



Article

Automated Characterization of Yardangs Using Deep Convolutional Neural Networks

Bowen Gao ¹, Ninghua Chen ^{1,*}, Thomas Blaschke ², Chase Q. Wu ³, Jianyu Chen ⁴, Yaochen Xu ¹, Xiaoping Yang ¹ and Zhenhong Du ¹

¹ Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, School of Earth Sciences, Zhejiang University, Hangzhou 310027, China; gaobw@zju.edu.cn (B.G.); 11838005@zju.edu.cn (Y.X.); xpyang@zju.edu.cn (X.Y.); duzhenhong@zju.edu.cn (Z.D.)

² Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria; thomas.blaschke@sbg.ac.at

³ Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA; chase.wu@njit.edu

⁴ State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China; chenjianyu@sio.org.cn

* Correspondence: geo316@zju.edu.cn; Tel.: +86-571-8795-3086

Abstract: The morphological characteristics of yardangs are the direct evidence that reveals the wind and fluvial erosion for lacustrine sediments in arid areas. These features can be critical indicators in reconstructing local wind directions and environment conditions. Thus, the fast and accurate extraction of yardangs is key to studying their regional distribution and evolution process. However, the existing automated methods to characterize yardangs are of limited generalization that may only be feasible for specific types of yardangs in certain areas. Deep learning methods, which are superior in representation learning, provide potential solutions for mapping yardangs with complex and variable features. In this study, we apply Mask region-based convolutional neural networks (Mask R-CNN) to automatically delineate and classify yardangs using very high spatial resolution images from Google Earth. The yardang field in the Qaidam Basin, northwestern China is selected to conduct the experiments and the method yields mean average precisions of 0.869 and 0.671 for intersection of union (IoU) thresholds of 0.5 and 0.75, respectively. The manual validation results on images of additional study sites show an overall detection accuracy of 74%, while more than 90% of the detected yardangs can be correctly classified and delineated. We then conclude that Mask R-CNN is a robust model to characterize multi-scale yardangs of various types and allows for the research of the morphological and evolutionary aspects of aeolian landform.

Keywords: aeolian landform; yardang; morphological characteristic; deep learning; Mask R-CNN; Google Earth imagery



Citation: Gao, B.; Chen, N.; Blaschke, T.; Wu, C.Q.; Chen, J.; Xu, Y.; Yang, X.; Du, Z. Automated Characterization of Yardangs Using Deep Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 733. <https://doi.org/10.3390/rs13040733>

Academic Editor: Carl Legleiter

Received: 31 December 2020

Accepted: 10 February 2021

Published: 17 February 2021

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



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

1. Introduction

Yardangs are wind-eroded ridges carved from bedrock or cohesive sediments, and are found broadly in arid environments on Earth [1–4], other planets including Mars [5–7], Venus [8,9] and Saturn's largest moon Titan [10]. Yardangs usually occur in groups and display a wide variety in scales and morphologies in different locations [11]. The morphologies of yardangs are the feedback results of the topographies, wind regimes and sediment types when they formed, which makes them critical paleoclimatic and paleoenvironmental indicators [12–14]. Significant work has been done regarding yardang morphologies and their spatial distributions, evolution processes and controlling factors [5,12,15–21].

The rapid development of remote sensing (RS) technologies has enabled the observation and measurement of yardangs at various spatial scales. De Silva et al. [22] studied the relationship between the yardang morphologies and the material properties of the host lithology using multi-source RS data combined with field surveys. Al-Dousari et al. [12]

produced a geomorphological map in the Um Al-Rimam depressions with all of the yardang-like morphologies based on a series of aerial photos with 60 cm resolution. Li et al. [20] mapped four types of yardangs in the Qaidam Basin from Google Earth images and analyzed the factors that control their formation and evolution. In addition, Wang et al. [7] used Google Earth images and unmanned aerial vehicle (UAV) images to apply a refined classification of yardangs to the same region. Hu et al. [21] manually extracted the geometries of yardangs such as length to width (l/w) ratios, orientations, spatial density and spacing to analyze their controlling factors. Xiao et al. [23] conducted an ultra-detailed characterization of whaleback yardangs using centimeter-resolution orthomosaics and digital elevation models (DEMs) derived from UAVs.

Due to the complicated morphometric features of yardangs such as multi-scaled forms and irregular shapes, the characterization of yardangs has relied heavily on visual interpretation and manual digitalization. These methods, however, are time-consuming and limit the size of a study area. To solve this problem, some researchers have proposed semi-automated approaches to extract yardangs from imagery. For example, Ehsani and Quiel [24] applied the Self Organizing Map (SOM) algorithm to characterize yardangs in the Lut Desert based on Shuttle Radar Topography Mission (SRTM) data. Zhao et al. [25] used the Canny edge detection (CED) algorithm to extract yardangs from Landsat 8 images. The ellipse-fitting algorithm was then used to calculate the morphological parameters of yardangs on 1.7 m DEM generated by UAV images. Yuan et al. [26] developed a semi-automated procedure to characterize yardangs from Google Earth images using object-based image analysis (OBIA) and CED. Although these methods can be used to extract yardangs from imagery, experimental parameter setting and adjustment (e.g., the scale of the segmentation in OBIA) are still needed in the application. This means that they may just be suitable for a fixed type of yardangs within a relatively small area. To detect and characterize yardangs of various forms and scales, there is a strong need to design a self-adaptive, fully automated method with general applicability. This method ought to quickly extract the yardangs' morphological features without requiring a priori specification of the scale. This will help geomorphologists in the future to study the formation mechanism and evolution process of yardangs from a pan-regional view.

Deep learning (DL) approaches [27] have recently received great attention in remote sensing and are widely utilized in tasks including pixel-level classification, object detection, scene understanding and data fusion [28–30]. In DL, convolutional neural networks (CNNs) are most commonly applied to analyze visual imagery. They use a series of convolutional layers to progressively extract higher level features from the input images. In contrast to conventional machine learning methods, CNNs have deeper and more complicated architectures with a huge number of parameters, resulting in a high capability of feature extraction [31]. Specifically in geomorphological research, many CNN-based methods are used for terrain feature detection and landform mapping. Li and Hsu [32] proposed a DL network based on a Faster region-based convolutional neural networks (R-CNN) model to identify natural terrain features (e.g., craters, sand dunes, rivers). However, simply localizing the terrain features with bounding boxes is not sufficient to study landform processes and their impacts on the environment. The introduction of fully convolutional neural networks (FCNs) could increase the accuracy of the result by upsampling images back to their original sizes via deconvolution, which allows the feature map to be produced, with each pixel belonging to its class label [33,34]. A number of the state-of-the-art semantic segmentation models such as UNet and DeepLabv3+ use the FCN architectures, while other authors also applied these models to map landforms by roughly delineating their borders [35–37]. Beyond that, DL shows a high ability to detect and delineate each distinct object at a pixel level in an image, also known as instance segmentation [38–40]. This means it has major potential for the precise characterization of micro-landforms and geological bodies while utilizing the advantages of very high spatial resolution (VHSR) remote sensing data.

In this study, we use the instance segmentation method to distinguish individual yardang as a separate instance and specifically adopt the Mask R-CNN model proposed by He et al. [39] to conduct the experiment. The Mask R-CNN is an extension of the Faster R-CNN [41], integrating the object detection and segmentation tasks by adding a mask branch into the architecture. This model can be described as a two-stage algorithm. Firstly, it generates the candidate object bounding boxes on feature maps produced by the convolutional backbone. Secondly, it predicts the class of the object, refines the bounding box and generates a binary mask for each region of interest (RoI). In the remote sensing domain, the Mask R-CNN is usually applied to detect man-made objects such as buildings [42], aircraft [43] and ships [44]. However, it has hardly been used for terrain feature extraction. Zhang et al. [45] used this method to map ice-wedge polygons from VHSR imagery and systematically assessed its performance. They reported that the Mask R-CNN can detect up to 79% of the ice-wedge polygons in their study area with high accuracy and robustness. Chen et al. [46] detected and segmented rocks along a tectonic fault scarp by using Mask R-CNN and estimated the spatial distributions of their traits. Maxwell et al. [47] extracted valley fill faces from LiDAR-derived digital elevation data using the Mask R-CNN. They used this method to map geomorphic features using LiDAR-derived data but stated poor transferability for photogrammetrically derived data. It should be noted that in former research, the spatial scales of geomorphic objects and their shape irregularity have not changed significantly.

