafrost thickness varies from a few meters to tens of meters; permafrost is ice-rich; and its temperature is within a fraction of a degree from the melting point [40,41]. The average thickness of the active layer is 1.32 ± 0.29 m with a range from 0.81 to 2.1 m [41,42]. The vegetation is dominated by alpine swamp meadow and alpine meadow, where hummocks are well developed and characterize the microtopography in this area [41,43].

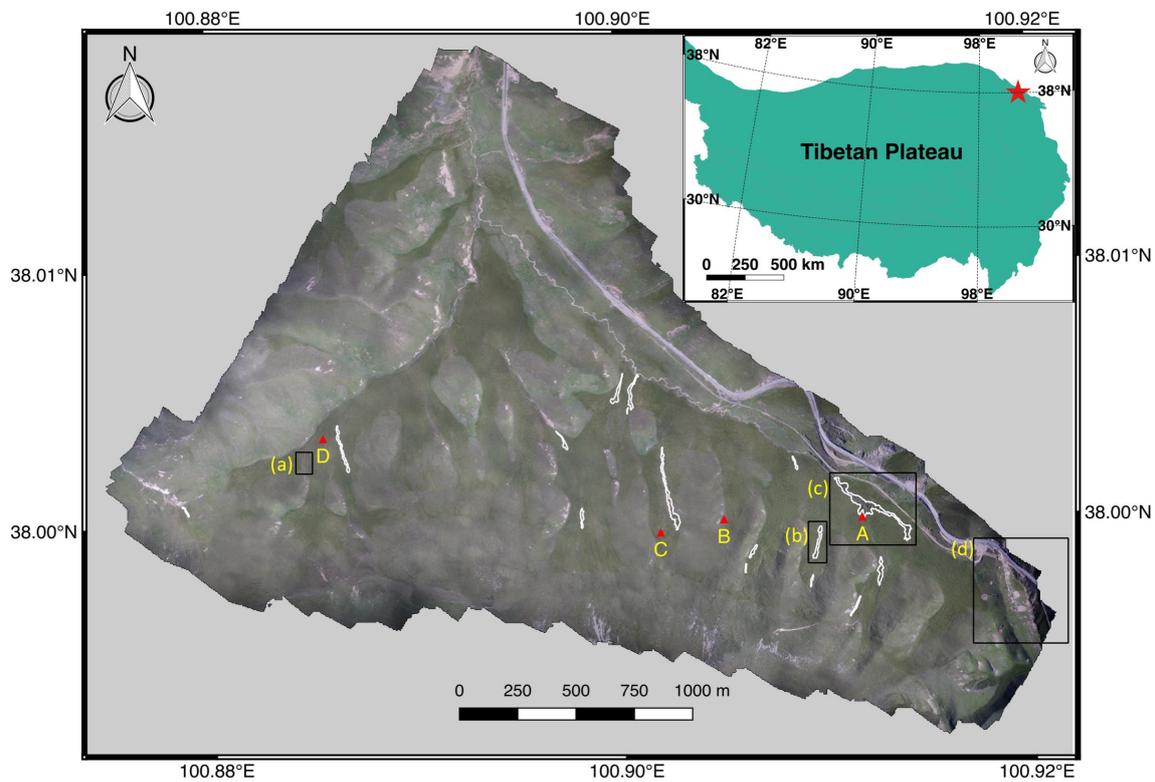


Figure 1. Digital Orthophoto Map (DOM) of the study area. The white polygons are the ground truth polygons. The red triangles and yellow uppercase letters indicate the positions and the names of ground photos shown in Figure 2, respectively. The black rectangles with lowercase letters indicate the extents of the DOM subsets shown in Figure 3. The red star in the inset shows the location of our study area on the Tibetan Plateau.

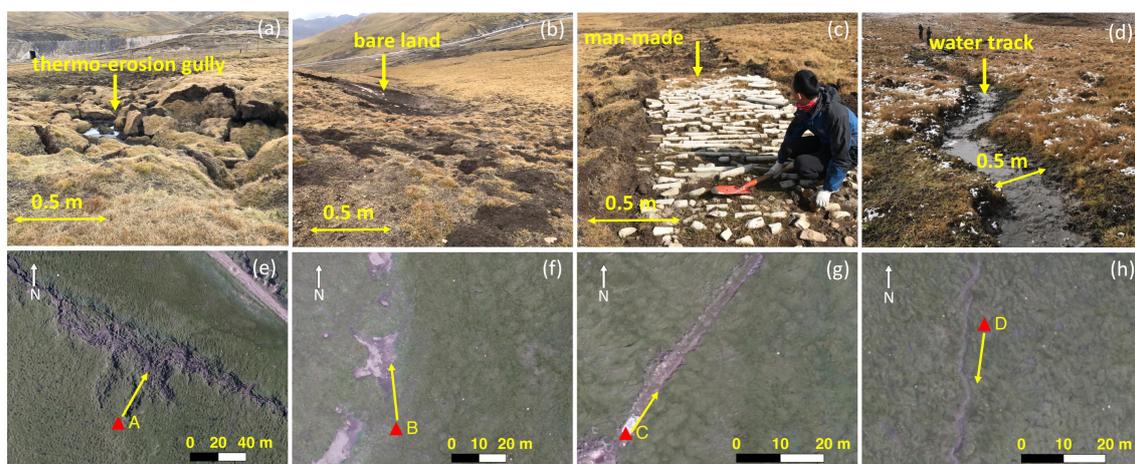


Figure 2. Ground photos (a–d); and their corresponding DOMs (e–h) of different land covers or landforms in the study area. In (e–h), the red triangles indicate the camera positions and the yellow arrows indicate the looking directions.

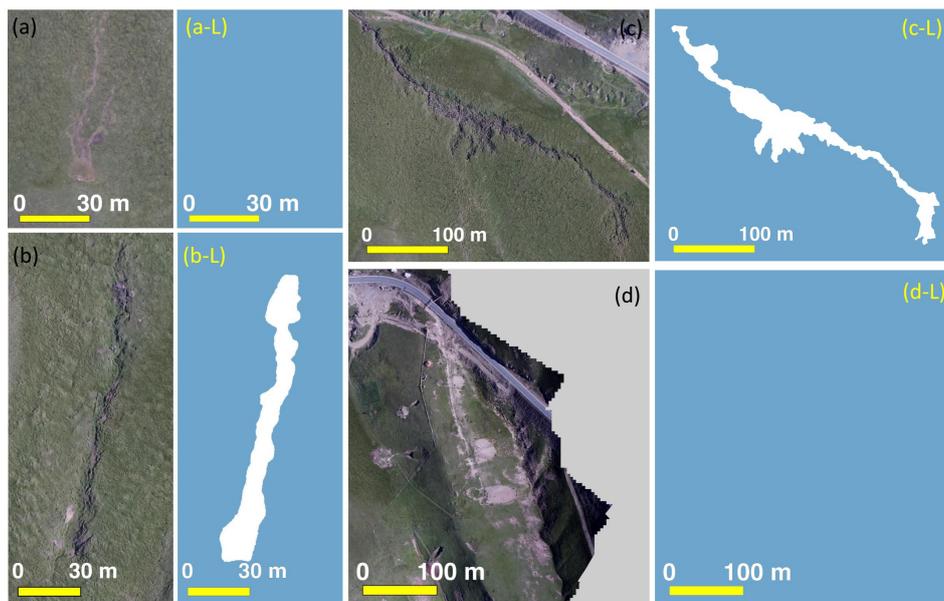


Figure 3. Examples of the DOM subsets used for training: (a,d) examples of sub-images containing non-gully land covers; (b,c) examples of sub-images containing gullies; and (a-L–d-L) their corresponding label images. The white pixels in (a-L–d-L) indicate gullies, while non-gully pixels appear as light blue. A label image appears entirely light blue (e.g., (a-L,d-L)) if the corresponding sub-image does not contain a gully.