We conclude that there is a need to investigate these methods for aeolian landforms and specifically for multi-scaled yardangs with various complexity. Therefore, the main purposes of this study are: (1) to automatically characterize and classify three types of yardangs using the Mask R-CNN model on VHSR remote sensing images provided by Google Earth; (2) to evaluate the accuracy and the generalization capacity of the model by applying it to the test dataset and extra image scenes at different sites. In this study, we will separately assess the correctness of detection, classification and delineation tasks, and analyze their performances regarding the different spatial resolutions (0.6, 1.2, 2.0 and 3.0 m) of the images.

2. Materials and Methods

2.1. Study Area and Data

The Qaidam Basin, located in the northeastern part of the Tibetan Plateau, is a hyper-arid basin with an average elevation of 2800 m above sea level. Covering an area of approximately 120,000 km², it is the largest intermontane basin enclosed by the Kunlun Mountains in the south, the Altyn-Tagh Mountains in the northwest and the Qilian Mountains in the northeast [48]. The East Asian Monsoons, which extend to the southeastern basin, carry the main moisture source and cause a gradual reduction in the annual precipitation from the southeast to the northwest, decreasing from 100 mm to 20 mm [7,25]. The northwesterly to westerly winds prevailing in the basin sculpt enormous yardang fields, covering nearly 1/3 of the whole basin [19,49]. The basin is also the highest-elevation yardang field on Earth.

The yardangs are mainly distributed in the central eastern and northwestern parts of the basin. The classification schemes for yardangs are different depending on the geographical regions. In this study, we focused on three major types—long-ridge, mesa and whaleback yardangs that are widely developed in the Qaidam Basin. Table 1 illustrates their characteristics that can be distinguished in remotely sensed images.

Table 1. Characteristics of the three main types of yardangs.

Type	Characteristic
Long-ridge	Long strip shape with flat and narrow tops and convex flanks
Mesa	Irregular in shape and unclear orientations with flats top and steep sides
Whaleback	Teardrop shape with blunt heads and tapered tails, some with sharp crests on their backs

In this study, we used very high spatial resolution images from Google Earth. Due to the vast and uneven distribution of yardangs, we firstly selected 24 rectangular image subsets of different areas that host representative yardangs (Figure 1). Ten of these image subsets with a spatial resolution of 1.19 m target areas with long-ridge yardangs, which are usually longer than 200 m. They are mainly distributed in the northernmost part of the basin. The other 14 subsets with resolution of 0.60 m cover areas where a significant amount of mesa and whaleback yardangs are located. All of the image subsets are composed of R (red) G (green) B (blue) bands and were downloaded using LocaSpaceViewer software.

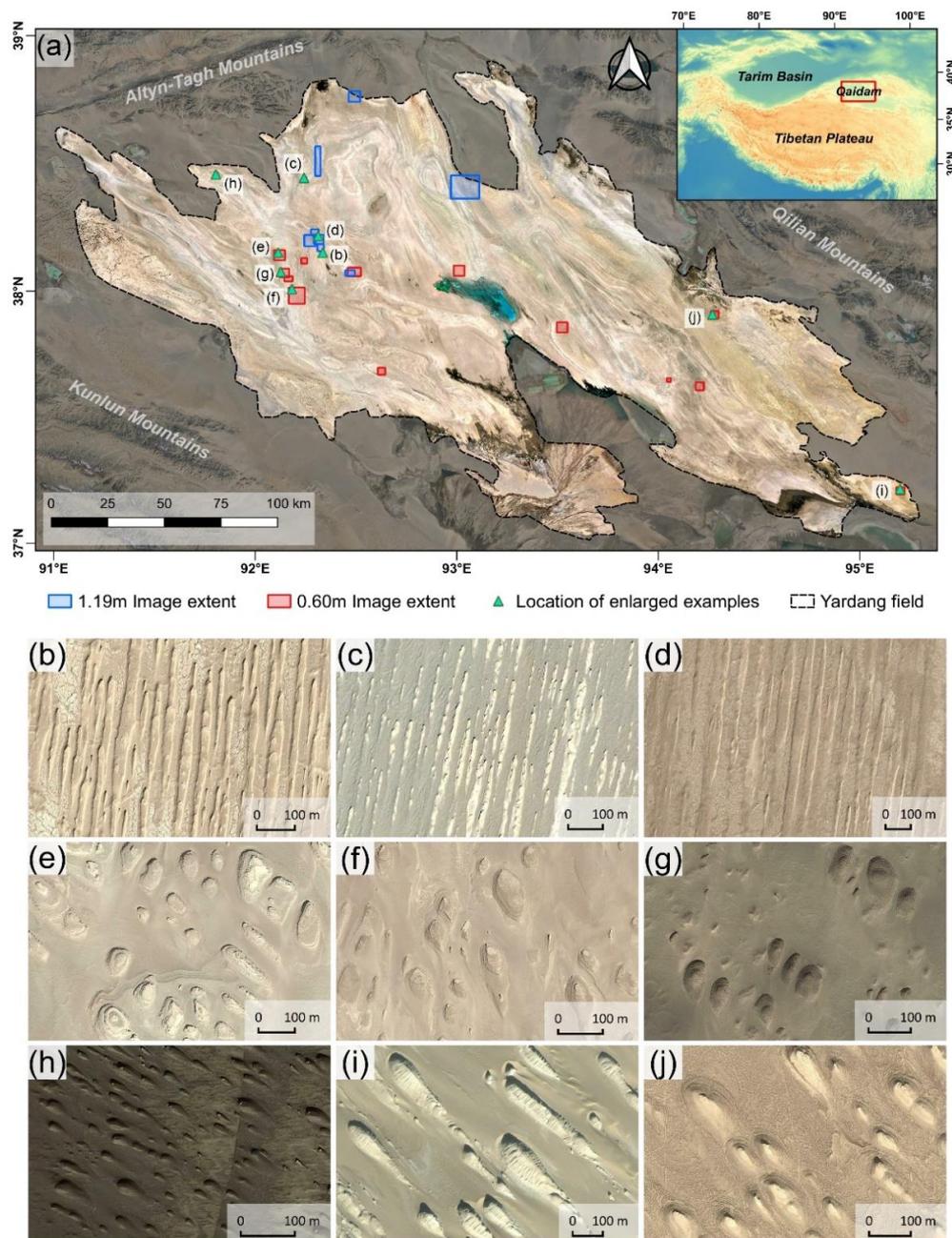


Figure 1. The study area and examples of yardangs. (a) Geographical settings of Qaidam Basin and locations of the original Google Earth image subsets. Images with resolution of 0.60 m and 1.19 m are displayed in red and blue rectangles, respectively. Enlarged images (true color) displayed examples of three types of yardangs, which are (b–d) long-ridge yardangs, (e–g) mesa yardangs and (h–j) whaleback yardangs.

2.2. Annotated Dataset

The DL model used in this study requires rectangular image subsets as the input. In order to generate an adequate dataset that can be used to train and test the Mask R-CNN algorithm, all 24 image subsets were clipped into 512×512 pixel image patches with an overlap of 25% in their X and Y directions, respectively, using Geospatial Data Abstraction Library (GDAL, <https://gdal.org/> (accessed on 31 December 2020)). The image patch containing at least one instance of yardang object with clear and complete boundary was considered valid. Finally, 899 image patches were selected for further manual annotation. We used the “LabelMe” image annotation tool [50] to annotate all the yardangs objects using polygonal annotations. For the instance segmentation step, every single object instance was required to be differentiated. Therefore, we distinguished each instance of the same class by adding sequence numbers after their class names. We delineated an approximately even number of yardang objects for long-ridge, mesa and whaleback yardangs (2568, 2539 and 2532 polygonal objects, respectively) to avoid class imbalance problems [51]. The annotations were saved in the format of JavaScript Object Notation (.json) and were converted into useable datasets using Python. These datasets were then split at random into three sub-datasets at the proportion of 8:1:1. The resulting training dataset consisted of 719 patches, the validation dataset yielded 90 patches and the test dataset yielded another 90 patches.