Three types of thermokarst landforms, including thermo-erosion gullies, sinkholes, and thermokarst pits, are present in this area. Sinkholes and thermokarst pits are rare and subtle on remote sensing images. Hence, we only target thermo-erosion gullies (hereafter referred to as “gullies”) in this study. These gullies are characterized by erosional features such as collapsed ground, ponds, disturbed vegetation, and exposed soil (Figure 2a). Most of them are narrow features on the images (Figure 2a,e). Their width and length range from around 1 to 30 m and 20 to 500 m, respectively. Most of these gullies also incise more than one meter into the ground (Figure 2a).

In addition to thermokarst landforms, diverse land cover types such as water tracks, bare land, and roads are found in the study area (Figures 1 and 2). Anthropogenic disturbance and natural processes contribute to bare land (Figure 2b,f) and man-made objects (Figure 2c,g). The low permeability of permafrost and saturated organic soil layer lead to many water tracks (Figure 2d), which are also narrow features on remote sensing images (Figure 2h).

3. Methods

3.1. Collection of UAV Images and Creation of Digital Orthophoto Map (DOM)

We collected the UAV images in the field in July 2016. The UAV operated on Trimble UX5 platform and was equipped with Sony A5100 camera. We used the Trimble Access Aerial Imaging application software for flight planning and monitoring as well as for collecting aerial photographs. We deployed 12 ground control points (GCPs) that are evenly distributed within the study area and measured their locations using Real-time Kinematic (RTK) with centimeter-level positioning accuracy. We used the Trimble Inpho UASmaster software to create a DOM using the measured GCPs and to generate a digital elevation model (DEM). Table S1 in the Supplementary Materials lists the software parameters and settings. The final DOM (Figure 1) has a spatial resolution of 0.15 m and contains three spectral bands of red, green, and blue. The DEM covers the entire study area. It was used for assisting the manual digitizing of gullies and simulating stream vectors. The horizontal resolution and vertical accuracy of the DEM are 0.41 m and 0.12 m, respectively.

3.2. Collection of Ground Truth Polygons

The ground truth polygons were collected using the RTK measurements in the field and manually delineated (digitizing) on the DOM. In the field, we walked along the edges of the thermo-erosion gullies and measured the boundaries of nine gullies in July 2016. We considered gully banks with collapse features as parts of gullies. We manually delineated seven gullies in QGIS (version 2.18.14, qgis.org). The UAV-based DEM also helped us to distinguish gullies from water tracks because the first ones incise more than one meter into the ground. We obtained 16 polygons (the white polygons in Figure 1) as the ground truths. Figure S1 in the Supplementary Materials gives the method of obtaining each polygon. We note that their boundaries are 1–2 m away from the gully edges because we could not walk exactly along the edges due to safety concerns.

3.3. Preparation of Training Data

We chose training polygons to prepare images to train our supervised learning method (Figure 4). Although some public datasets such as the PASCAL VOC 2012 images [44] are suitable for training and evaluating deep learning algorithms, none of them contain thermo-erosion gullies. Therefore, we prepared the training data specifically for thermo-erosion gullies. Among these training polygons, 11 of them are from the ground truths and 15 from manually identifying non-gully regions (see Figure S2 in the Supplementary Materials for their distribution and sources). The 11 ground truths are representative of all gullies in the study area. The rest ground truths and the other regions on the DOM were used as test data. The 15 non-gully polygons are essential because they help distinguish the gullies from the similar non-gully land covers during the training. To prove that our choice of training polygons is representative, we conducted 21 (three groups) independent experiments using different combinations of training data, which is similar to bootstrapping in statistics. All experiments had the same setting except their training polygons. The first experiment (first group) used all ground truths and non-gully polygons. Both the second and third groups of experiments randomly selected 70% of ground truths. The second group randomly selected 70% of non-gully polygons while the third group used all of them. Using 70% of ground truths as training data is common in machine learning application. In the third group, the number of training polygons equaled the number of representative selection we described above. Lastly, we compared the results of these 21 experiments with the final results reported in this study.

We utilized the training polygons to extract the subsets of the DOM (termed as sub-images) and to generate the corresponding label images. For each training polygon, we adopted the following two steps to extract a sub-image. (1) We expanded the training polygon by adding a 20-m-wide buffer area. (2) We utilized the minimum bounding rectangle of the expanded area to extract a subset of the DOM by using the Geospatial Data Abstraction Library (GDAL, www.gdal.org). After extracting operation, we obtained 26 sub-images. Among them, 15 contain non-gully land covers such as a water track, bare land, and road, as shown in Figure 3a,d. The other 11 sub-images (e.g., Figure 3b,c) contain both gullies and their surrounding non-gully pixels. To obtain label images, we first rasterized all the ground truth polygons into one single image using GDAL. Then, we extracted 26 label images such as those shown in Figure 3a-L to d-L from this image.

We needed to further subdivide the sub-images and label images into small patches due to the input requirement of our method. The deep learning algorithm (more in Section 3.4) was originally developed to process images from everyday scenes such as the ones in the PASCAL VOC 2012 dataset. These images have small dimensions (on the order of 500×500 pixels) and file sizes (smaller than one megabyte). High-resolution remote sensing images are larger than everyday images. For example, the dimension and file size of the DOM are $28,162 \times 21,003$ pixels and 2.4 gigabytes, respectively. The dimension of sub-images ranges from 313×455 to 3011×2706 pixels depending on the size and shape of the training polygons. Moreover, because of the limitation of Graphics Processing Unit (GPU) and the algorithm, it is not possible to directly input an image with large dimension (such as greater

than 1000×1000 pixels) to the deep learning algorithm. Therefore, we subdivided each sub-image and the corresponding label image into image patches and label patches.