2.3. Mask R-CNN Algorithm

The Mask R-CNN has been one of the most popular deep neural network algorithms and aims to complete object instance segmentation [39]. As an extension of the Faster R-CNN [41], a branch that outputs the object mask was constructed in the Mask R-CNN, resulting a much finer localization and extraction of objects. The Mask R-CNN is a two-stage procedure. In the first stage, the Residual Learning Network (ResNet) [52] backbone architecture extracts features from raw images, and the Feature Pyramid Network (FPN) [53] improves the representation of the object by generating multi-scale feature maps. Next, the Region Proposal Network (RPN) scans the feature maps and proposes regions of interest (RoI) which may contain objects. In the second stage, the proposed RoIs will be assigned to the relevant area of the corresponding feature map and generate the fixed size feature map via RoIAlign. Then for each RoI, the classifier predicts the class and refines the bounding box whilst the FCN [33] predicts the segmentation mask. For a detailed discussion of the Mask R-CNN, readers can consult He et al. [39].

2.4. Implementation

In this study, we specifically used an open-source package developed by the Matterport team to implement the Mask R-CNN method [54]. The model is built on Python 3, Keras and TensorFlow, and the source codes are available on GitHub (https://github.com/matterport/Mask_RCNN (accessed on 31 December 2020)). The experiments were conducted on a personal computer equipped with an Intel i5-8600 CPU, 8 GB of RAM and a NVIDIA GeForce GTX 1050 5 GB graphic card.

To train the model on our dataset, we created a subclass from the Dataset class and modified the loading dataset function provided in the source codes. We concluded from many studies [45,47,55–57] that the utilization of pre-trained weights learned from other tasks to initialize the model will benefit the training efficiency and result. Although different data and classes are used and targeted in other tasks (e.g., everyday object detection), some common and distinguishable features can be learned and applied in another disparate dataset (known as transfer learning) [30,56]. As a result, we used the pre-trained weights learned from the Microsoft Common Object in Context (MS-COCO) dataset [58] to train our model instead of building the Mask R-CNN from scratch. In the training process, we adopted the ResNet-101 as the backbone and used a learning rate of 0.002 to train the head layers for 2 epochs. This was followed by training all layers at the learning rate of 0.001 for 38 epochs. The default learning momentum of 0.9 and the weight decay of 0.0001 were

used for all epochs. We adopted 240 training and 30 validation steps for each epoch with a mini-batch size of one image patch. We set all the losses (RPN class loss, RPN bounding box loss, Mask R-CNN class loss, Mask R-CNN bounding box loss and Mask R-CNN mask loss) to be equally weighted and kept the default configuration for other parameters. The full description of parameter settings can be found in the above-mentioned Matterport Mask R-CNN repository. Data augmentation is a strategy to artificially increase the amount and diversity of data and is commonly used to minimize overfitting during the training for neural networks [28,30,59–61]. Therefore, we applied random augmentations for training dataset including horizontal flips, vertical flips and rotations at 90°, 180° and 270° using the imgaug library [62]. The validation dataset was used to avoid overfitting and to decide which model weights to use. Once a final model was obtained, it was used to detect yardangs in the test datasets and in other study sites. The detection confidence threshold was set at 70%, which meant that objects with confidence scores below 0.7 were ignored.

2.5. Accuracy Assessment

We assessed the trained Mask R-CNN model on the test dataset based on the mean average precision (mAP) over the different intersection of union (IoU) thresholds ranging from 0.50 to 0.95 in 0.05 increments. The mAP computes the mean average precision value in each category over multiple IoUs and represents the area under the precision-recall curve. The IoU is the ratio of the intersection area and the union area of predicted mask and ground truth mask of an object [61,63]. This can be described in Equation (1) where A is the inferred polygon of yardang and B is the manually digitalized one.

$$\text{IoU}(A, B) = \frac{\text{area}(A \cap B)}{\text{area}(A \cup B)} \quad (1)$$

We calculated the IoU value for each image patch in the test dataset while the confusion matrix was used to analyze the efficiency of the method on specific categories. In order to test the transferability and robustness of the Mask R-CNN method, we applied the model on 80 randomly selected image subsets (spatial resolution of 0.60 m and spatial extent of 1 km × 1 km) excluding previously annotated datasets in the Qaidam Basin (Figure 2). We then manually assessed the accuracies of detecting, classifying and delineating yardangs. The criteria of accuracy evaluation of the three sub-tasks are as follows:

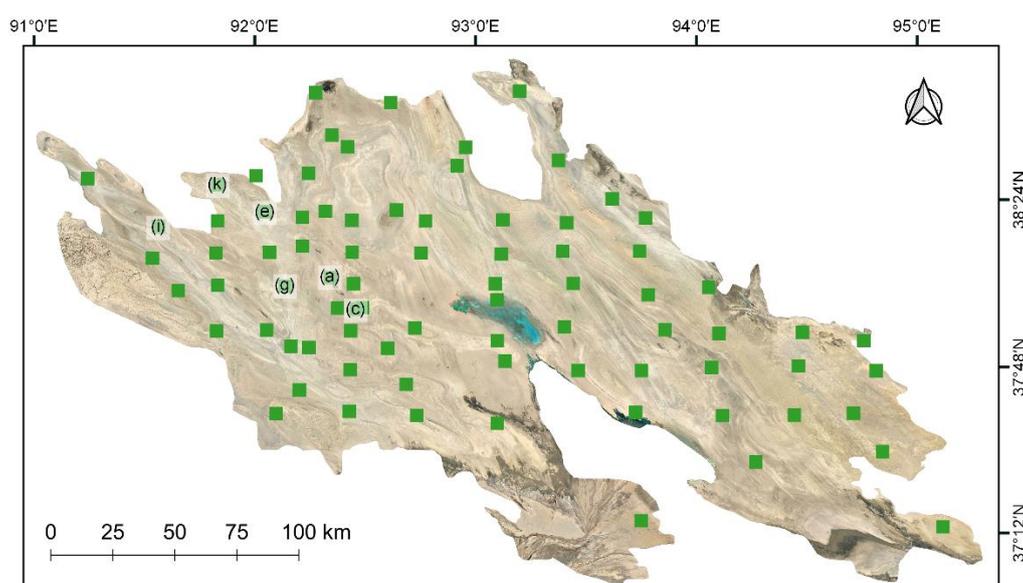


Figure 2. The 80 case study sites (marked by green squares) selected in the yardang field of the Qaidam Basin. The squares marked with letters correspond to the enlarged image scenes in Figures 5–8.

For the detection task, a true positive (TP) indicates a correctly detected yardang and a false positive (FP) indicates a wrongly detected yardang, which in fact belongs to the background. A missed yardang ground truth is a false negative (FN). We then calculated the precision, recall and overall accuracy (OA) of the method for detecting yardangs using the equations below.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Overall Accuracy} = \frac{TP}{TP + FP + FN} \quad (4)$$

For classification and delineation tasks, we focused on those instances that were correctly detected as yardangs. A positive classification refers to a correct type of a detected yardang. Likewise, a positive delineation means the Mask R-CNN successfully outlined a yardang based on the interpreter's judgment. We then calculated the accuracies of these tasks.

3. Results

3.1. Model Optimization and Accuracy

We recorded the learning curve of the model and evaluated its accuracy. As shown in Figure 3, the overall loss is the sum of the other five losses. The validation loss shows a fluctuation and reaches its lowest value after 27 epochs (0.527). Although the training loss presents a continuous decline trend, it declines very slowly and no substantial improvement of validation loss is recorded in the following epochs. This indicates overfitting, where after 27 epochs, the model tends to learn the detail and noise in the training data. This negatively affects the performance of detecting yardangs. Therefore, we selected the Mask R-CNN model that was trained for 27 epochs. The convergence after several epochs was also noted in some related works [45,47,57]. It can be attributed to the usage of pre-trained weights and the relatively small amount of training data for a specific object, which have fewer features to learn compared to a common dataset.