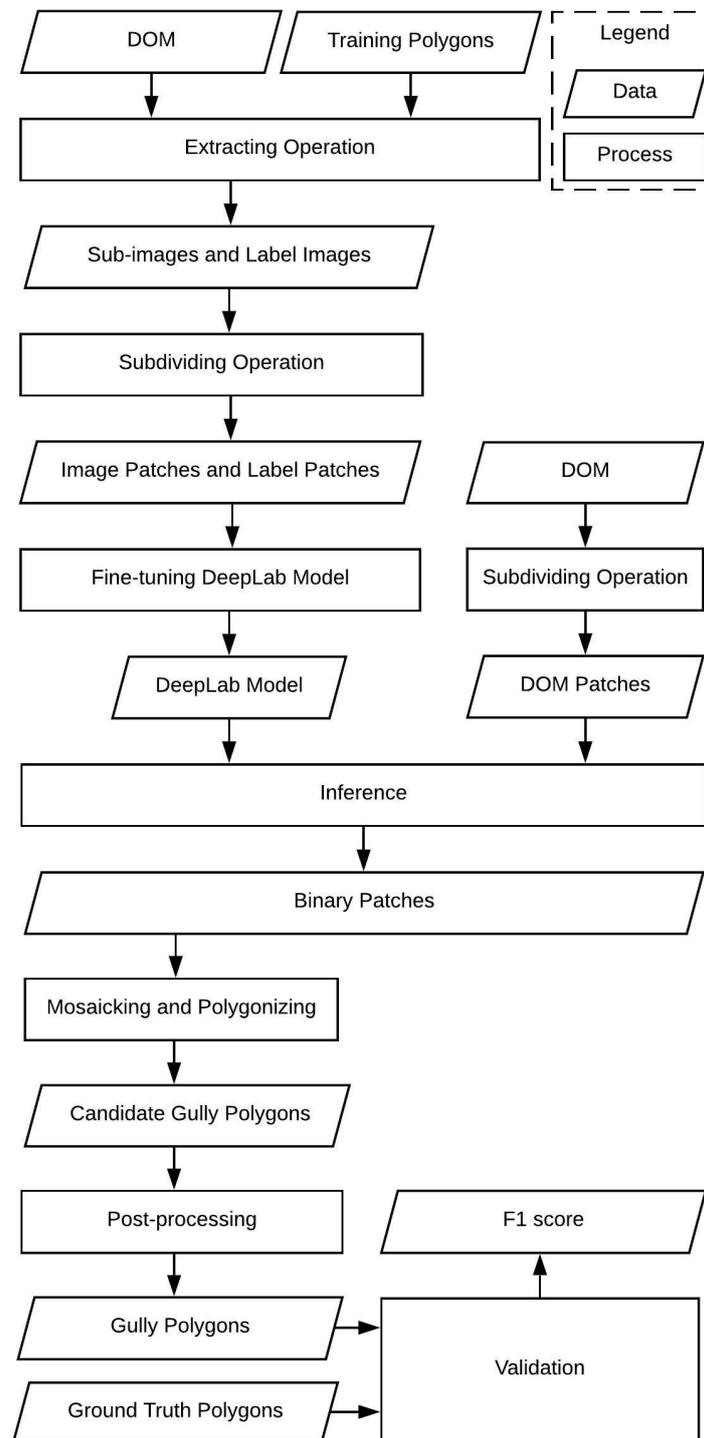


Figure 4. Flowchart of our mapping method based on DeepLab.

We created a grid for each sub-image, and then subdivided sub-images based on their corresponding grids. Figure 5 gives an example of how we subdivided the sub-image shown in Figure 3c. The size of each black square is 321×321 pixels. To utilize the contextual information, we extended each black square in up, down, left, and right directions by 150 pixels (red arrows in Figure 5). The yellow square in Figure 5 shows an example of how we extended one black square (ID #8). The extending operation ensures that each patch overlaps its adjacent ones. Moreover, the overlap is

necessary for producing consistent results among the adjacent patches. Then, we extracted the image patches from the sub-images based on the extended squares. In the case the image patches in the last column or row of the grid are too small (the length of the yellow double arrow is smaller than one-third of the black square), we merged them to the patches in the adjacent column (e.g., #27 into #26 in Figure 5) or row (e.g., #52 into #45 in Figure 5). We did not subdivide those sub-images smaller than 321×321 pixels and considered them as image patches directly. Lastly, we applied the same subdividing operation to label images (e.g., Figure 3c-L) and obtained the corresponding label patches.

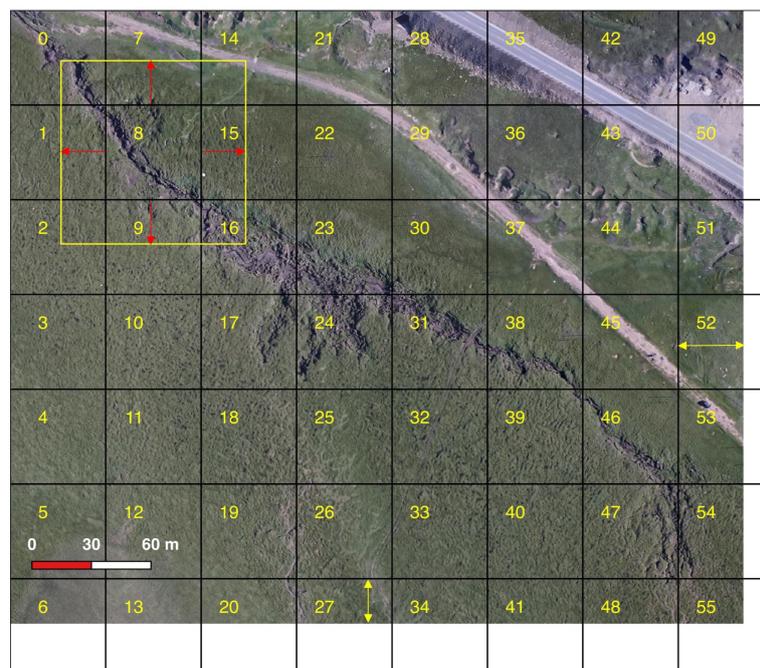


Figure 5. The grid for subdividing a sub-image (same as Figure 3c). The yellow number in each black square is the ID of the subdivided image patch. The yellow square is an example showing how we extend the boundary of a subdivided image patch (#8) from the black square to the yellow square. The yellow arrows in #52 and #27 indicate the patch width and height, respectively.

To increase the number of the training data and the generalization of the deep learning algorithm, we performed data augmentation on the subdivided image patches and the corresponding label patches. The data augmentation included flipping, blurring, and rotating (see Figure 6 for examples of data augmentation).

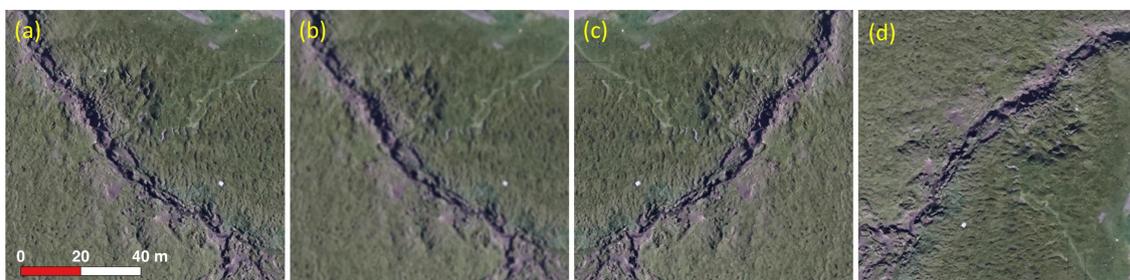


Figure 6. Data augmentation: (a) the patch marked by the yellow square in Figure 5; and (b–d) the patches after blurring, flipping left and right, and rotating 90° clockwise, respectively.

3.4. Mapping of Thermo-Erosion Gullies Using DeepLab

The core of our method is based on DeepLab, a state-of-the-art semantic image segmentation algorithm building on deep CNN [45]. Semantic segmentation implies that the algorithm assigns each

pixel in images to a specific class. We only have two classes, namely gullies and non-gullies, in the semantic segmentation task. DeepLab was developed to label specific objects (e.g., cars, cats, and dogs) on RGB images and outperforms many other deep learning algorithms [45]. Therefore, we developed our method based on DeepLab. We input three-band (i.e., RGB) images to DeepLab, which can learn features automatically. However, it is unclear what features it learned (more discussion in Section 5.5). The network architecture we used is DeepLab-LargeFOV. All parameters in the neural network can be considered as a model.