We applied the final model on the test dataset that contained 90 image patches including 768 yardangs (291 long-ridge yardangs, 263 mesa yardangs and 214 whaleback yardangs). The total numbers of predicted long-ridge, mesa and whaleback yardangs are 306, 285 and 197, respectively. The IoU measures how much the predicted yardang boundaries overlap with the ground truth. We received a mean IoU value of 0.74 for the test dataset and 90% of the test patches yielded IoU values greater than 0.5 (Figure 4a). Mean average precisions (mAPs) at different IoU thresholds are plotted in Figure 4b, where the averaged mAP of 10 IoU thresholds is 0.588. The mAP⁵⁰ (mAP at IoU = 0.50) is 0.869, indicating that the model performs well with a common standard for object detection. The mAP⁷⁵ (mAP at IoU = 0.75) is 0.671 when a more stringent metric is used, showing a relatively high accuracy and the effectiveness of this model. The normalized confusion matrix (Figure 4c) displays the fine classification results of yardangs, showing that three types of yardangs can be clearly distinguished. However, the prediction accuracy of the whaleback yardangs is 0.76, which is the lowest value compared to the other two types. The morphological features of whaleback yardangs are more complex due to their diverse variance in sizes and shapes [7], which results in relatively more false and failed detections of this kind of yardangs.

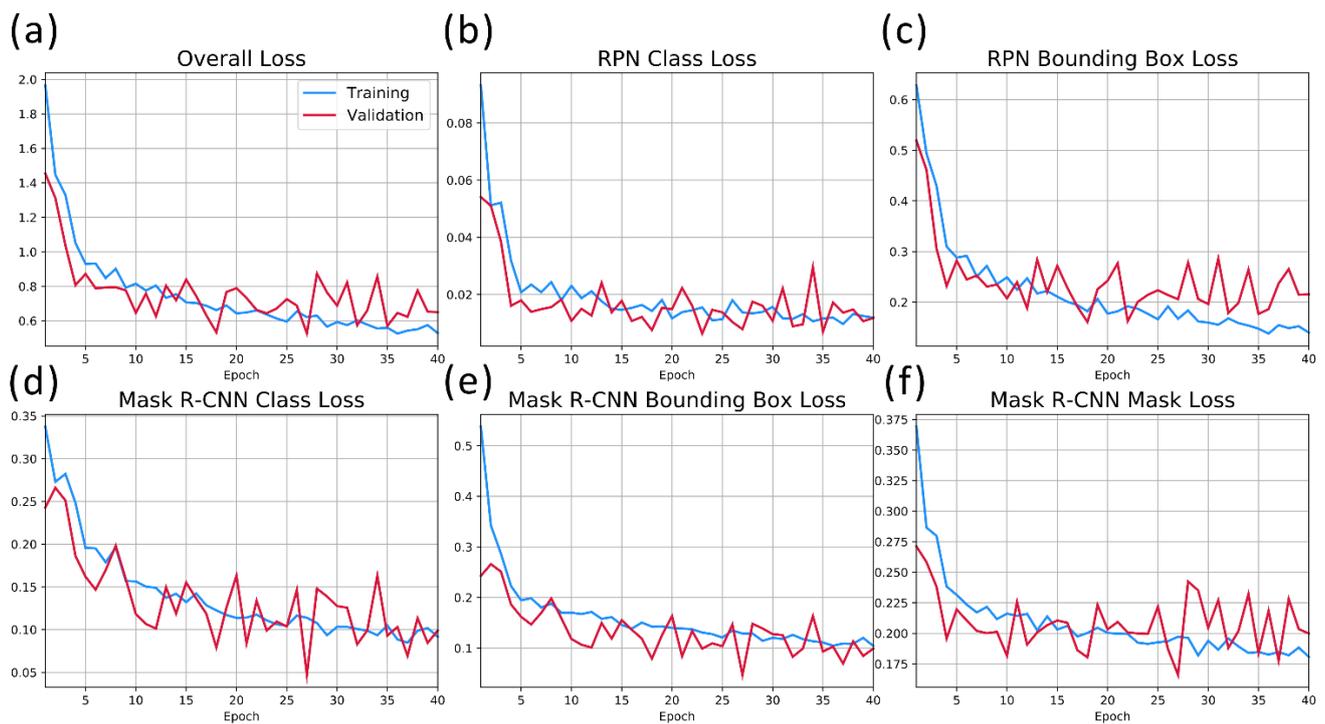


Figure 3. Loss curves of Mask region-based convolutional neural networks (Mask R-CNN) on training and validation datasets. (a) Overall loss is the sum of losses in (b–f). (b) Region Proposal Network (RPN) class loss measures how well the RPN anchor classifier separates the objects with background. (c) RPN bounding box loss represents how well the RPN localizes objects. (d) Mask R-CNN class loss is used to measure how well the Mask R-CNN recognizes the category of object. (e) Mask R-CNN bounding box loss measures how well the Mask R-CNN localizes objects. (f) Mask R-CNN mask loss evaluates how well the Mask R-CNN segments each object instance.

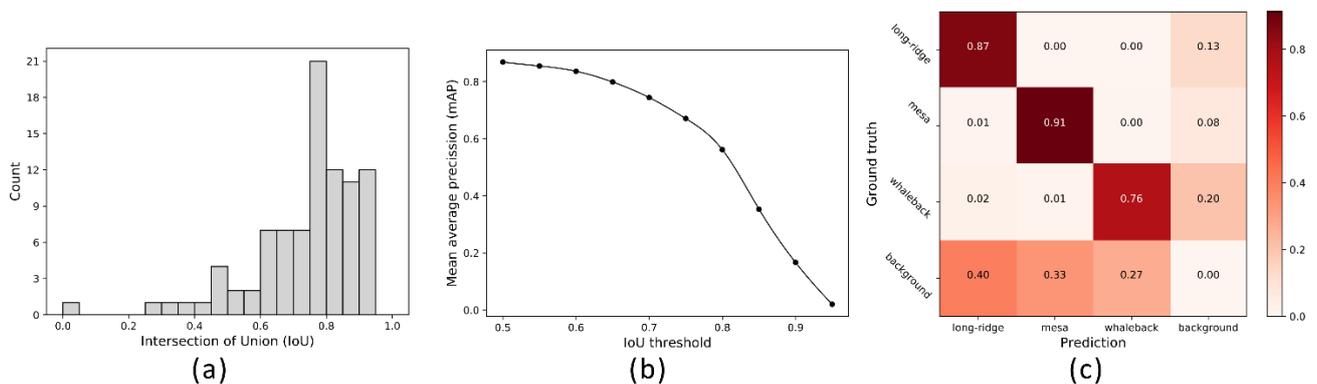


Figure 4. Metrics used to evaluate the accuracy of the Mask R-CNN model to detect yardangs on test dataset. (a) Histogram of IoU values between predicted masks and ground truths. (b) mAP over IoU thresholds of 0.50–0.95. (c) Normalized confusion matrix of classification results of three types of yardangs and the background.

3.2. Case Studies and Validation

To assess the transferability of this method, we conducted the case studies using the 80 Google Earth image subsets described in 2.5 with a spatial resolution of the original images of 0.6 m. We resampled these image subsets to three different spatial resolution images of 1.2 m, 2.0 m and 3.0 m using the nearest neighbor method. We applied the same trained model on these four groups of images with spatial resolutions from 0.6 to 3.0 m and analyzed their performances.