We fine-tuned the DeepLab model by utilizing image patches and the corresponding label patches. Fine-tuning implies that we initialized the network using a pre-trained model instead of random values, and then we made small adjustments to the model in the training process. The pre-trained model consisted of a set of network parameters that have been obtained by training with the PASCAL VOC 2012 images [45]. It has already learned image features that can be used for identifying targets. Such ability is termed as transfer learning, as the features that have been learned in one setting can be exploited to another setting [46]. By adopting the fine-tuning strategy, we could reduce the training time and lower the requirement of training data quantity. The fine-tuning step took around three hours on one GPU (NVIDIA Quadro P5000) after 6000 iterations and reached convergence (the loss value stabilized at about 0.05).

We utilized the fine-tuned model to map thermo-erosion gullies. We subdivided the entire DOM into many DOM patches (smaller than 650×650 pixels) and performed inference on them. Inference implies that we input a DOM patch into the network with the learned parameters, and then obtain a feature map. The feature map indicates the probability of thermokarst landforms at each pixel. It took around 23 min in the inference step for all the DOM patches on GPU Quadro P5000. We converted the feature map to a binary (gully or non-gully) patch and mosaicked them into a single file for the entire study area. Lastly, we converted (polygonized) binary results to polygons using GDAL. Hereafter, we refer to these polygonized results as “candidate gully polygons”.

3.5. Post-Processing

We set criteria based on the area and shape of thermo-erosion gullies to remove a few erroneously candidate polygons. We set an area threshold of 40 m^2 to remove polygons smaller than it. Moreover, we defined a metric to quantify the narrowness of a gully as:

$$N = \frac{P^2}{S} \quad (1)$$

where N is the narrowness metric, and P and S are the perimeter and area of the polygon, respectively. A larger N means a narrower polygon. We removed the candidate polygons with $N < 40$. The remaining polygons after post-processing are considered as the final output, namely the mapped gully polygons (also simply termed as ‘gully polygons’).

3.6. Validation

We used the ground truth polygons to validate the gully polygons. We assigned each pixel on the DOM to gully or non-gully, but only gullies are our targets. Therefore, we calculated the intersection over union (*IoU*) value of each gully polygons, then counted the numbers of true positives and false positives. We calculated the *IoU* value as:

$$IoU(A, B) = \text{area}(A \cap B) / \text{area}(A \cup B) \quad (2)$$

where A is a gully polygon and B is any ground truth polygon. In cases when multiple *IoU* values exist for one gully polygon (i.e., it intersects with several ground truth polygons), we only keep the maximum one. The value of *IoU* ranges from 0 to 1. A higher *IoU* value means a higher delineation accuracy for a given gully polygon. We set the *IoU* threshold as 0.4, slightly lower than the conventional

threshold of 0.5 because the ground truth polygons contain uncertainties (e.g., their boundaries are wider than the truth ones). If the *IoU* value of a gully polygon is greater than the threshold, we count it as a true positive. Otherwise, it is a false positive. A false negative means that a ground truth polygon does not have any corresponding true positive. The number of false negatives is equal to the number of ground truth polygons minus the number of true positives. We then used these numbers to calculate the *F1* score as:

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = TP / (TP + FN) \quad (4)$$

$$F1 = 2 \times Precision \times Recall / (Precision + Recall) \quad (5)$$

where *TP*, *FP*, and *FN* are the number of true positives, false positives, and false negatives, respectively. The *F1* score provides a quantitative rubric for assessing the overall accuracy of our mapped results. We also conducted a sensitivity analysis on how the delineation accuracy depends on the choices of *IoU* threshold (see Table S2 in the Supplementary Materials). The analysis shows that this threshold is reasonable and produces the best results in terms of the overall *F1* score.

4. Results

From the DOM, our method allowed us to delineate 27 thermo-erosion gullies (Figure 7). Among them, 16 are true positives and 11 are false positives. The number of ground truth polygons is 16, thus *TP* = 16, *FP* = 11, and *FN* = 0. Based on Equations (3)–(5), the precision, recall, and *F1* score are 0.59, 1.0, and 0.74, respectively. Visual inspection of gullies on the DOM (e.g., Figure 8a–d) has shown that the boundaries of most of the gully polygons are close to the ones of the ground truth polygons. The false positives correspond to water tracks and shadows, which appear similar to thermo-erosion gullies (more discussion in Section 5.3).

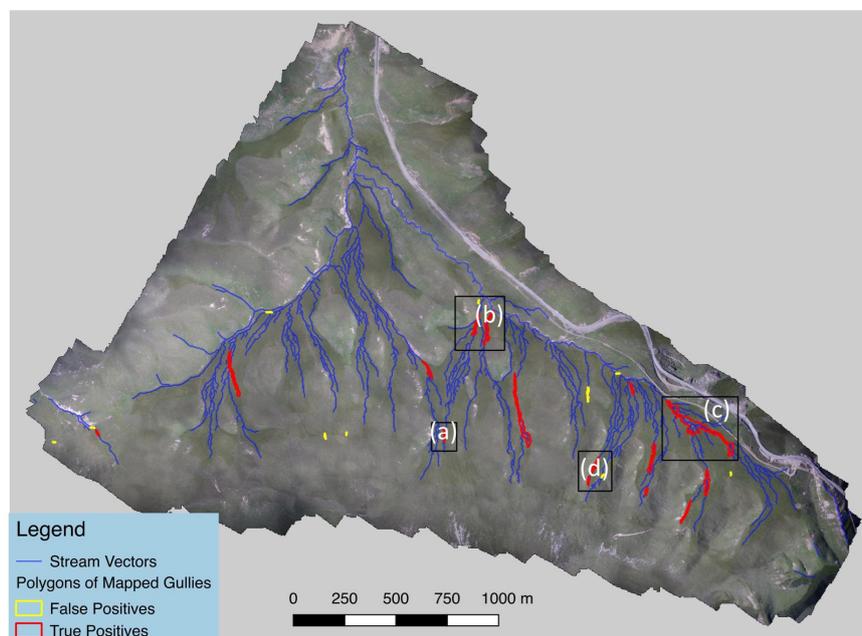


Figure 7. Distribution of the mapped gully polygons (the red and yellow ones are true and false positives, respectively). The blue lines are the simulated surface streams (see Section 5.1). The boxes (a)–(d) denote the extents of the subfigures in Figure 8.

The experiments (Table 1) using varying combinations of training polygons show that the highest achievable *F1* score is within 0.7–0.8, which also proves the robustness of our main results. Moreover, our choice of representative training polygons outperforms all other random combinations in terms of

the $F1$ score and false negatives. Only one experiment (#16) has a higher $F1$ score than our final result. However, its FN equals one, which implies that one ground truth is missed in the result. The first group of an experiment (#1) shows that using all ground truths cannot guarantee the highest $F1$ score. The possible reason is that a larger number (16 gullies vs. 11 non-gullies) of gully samples in the training polygons makes the algorithm more sensitive to gully-like land cover, thus the algorithm produced more false positives. The $F1$ scores in Groups 2 and 3 are similar, which implies that the same percentage of ground truths can result in similar accuracies.

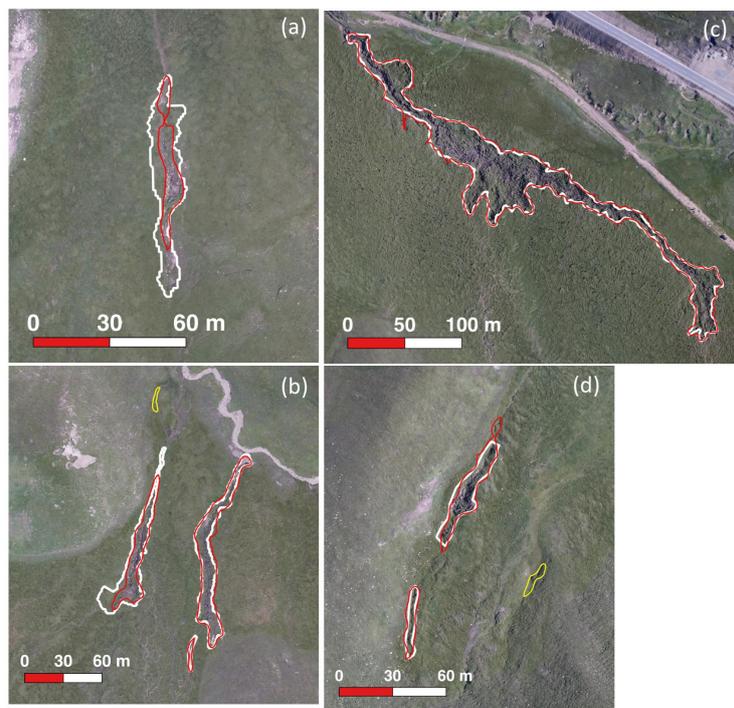


Figure 8. Zoomed-in comparisons between the gully polygons (red) and ground truth polygons (white). The yellow ones are false positives (same as in Figure 7). The extents of (a)–(d) are marked in Figure 7. See more discussion in Sections 5.2 and 5.3.

Table 1. Accuracies of experiments using varying combinations of training polygons.

Group	#	TP	FP	FN	Precision	Recall	F1 Score
1	1	16	17	0	0.48	1.00	0.65
	2	12	15	4	0.44	0.75	0.56
	3	13	13	3	0.50	0.81	0.62
	4	14	12	2	0.54	0.88	0.67
	5	15	18	1	0.45	0.94	0.61
	6	13	22	3	0.37	0.81	0.51
	7	15	17	1	0.47	0.94	0.63
	8	15	15	1	0.50	0.94	0.65
	9	12	22	4	0.35	0.75	0.48
	10	15	20	1	0.43	0.94	0.59
	11	14	15	2	0.48	0.88	0.62
2	12	13	24	3	0.35	0.81	0.49
	13	13	21	3	0.38	0.81	0.52
	14	13	22	3	0.37	0.81	0.51
	15	14	22	2	0.39	0.88	0.54
	16	15	7	1	0.68	0.94	0.79
	17	13	16	3	0.45	0.81	0.58
	18	15	21	1	0.42	0.94	0.58
	19	12	8	4	0.60	0.75	0.67
	20	13	13	3	0.50	0.81	0.62
	21	15	13	1	0.54	0.94	0.68

The thermo-erosion gullies are concentrated at moderate elevations and slopes within the study area. The elevation of gully polygons (true positive) ranges from 3526 m to 3667 m (Figure 9a), while the elevation of the entire study area ranges from 3484 m to 3726 m. Moreover, 94% of the gullies are located on slopes gentler than 17° but steeper than 8° (Figure 9b), while the slope values of more than half (55.6%) of the study area are greater than 17° or smaller than 8°.

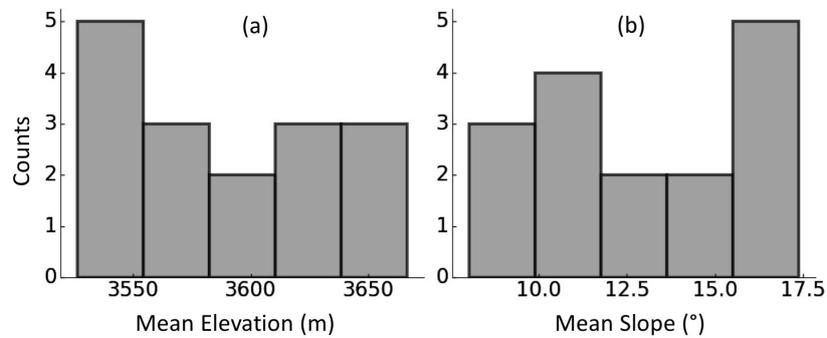


Figure 9. Topographical statistics of the mapped thermo-erosion gullies. (a) and (b) are statistics of mean elevation and mean slope, respectively.

Their circularity indexes (namely, $\frac{4\pi S}{P^2}$) are smaller than 0.25 (Figure 10a). Compared to thermokarst lakes, which have been well studied, the gullies are small: 87% of them are smaller than 2050 m² (Figure 10b) and 94% of them have perimeters shorter than 1000 m (Figure 10c).

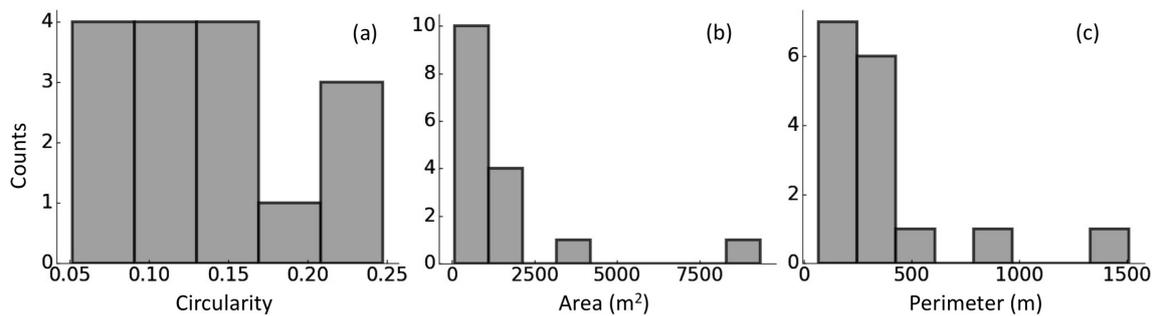


Figure 10. Geometric statistics of the mapped thermo-erosion gullies. (a), (b), and (c) are statistics of circularity, area, and perimeter, respectively.

The *IoU* values of the true positives range from 0.41 to 0.91 (Figure 11) with a mean of 0.73. The minimum and maximum *IoU* values correspond to the gully polygons in Figure 8a,c, respectively.

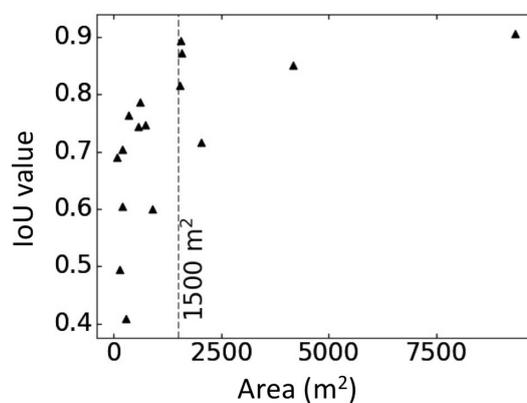


Figure 11. Scatter plot between the *IoU* values of the mapped gullies (the true positives) and their areas.