3.2.1. Case Study Results of 0.6 m Resolution Images

A total of 3361 yardangs are detected in this experiment, while 447 of them are false positives and 556 yardangs are missed. Table 2 shows the accuracies of detection, classification and delineation where the overall accuracy (OA) of detection is 74% with a precision and recall of 0.87 and 0.84, respectively. For successfully detected yardangs, the Mask R-CNN shows an ability to distinguish the fine classes of yardangs with a classification OA of 91%. The model can infer the outline of yardangs with a delineation OA of 95%, which proves that this method can be used for a precise characterization of yardangs. Figure 5 presents enlarged examples of classification and delineation results, with their locations marked in Figure 2. The long-ridge yardangs are arranged densely with very narrow spacing and show a wide range of lengths (from meters to hundreds of meters). Under these circumstances, there may be some missing instances or overlapping boundaries (Figure 5b,d). Although the shapes of mesa yardangs are mostly irregular, they can be well depicted because of their complete and clear boundaries (Figure 5f,h). For most of the whaleback yardangs, although their sizes vary widely, the delineation results are satisfactory. In some cases, the boundaries between the whaleback yardangs and their background can be very blurred and can thus cause some oddly shaped polygons (Figure 5j,l).

Table 2. Accuracy assessment for original images (0.6 m).

		Count	Precision	Recall	Overall Accuracy (%)
Detection	TP	2914	0.87	0.84	74
	FP	447			
	FN	556			
Classification	T	2648	-	-	91
	F	266			
Delineation	T	2768	-	-	95
	F	146			

3.2.2. Case Study Results of 1.2 m Resolution Images

A total of 1939 yardangs are detected in this case study. Based on visual inspection, 189 of them are false positives. The number of false negatives is 1719, which means approximately half of the yardangs are not detected. The down-sampled images lose some key features to detect small targets and, as expected, the recall of detecting decreases from 0.84 to 0.5. This results in a significant reduction of detection OA from 74% to 48%. Table 3 provides the detection, classification and delineation accuracies. Compared to the original images, the OA of classification and delineation drop by 6% and 2%, respectively. Some long-ridge yardangs shorter than 50 m are missed out and some overlapping polygons are created because of the ambiguous boundaries between yardangs (Figure 6b,d). The delineation results of mesa yardangs are still satisfying but several small residual yardangs are missed (Figure 6f,h). However, the detection of whaleback yardangs is drastically affected by the reduced spatial resolution. In Figure 6l, the number of the detected whaleback yardangs decreases by nearly 80% compared to the result of 0.6 m spatial resolution images (Figure 5l). It is noted that most of the missed yardangs were shorter than 40 m.

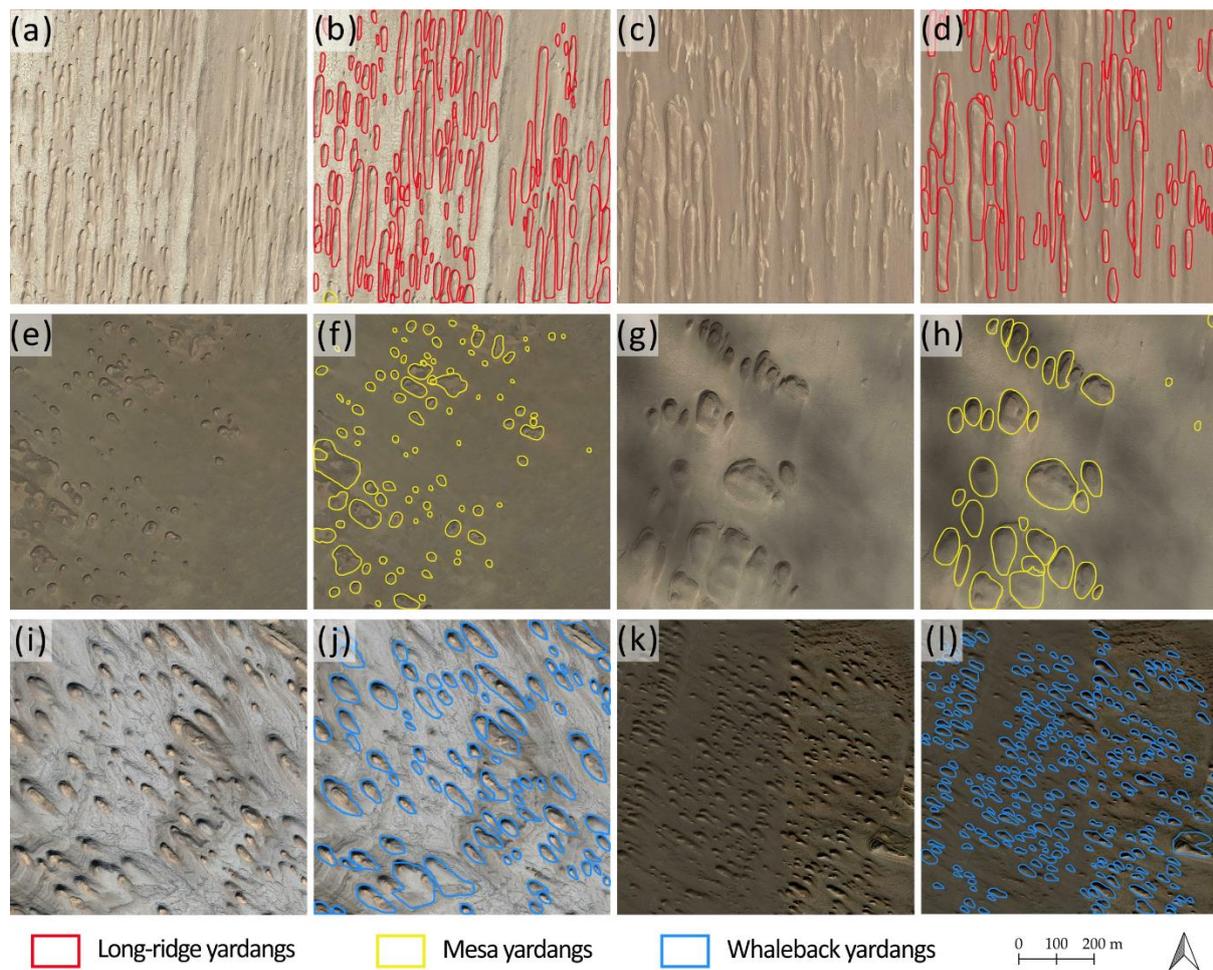


Figure 5. Examples of delineation and classification results on 0.6 m images. (a,c,e,g,i,k): enlarged 0.6 m spatial resolution images with their locations marked in Figure 2. (b,d,f,h,j,l): corresponding characterization results by Mask R-CNN where long-ridge, mesa and whaleback yardangs are depicted using polygons with red, yellow and blue borders, respectively.

Table 3. Accuracy assessment for original images (1.2 m).

		Count	Precision	Recall	Overall Accuracy (%)
Detection	TP	1750	0.90	0.50	48
	FP	189			
	FN	1719			
Classification	T	1492	-	-	85
	F	258			
Delineation	T	1627	-	-	93
	F	123			