5. Discussion

5.1. Spatial Distribution of the Thermo-Erosion Gullies

Our results provide insights into the controls of gully appearance. Typically, the spatial distribution of thermo-erosion gullies in the Arctic is controlled by ice-wedge polygon networks and topographic gradient [18]. The growth of ice-wedges is not typical in the study area. Therefore, we focused on examining the topographic gradient as represented by simulated surface streams. We obtained the stream vectors as shown in Figure 7 from a watershed simulation based on the DEM using the Geographic resources analysis support system (GRASS) toolbox (<http://grass.osgeo.org>). Only counting the true positives, we found that all gully polygons are co-located with the simulated surface streams (Figure 7). Two mechanisms can result in such co-location. Firstly, water run-off heats the ground and accelerates the thawing of ice-rich permafrost. Secondly, stream erosion disturbs or even removes the vegetation and soils, which exposes the organic layer to solar radiation and leads to permafrost thawing. We also note that not all the simulated stream vectors have co-located gullies, implying that ground ice content in the near-surface permafrost can also influence the distribution. However, the exact distribution of ground ice is unknown and difficult to quantify.

5.2. Delineation Accuracies of the True Positives

The *IoU* values indicate high delineation accuracies for most of the mapped gullies. A low *IoU* value (e.g., 0.41) could be due to the difference between the microtopography observed in the field and on the DOM. In the field, we considered the collapsed ground on the gully bank as part of the gully; however, the collapse features appear subtle at the gully in Figure 8a as well as a few others on the DOM. Therefore, the ground truth polygons of these gullies are wider than the corresponding mapped polygons from the DOM.

The gully size affects the delineation accuracy. Among the six mapped gullies whose areas are greater than 1500 m², one has an *IoU* value around 0.7 and the other five have *IoU* values greater than 0.8 (Figure 11). Conversely, the *IoU* values of the gullies smaller than 1500 m² are lower than 0.8. Obviously, a bigger gully is easier to be identified from the DOM. However, as shown in Figure 11, the dependency of the *IoU* values on the sizes is non-linear.

5.3. Possible Causes of False Positives

The false positives are mainly delineated due to the limitations of the training data and diverse land covers in our study area. Firstly, the formation of thermo-erosion gullies involves complex thermal, hydrological, and geomorphic processes [18,25]. Some of the gullies in our study area may be at an early stage of development. The training polygons we used may not include all the gullies at different development stages because we only selected the ones that we easily identified on the DOM and in the field. Secondly, the non-gully training polygons we selected may not cover all the land cover types. The study area is close to settlements. The residents and their livestock may have disturbed the land surface at some locations (e.g., vegetation damage and vehicle tracks). These anthropogenic disturbances may have been mistakenly mapped as thermo-erosion gullies because of their similar appearance on the DOM. Lastly, as we pointed out in Figure 8, some water tracks, shadows, and bared land also contribute to the false positives.

5.4. Mapping of Thermo-Erosion Gullies on the DEM

We tested the capability of mapping thermo-erosion gullies on the DEM because it is natural to use DEM data to map geomorphological features, although we wanted to solely use the three band (RGB) images. We applied the same processing chain on the DEM and obtained mapped gully polygons (as shown in Figure S4). In these results, we obtained the accuracy of *F1* score = 0.49 (precision is 0.35 and recall is 0.81). Compared to the results solely using the RGB image, there are more false positives and false negatives. The reasons for this could be: (1) many other landforms or disturbances

are similar to thermo-erosion gullies on the DEM in the study area; and (2) many erosional features (e.g., collapsed ground, ponds, and exposed soil) that are critical for identifying thermo-erosion gullies are not presented on the DEM. Further investigation on using DEM is worth exploring but were not included in this study because including DEM in the method would limit its transferability. In many other regions, collecting remote sensing images with the high spatial resolution from airborne platforms or satellites is easier than obtaining or constructing a high-resolution DEM.

5.5. Advantages and Limitations of Deep Learning for Mapping Thermokarst Landforms

The most significant advantage of deep learning methods is that they can automatically learn features [31]. Hand-designed features are essential and critical in the traditional automatic mapping methods. These methods require time and expertise to design good features and choose the best combination of them. However, with deep learning methods, we can easily extend the method to other application without corresponding expertise. In this study, we only prepared the training images according to the requirement of DeepLab and adopted a network architecture from other application (i.e., DeepLab-LargeFOV). We did not design features or conduct experiments on different combinations of features.

Our method based on deep learning can output results close to manually identified ones. Typically, traditional automatic or semi-automatic mapping methods detect the boundaries based on the differences in the features including edge, texture, and spectral information. Consequently, these methods output results that cannot delineate the thermokarst landform due to the various spectral reflection inside it [27,28]. Figure S3 of the Supplementary Materials gives one example that compares the results obtained from a supervised classification with the ones presented here. Our method can delineate the landforms and acquire high *IoU* values (as discussed in Section 5.2).

However, many studies claim that deep learning is a black box, namely the learned features are unknown, and the factors controlling the output are unclear [47,48]. For instance, we used different data augmentation strategies but found that not all strategies gave the same *F1* score; some even introduced more false positives (and reduced the *F1* score). However, it is unclear how exactly the data augmentation strategies yield the different mapping results. Nonetheless, we need to point out that data augmentation was essential in this study as our training data were limited.

Our deep learning method did not produce exactly repeatable results. We conducted several processes of fine-tuning and inference using the same network setting and training data. However, these processes gave different, yet very similar results; and the *F1* score varied between 0.7 and 0.78. Even if we doubled the number of training iterations, the *F1* score was still not repeatable. We speculate that the random initialization of parameters (of input and output layers) and dropout (which is a strategy to prevent overfitting) led to this phenomenon. Moreover, the small number (16) of ground truth polygons made the *F1* score sensitive to *TP*, *FP*, and *FN*.

5.6. Prospects and Future Work for Mapping Thermokarst Landforms in a Large Area

With deep learning, the achievement of image processing has reached a new stage [30,31]. For example, identification of objects such as cats in images was challenging a few years ago, but it is much easier nowadays. Automatic mapping of thermokarst distribution in permafrost areas was regarded as extremely difficult [49]. However, our results show that it is feasible to map thermokarst landforms on remote sensing images with deep learning algorithms automatically.

Despite the great potentials, we need to make a few major improvements to the current method and training data towards mapping thermokarst landforms at regional scales such as the Tibetan Plateau. When extending the method to a large area, we need to collect corresponding high-resolution remote sensing images and training polygons. Due to the diverse types of thermokarst landforms in a large area, obtaining training data is labor-intensive. Therefore, it may require a community effort (from permafrost scientists and engineers). Variation in the size of thermokarst landforms requires images with different spatial resolution, which will increase image heterogeneities (different sources,

bands, acquisition times, and spatial resolution). Moreover, a large area also means a large volume of data. To overcome these challenges, we need to innovate and improve the algorithm in the following aspects: (1) adopting state-of-the-art algorithms and trained models because deep learning methods advance very quickly; (2) utilizing multiple bands information by integrating the recurrent neural network and CNN; and (3) improving the algorithm to handle the dataset covering the large area by using multiple GPUs.

6. Conclusions

We designed and applied an automatic mapping method based on deep learning to map the thermokarst landforms from high-resolution UAV images. Despite a few false positives, we successfully mapped all 16 thermo-erosion gullies in a local area in the Northeastern Tibetan Plateau (the overall *F1* score is 0.74, precision is 0.59, and recall is 1.0). These gullies are relatively small and narrowly-shaped. All the gullies are co-located with simulated surface streams, confirming that the formation and development of the gullies are largely affected by the surface streams.

In the future, we will consider more factors to quantitatively analyze the spatial distribution of these gullies. We can extend our method to other thermokarst landforms and larger areas by collecting corresponding training data and remote sensing images. With automatic mapping methods, it is possible to map the thermokarst landforms in a large area such as the Tibetan Plateau and the circumpolar region. Products from such comprehensive mapping efforts will significantly improve observations and understanding of permafrost degradation and its environmental and socio-economic impacts.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-4292/10/12/2067/s1>, Figure S1: Map of the ground truth polygons and their obtaining methods. Nine of them were collected in the field using GPS Real-Time Kinematic. Seven of them were manually delineated (digitizing) on the DOM, Figure S2: Distribution of the training polygons: 15 of them are non-gullies; seven were collected from field GPS measurements; and four were from manual delineation (digitizing), Figure S3: Comparison between the DeepLab-based results and those obtained from a supervised classification using support vector machine (SVM). The mapped gullies are marked by filled red polygons on all sub-figures. (a) and (c) are the results of SVM, which contain numerous false positives. (b) and (d) are the results of DeepLab, which only contain a few false positives. The white boundary in (c) and (d) is a ground truth polygon, Figure S4: Distribution of mapped gully polygons (the red and yellow ones are true and false positives, respectively) solely using the UAV-based digital elevation model, Table S1: Parameters and settings of the Trimble Inpho UASmaster software for producing the digital elevation model and digital orthophoto map, Table S2: Delineation accuracies obtained with different IoU thresholds.

Author Contributions: L.H. designed and developed the algorithm of this study, conducted the analysis, and wrote the manuscript. L.L. supervised L.H. and co-designed the algorithm. L.J. and T.Z. provided the data and valuable knowledge and revised the manuscript. All co-authors helped write the manuscript.

Funding: Strategic Priority Research Program of the Chinese Academy of Sciences: XDA19070204, Hong Kong Research Grants Council: CUHK14300815, CUHK Direct Grant for Research: 4053206 and 4053282, Key Research Program of Frontier Sciences, CAS: QYZDB-SSW-DQC027 and XDA19070104.

Acknowledgments: We acknowledge the support of NVIDIA Corporation with the donation of the Quadro P5000 GPU used for this research. We thank the editor and three anonymous reviewers for their constructive and insightful comments. We thank Yafei Sun, Dewei Li, and Daan Li for collecting the RTK data and UAV images; and Jie Chen, Xiyu Chen, Lili Li, and Wen Zhong for helping collect data in the fieldwork of July 2016. We also thank Daan Li and Zhiwei Zhou for producing the DOM; Liang Linlin, Chen Yuxing, Wang GuanXin, Sun Wen and Fan Chengyan for helping collect the ground photos; and Mu Cuicui and Shao Wanwan for arrangement and help in the fieldwork. The work at The Chinese University of Hong Kong was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19070204), the Hong Kong Research Grants Council (CUHK14300815), and CUHK Direct Grant for Research (4053206 and 4053282). The work at the Institute of Geodesy and Geophysics, CAS was supported by the Key Research Program of Frontier Sciences, CAS (QYZDB-SSW-DQC027), and the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19070104).

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

- Zhang, T.; Barry, R.G.; Knowles, K.; Heginbottom, J.A.; Brown, J. Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere 1. *Polar Geogr.* **1999**, *23*, 132–154. [[CrossRef](#)]
- Marchenko, S.S.; Gorbunov, A.P.; Romanovsky, V.E. Permafrost warming in the Tien Shan mountains, central Asia. *Glob. Planet. Chang.* **2007**, *56*, 311–327. [[CrossRef](#)]
- Osterkamp, T.E. Characteristics of the recent warming of permafrost in Alaska. *J. Geophys. Res. Earth Surf.* **2007**, *112*. [[CrossRef](#)]
- Romanovsky, V.E.; Smith, S.L.; Christiansen, H.H. Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: A synthesis. *Permafr. Periglac. Process.* **2010**, *21*, 106–116. [[CrossRef](#)]
- Romanovsky, V.E.; Drozdov, D.S.; Oberman, N.G.; Malkova, G.V.; Kholodov, A.L.; Marchenko, S.S.; Moskalenko, N.G.; Sergeev, D.O.; Ukraintseva, N.G.; Abramov, A.A. Thermal state of permafrost in Russia. *Permafr. Periglac. Process.* **2010**, *21*, 136–155. [[CrossRef](#)]
- Wu, Q.; Zhang, T. Recent permafrost warming on the Qinghai-Tibetan Plateau. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
- Zhao, L.; Wu, Q.; Marchenko, S.S.; Sharkhuu, N. Thermal state of permafrost and active layer in Central Asia during the International Polar Year. *Permafr. Periglac. Process.* **2010**, *21*, 198–207. [[CrossRef](#)]
- Åkerman, H.J.; Johansson, M. Thawing permafrost and thicker active layers in sub-arctic Sweden. *Permafr. Periglac. Process.* **2008**, *19*, 279–292. [[CrossRef](#)]
- Czudek, T.; Demek, J. Thermokarst in Siberia and its influence on the development of lowland relief. *Quat. Res.* **1970**, *1*, 103–120, doi:10.1016/0033-5894(70)90013-X. [[CrossRef](#)]
- Jorgenson, M.T.; Osterkamp, T.E. Response of boreal ecosystems to varying modes of permafrost degradation. *Can. J. For. Res.* **2005**, *35*, 2100–2111. [[CrossRef](#)]
- Jorgenson, M.T. Thermokarst terrains. *Treatise Geomorphol.* **2013**, *8*, 313–324.
- Shur, Y.L.; Jorgenson, M.T. Patterns of permafrost formation and degradation in relation to climate and ecosystems. *Permafr. Periglac. Process.* **2007**, *18*, 7–19. [[CrossRef](#)]
- Grosse, G.; Harden, J.; Turetsky, M.; McGuire, A.D.; Camill, P.; Tarnocai, C.; Froelking, S.; Schuur, E.A.; Jorgenson, T.; Marchenko, S. Vulnerability of high-latitude soil organic carbon in North America to disturbance. *J. Geophys. Res. Biogeosci.* **2011**, *116*. [[CrossRef](#)]
- Olefeldt, D.; Goswami, S.; Grosse, G.; Hayes, D.; Hugelius, G.; Kuhry, P.; McGuire, A.D.; Romanovsky, V.E.; Sannel, A.B.K.; Schuur, E. Circumpolar distribution and carbon storage of thermokarst landscapes. *Nat. Commun.* **2016**, *7*, 13043. [[CrossRef](#)] [[PubMed](#)]
- Schuur, E.; McGuire, A.D.; Schdel, C.; Grosse, G.; Harden, J.W.; Hayes, D.J.; Hugelius, G.; Koven, C.D.; Kuhry, P.; Lawrence, D.M. Climate change and the permafrost carbon feedback. *Nature* **2015**, *520*, 171–179. [[CrossRef](#)] [[PubMed](#)]
- Tarnocai, C.; Canadell, J.G.; Schuur, E.; Kuhry, P.; Mazhitova, G.; Zimov, S. Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **2009**, *23*. [[CrossRef](#)]
- Jorgenson, M.T.; Yoshikawa, K.; Kanevskiy, M.; Shur, Y.; Romanovsky, V.; Marchenko, S.; Grosse, G.; Brown, J.; Jones, B. Permafrost characteristics of Alaska. In Proceedings of the Ninth International Conference on Permafrost, Fairbanks, AK, USA, 20 June–3 July 2008; Volume 3, pp. 121–122.
- Fortier, D.; Allard, M.; Shur, Y. Observation of Rapid Drainage System Development by Thermal Erosion of Ice Wedges on Bylot Island, Canadian Arctic Archipelago. *Permafr. Periglac. Process.* **2007**, *18*, 229–243. [[CrossRef](#)]
- Luo, J.; Niu, F.; Lin, Z.; Liu, M.; Yin, G. Thermokarst lake changes between 1969 and 2010 in the Beilu River Basin, Qinghai–Tibet Plateau, China. *Sci. Bull.* **2015**, *60*, 556–564. [[CrossRef](#)]
- Niu, F.; Lin, Z.; Lu, J.; Luo, J.; Wang, H. Assessment of terrain susceptibility to thermokarst lake development along the Qinghai–Tibet engineering corridor, China. *Environ. Earth Sci.* **2015**, *73*, 5631–5642. [[CrossRef](#)]
- Ramage, J.L.; Irrgang, A.M.; Herzsuh, U.; Morgenstern, A.; Couture, N.; Lantuit, H. Terrain controls on the occurrence of coastal retrogressive thaw slumps along the Yukon Coast, Canada. *J. Geophys. Res. Earth Surf.* **2017**, *122*, 1619–1634. [[CrossRef](#)]
- Nitze, I.; Grosse, G. Detection of landscape dynamics in the Arctic Lena Delta with temporally dense Landsat time-series stacks. *Remote Sens. Environ.* **2016**, *181*, 27–41. [[CrossRef](#)]