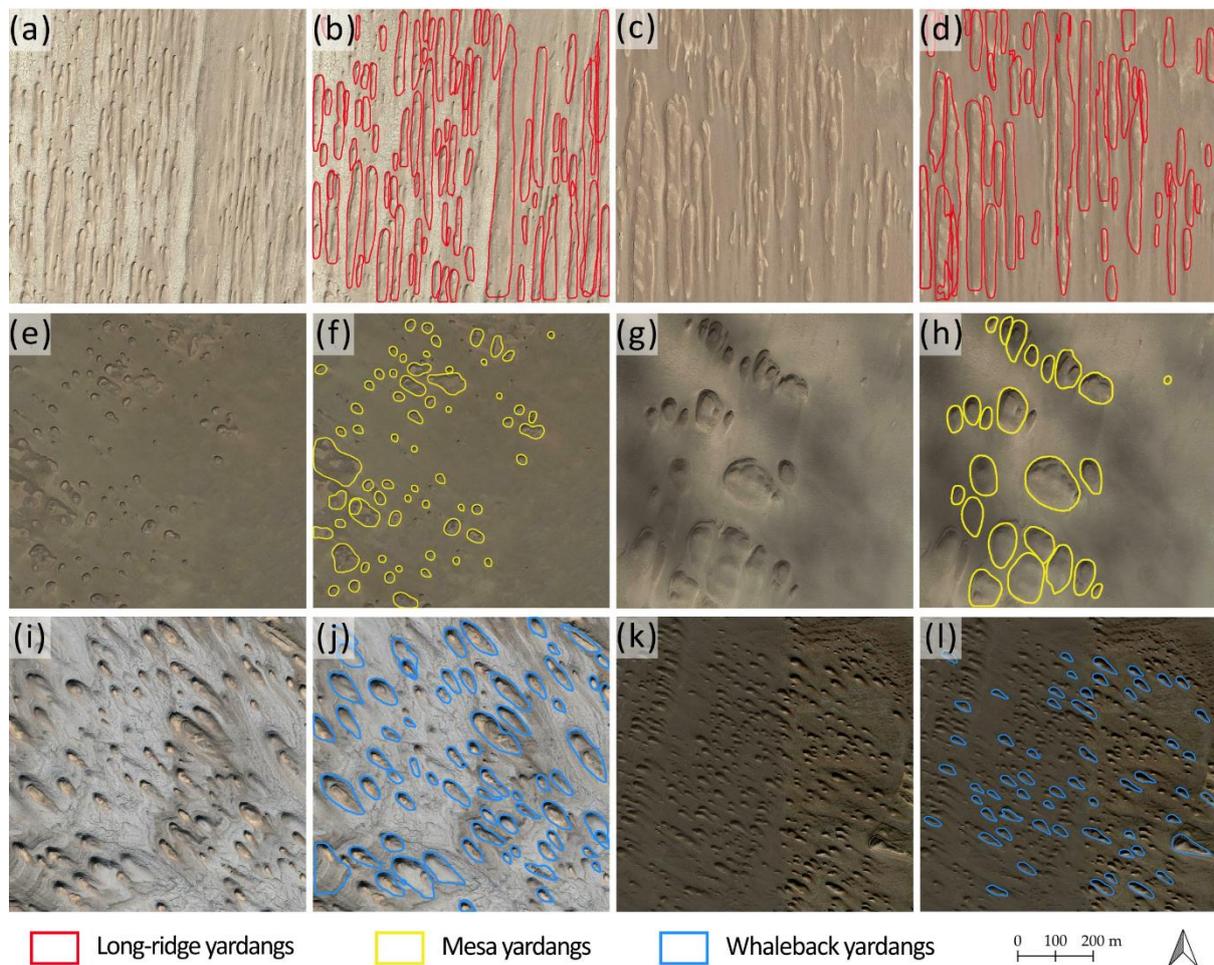


Figure 6. Example delineation and classification results on 1.2 m images. (a,c,e,g,i,k): enlarged 1.2 m spatial resolution images with their locations marked in Figure 2. (b,d,f,h,j,l): corresponding characterization results by Mask R-CNN where long-ridge, mesa and whaleback yardangs are depicted using polygons with red, yellow and blue borders, respectively.

3.2.3. Case Study Results of 2.0 m Resolution Images

A total of 1415 yardangs are detected on this set of images with 2.0 m spatial resolution. The number of false positives and false negatives are 109 and 2162, respectively (Table 4). Due to the increase of the missed detections, the OA of detection drops to 37%. Similarly, the OA of classification slightly decreases 1% compared to the result of 1.2 m resolution images. It is found that 94% of all detected yardangs can be correctly delineated regardless of the resampling issues, which indicates the image segmentation stability of this model. In this case study, many small end-to-end long-ridge yardangs are detected as one long yardang as a result of blurred boundaries (Figure 7b,d). The mesa yardangs can still be well delineated but more relatively small yardangs are missed (Figure 7f,h). The number of the detected whaleback yardangs decreases as the image pixel size increases and yardangs that are longer than 50 m are more likely to be detected (Figure 7j,l).

Table 4. Accuracy assessment for original images (2.0 m).

		Count	Precision	Recall	Overall Accuracy (%)
Detection	TP	1306	0.92	0.38	37
	FP	109			
	FN	2162			
Classification	T	1100	-	-	84
	F	206			
Delineation	T	1229	-	-	94
	F	79			

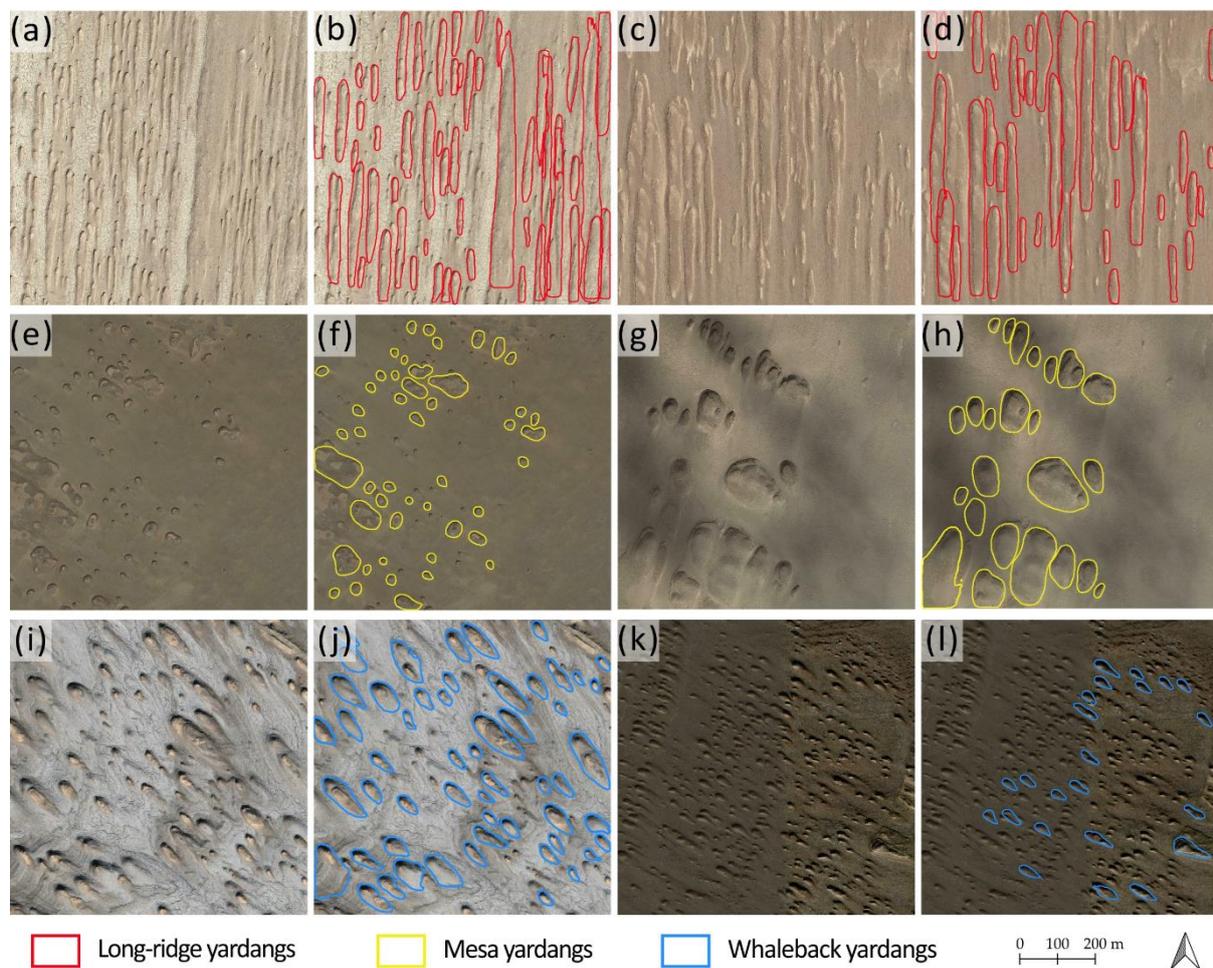


Figure 7. Example delineation and classification results on 2.0 m images. (a,c,e,g,i,k): enlarged 2.0 m spatial resolution images with their locations marked in Figure 2. (b,d,f,h,j,l): corresponding characterization results by Mask R-CNN where long-ridge, mesa and whaleback yardangs are depicted using polygons with red, yellow and blue borders, respectively.

3.2.4. Case Study Results of 3.0 m Resolution Images

The numbers of true positives and false negatives equate to 1042 and 2426, respectively (Table 5), which means that only 30% of the ground truth features are predicted on the 3.0 resolution images. The OA of detection, classification and delineation are 29%, 83% and 93%, respectively. Even though the outlines of yardangs are still recognizable for human eyes, the losses of boundary information are significant and make it hard for Mask R-CNN to detect and depict them. Therefore, the reduction of detected yardangs happen to all types, especially for long-ridge yardangs (Figure 8b,d) and whaleback yardangs

(Figure 8j,l), which are densely aligned and accompanied by many small yardang objects. Similar to the other case studies, the detection and delineation results of mesa yardangs are stable while benefitting from their relatively simple and clear morphological characteristics (Figure 8f,h).

Table 5. Accuracy assessment for original images (3.0 m).

		Count	Precision	Recall	Overall Accuracy (%)
Detection	TP	1042	0.94	0.30	29
	FP	67			
	FN	2426			
Classification	T	869	-	-	83
	F	173			
Delineation	T	974	-	-	93
	F	68			

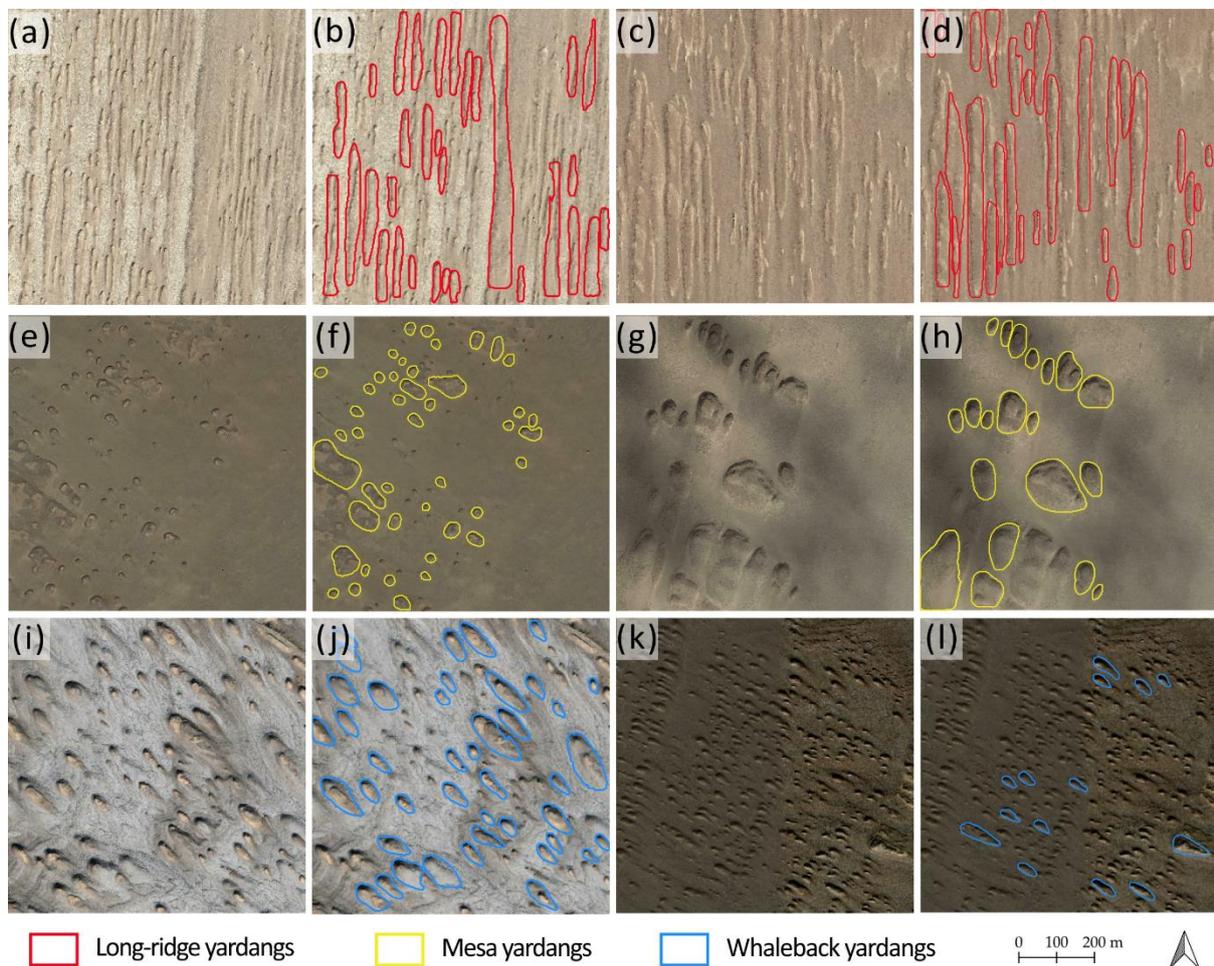


Figure 8. Example delineation and classification results on 3.0 m images. (a,c,e,g,i,k): enlarged 3.0 m spatial resolution images with their locations marked in Figure 2. (b,d,f,h,j,l): corresponding characterization results by Mask R-CNN where long-ridge, mesa and whaleback yardangs are depicted using polygons with red, yellow and blue borders, respectively.

3.3. Transferability

The four case studies evaluate the transferability of Mask R-CNN with respect to the image content and spatial resolution. Although the study sites are geographically close to the locations of training data, considering the complex and various development

environments of yardangs, they still exhibit variance in micro-landform at different degrees. The trained Mask R-CNN still achieved 74% OA of detection on 0.6 m spatial resolution images of new study sites. The OA of Mask R-CNN decreases as the pixel size of the image increases. However, the OA of detection, classification and delineation show different reactions to the increasing pixel size. The OA of detection drops by 45% (Figure 9a) when downsampling the image from 0.6 m to 3.0 m, which is a great degradation level compared to the 8% reduction of classification OA (Figure 9b) and 2% reduction of delineation OA (Figure 9c). This is mainly because the coarser resolution images lose some key information for effectively identifying yardangs and the background. Nonetheless for those yardangs who have been successfully detected, their boundaries are clear enough to be depicted and the variance of geomorphic characteristics of three types of yardangs are still significant.

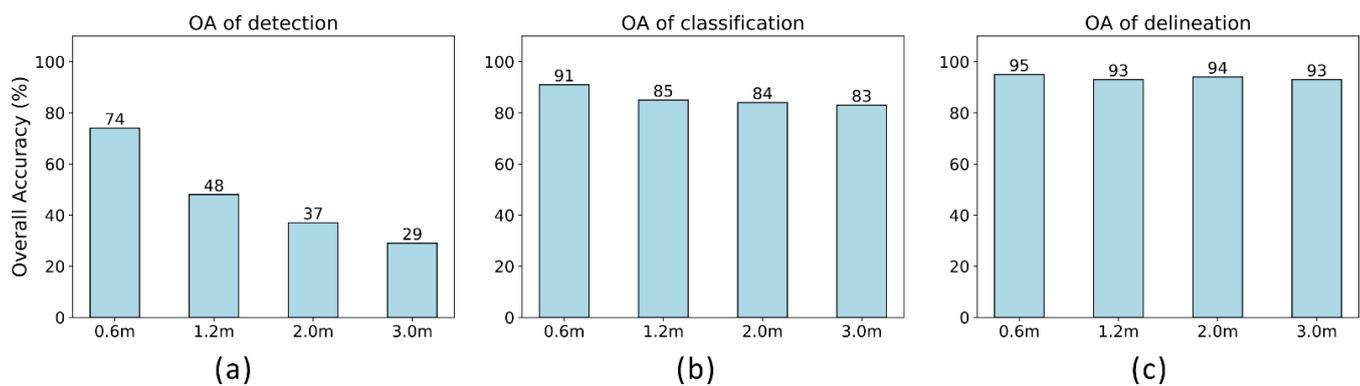


Figure 9. Statistical summary of overall accuracies of (a) detection, (b) classification and (c) delineation tasks on original and resampled images.