23. Nitze, I.; Grosse, G.; Jones, B.M.; Arp, C.D.; Ulrich, M.; Fedorov, A.; Veremeeva, A. Landsat-based trend analysis of lake dynamics across northern permafrost regions. *Remote Sens.* **2017**, *9*, 640. [[CrossRef](#)]
24. Lacelle, D.; Brooker, A.; Fraser, R.H.; Kokelj, S.V. Distribution and growth of thaw slumps in the Richardson Mountains–Peel Plateau region, northwestern Canada. *Geomorphology* **2015**, *235*, 40–51. [[CrossRef](#)]
25. Godin, E.; Fortier, D. Geomorphology of thermo-erosion gullies—case study from Bylot Island, Nunavut, Canada. In Proceedings of the 6th Canadian Permafrost Conference and 63rd Canadian Geotechnical Conference, Calgary, AB, Canada, 12–16 September 2010; pp. 1540–1547.
26. Godin, E.; Fortier, D. Fine Scale Spatio-Temporal Monitoring of Multiple Thermo-Erosion Gullies Development on Bylot Island, Eastern Canadian Archipelago. In Proceedings of the Tenth International Conference on Permafrost (TICOP), Salekhard, Russia, 25–29 June 2012; Volume 1, pp. 1–7.
27. Belshe, E.F.; Schuur, E.; Grosse, G. Quantification of upland thermokarst features with high resolution remote sensing. *Environ. Res. Lett.* **2013**, *8*, 035016. [[CrossRef](#)]
28. Rudy, A.C.; Lamoureux, S.F.; Treitz, P.; Collingwood, A. Identifying permafrost slope disturbance using multi-temporal optical satellite images and change detection techniques. *Cold Reg. Sci. Technol.* **2013**, *88*, 37–49. [[CrossRef](#)]
29. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NY, USA, 3–8 December 2012; pp. 1097–1105.
31. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
32. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359. [[CrossRef](#)]
33. Guo, W.; Yang, W.; Zhang, H.; Hua, G. Geospatial Object Detection in High Resolution Satellite Images Based on Multi-Scale Convolutional Neural Network. *Remote Sens.* **2018**, *10*, 131. [[CrossRef](#)]
34. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [[CrossRef](#)]
35. Jean, N.; Burke, M.; Xie, M.; Davis, W.M.; Lobell, D.B.; Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science* **2016**, *353*, 790–794. [[CrossRef](#)] [[PubMed](#)]
36. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
37. Mu, C.; Zhang, T.; Cao, B.; Wan, X.; Peng, X.; Cheng, G. Study of the organic carbon storage in the active layer of the permafrost over the Eboling Mountain in the upper reaches of the Heihe River in the Eastern Qilian Mountains. *J. Glaciol. Geocryol.* **2013**, *35*, 1–9.
38. Mu, C.; Zhang, T.; Wu, Q.; Cao, B.; Zhang, X.; Peng, X.; Wan, X.; Zheng, L.; Wang, Q.; Cheng, G. Carbon and nitrogen properties of permafrost over the Eboling Mountain in the upper reach of Heihe River basin, Northwestern China. *Arct. Antarct. Alpine Res.* **2015**, *47*, 203–211. [[CrossRef](#)]
39. Mu, C.; Abbott, B.W.; Wu, X.D.; Zhao, Q.; Wang, H.J.; Su, H.; Wang, S.F.; Gao, T.G.; Guo, H.; Peng, X.Q. Thaw Depth Determines Dissolved Organic Carbon Concentration and Biodegradability on the Northern Qinghai-Tibetan Plateau. *Geophys. Res. Lett.* **2017**, *44*, 9389–9399. [[CrossRef](#)]
40. Wang, Q.F.; Zhang, T.J.; Wu, J.C.; Peng, X.Q.; Zhong, X.Y.; Mu, C.C.; Wang, K.; Wu, Q.B.; Cheng, G.D. Investigation on permafrost distribution over the upper reaches of the Heihe River in the Qilian Mountains. *J. Glaciol. Geocryol.* **2013**, *35*, 19–25.
41. Cao, B.; Gruber, S.; Zhang, T.; Li, L.; Peng, X.; Wang, K.; Zheng, L.; Shao, W.; Guo, H. Spatial variability of active layer thickness detected by ground-penetrating radar in the Qilian Mountains, Western China. *J. Geophys. Res. Earth Surf.* **2017**, *122*, 574–591. [[CrossRef](#)]
42. Cao, B.; Zhang, T.; Peng, X.; Mu, C.; Wang, Q.; Zheng, L.; Wang, K.; Zhong, X. Thermal Characteristics and Recent Changes of Permafrost in the Upper Reaches of the Heihe River Basin, Western China. *J. Geophys. Res. Atmos.* **2018**, *123*, 7935–7949. [[CrossRef](#)]
43. Mu, C.; Zhang, T.; Zhang, X.; Li, L.; Guo, H.; Zhao, Q.; Cao, L.; Wu, Q.; Cheng, G. Carbon loss and chemical changes from permafrost collapse in the northern Tibetan Plateau. *J. Geophys. Res. Biogeosci.* **2016**, *121*, 1781–1791. [[CrossRef](#)]

44. Everingham, M.; Eslami, S.A.; Gool, L.V.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [[CrossRef](#)]
45. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv* **2016**, arXiv:1606.00915.
46. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1.
47. Castelvetti, D. Can we open the black box of AI? *Nat. News* **2016**, *538*, 20. [[CrossRef](#)] [[PubMed](#)]
48. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European conference on computer vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 818–833.
49. Council, N.R. *Opportunities to Use Remote Sensing in Understanding Permafrost and Related Ecological Characteristics: Report of a Workshop*; National Academies Press: Washington, DC, USA, 2014.



© 2018 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/>).