4. Discussion

4.1. Advantages and Limitations of Using Google Earth Imagery

Google Earth is a large collection of remotely sensed imagery including satellite and aerial images. The full coverage of the yardang fields in the study area and the availability of VHSR images make it possible to characterize yardangs with fine differentiations. The Google Earth images are acquired with different sensors and for different points in time, causing the variance in color and texture of yardangs in different images. This introduces more diversity to the training data and helps the model to learn more features of yardangs. However, the distribution of VHSR images in the study area is uneven. For example, the VHSR image subsets we used for annotation can be the mosaic results of images which have been resampled to the same resolution. Some of them are up-sampled from coarser images; however, the content and information of the images are not or very limitedly promoted as the pixel size of the image decreases. In addition, some yardangs are partly integrated with the background or are covered by gravel and sand. In such cases, it is hard for human interpreters to make an integral and accurate annotation of yardangs on images. As a result, the model may not sufficiently extract the detailed features of some yardangs during the training stage.

4.2. Advantages and Limitations of the Method

The results of this study that focus on characterizing aeolian yardangs demonstrate that the Mask R-CNN can be used to extract multi-scale and multi-type terrain features synthetically. The model learns the multi-level feature representations from the training data and could be successfully applied to the test dataset and other regions. Thanks to transfer learning, which is a notable advantage of deep learning, this method can potentially be applied to the other yardang fields on the Earth or even on the other planets. To the best of our knowledge, this is the first study that used DL method to automatically detect,

classify and delineate yardangs on remotely sensed imagery. However, there are still some limitations.

Methodologically, the performance of the Mask R-CNN largely depends on the quantity and quality of the training data, which require massive time and labor inputs for data selecting and labeling. In addition, as a supervised learning method, a large number of hyperparameters can influence the performance of the model. Considering the existing equipment condition in this study, it seems impossible to fully optimize the Mask R-CNN model. We adjusted some of the settings to make this heavy-weighted model easier to train on a smaller GPU. Despite this, the training process still took around 10 hours for 40 epochs. This was a much longer time for completing one experiment than needed for traditional machine learning methods and caused extremely high costs to be incurred for comparative studies. Therefore, the optimization relied on a limited series of experiments by tuning a small number of model parameters. Based on the case studies, we analyzed the effect of spatial resolution on mapping yardangs and found a negative correlation between image downsampling and OA of detection (Figure 9). Due to the relatively high spatial resolution (0.60 m and 1.19 m) of the training data, the prediction performance was very poor on coarser images. This indicates that the model requires a certain consistency between training data and predicting data.

Another crucial factor that affects the model performance lies in the yardangs' intrinsic feature complexity. To improve the quality of the training data, we select representative yardangs with very distinctive features. We also purposely balance the number of annotated instances for three types of yardangs to avoid potential biases, namely that the model is not more sensitive to a certain class. Although we collect training data from different sites and used data augmentation, the training dataset is not fully adequate considering the wide range of possible variations of yardang shapes. When applied to regions that host atypical yardangs, the prediction results are usually unsatisfactory. For example, some long-ridge yardangs show an ambiguous morphology. Their upwind sides are eroded into small ridges with tapered heads while their downwind sides gradually get wider and converge together without clear boundaries [7]. Under such circumstances, the Mask R-CNN may only delineate parts of the yardang or may generate a bizarre polygon that does not appropriately represent its direction and geometry. This also happens to whaleback yardangs when they occur in dense arrangements that are hard to be identified separately. The surfaces of yardangs are not always flat and smooth. In some places, remnant cap layers grow on the top surfaces of some long-ridge and mesa yardangs. Sometimes, these portions of a whole yardang are recognized as small yardangs by Mask R-CNN and thus generate several intersected or contained polygons, which does not reflect the real situation.

Despite being incredibly slow, the development of yardangs is a dynamic process and it is common for terrain features to have gradational boundaries in the transitional zone, which raises new issues for precise classification of yardangs. For instance, the downwind part of some long-ridge yardangs are cut off by fluvial erosion and gradually turn into mesa yardangs [20]. Once separated blocks are observed, they are recognized as mesa yardangs by interpreters. However, some of the yardangs remain the features of long-ridge yardangs and mislead the Mask R-CNN to predict them as long-ridges with relatively shorter lengths. It can be explained that the R-CNN (Region-based Convolutional Neural Network) model only uses information within a specific proposal region of an image but ignores the contextual information which is also crucial for object detection tasks [40]. Specifically, yardangs are spatially clustered and the visual identification for an ambiguously featured yardang would take the landscape pattern and the types of surrounding yardangs as references. These are the global and local surrounding contexts that are not considered in the Mask R-CNN. As a result, the model generated slightly confused classification results of yardangs in such situations.

4.3. Recommendations and Future Work

A limited amount of training data has always been a large obstacle for applying deep learning in the remote sensing domain especially on the extraction of natural terrain features [36,37]. Therefore, a large image dataset covering aeolian landforms with labeled data from multiple spatial resolutions and different sources is needed. The model trained on the thematic data may improve the performance of transfer learning compared to the utilization of pre-trained weights on normal photographs. The Mask R-CNN shows its advantage in extracting objects with unique color and textural features [42,45,57,64] on optical images. However, the elevation information of yardangs, which is an important parameter to distinguish between different types of yardangs and to differentiate them from other classes, is not involved in this study. Previous studies verified a high potential of CNN-based deep learning models for learning features from digital elevation data and other terrain derivatives [37,47,65]. Therefore, more attention should be given to additional information (e.g., DEMs) or the combined representations that integrate multi-source data as inputs for deep learning models. This can foreseeably promote the detection and delineation results of yardangs if their spectral and geometric characteristics are similar to the other aeolian landform. Further ways to improve terrain feature mapping could be the implementation of unsupervised [66], semi-supervised [67] and weakly-supervised [68] deep learning models for remote sensing image classification and segmentation. These approaches are an efficient way to extract features from unlabeled data and can break through the limitations caused by the scarcity of annotation data.

To verify the effectiveness of our method to characterize yardangs, future work is necessary to conduct comparative experiments using different algorithms, including object-based image analysis (OBIA) [69] and other conventional machine learning methods, in future studies. We also have to compare different CNN-based semantic segmentation methods like UNet and DeepLabv3+ while considering the recent progress within the computer vision domain. As a black-box model, the features that are learned by deep neural networks are unclear and the prediction results are usually of poor interpretability. Additional refinement of the current models may solve this problem, such as by introducing domain knowledge into the training process.

5. Conclusions

In this project, we apply the Mask R-CNN, a deep learning model for instance segmentation. The experimental results indicate that the Mask R-CNN can successfully extract the features of yardangs from the training dataset. In particular, the method can efficiently localize yardangs, depict their boundaries and classify them into correct fine categories simultaneously during the prediction stage. Our method yields a mean IoU value of 0.74 and when the IoU threshold is 0.5 for the testing dataset, there is a mean average precision of 0.869. Based on visual inspection, the automatically delineated boundaries of yardangs are very similar to manually annotated boundaries as they are accurate enough to support further morphological analysis. When applied to additional images with a wider geographical extent and complex background, the Mask R-CNN shows good transferability and can detect up to 74% of yardangs on 0.6 m spatial resolution images. The model trained on VHRS images has the potential to be applied to coarser images. However, the accuracy of detecting yardangs is more sensitive to the changes in image resolution compared to the classification and delineation tasks.

The extraction ability of the model for three types of yardangs differs slightly. In tendency, it yields better results for mesa yardangs, which have less complicated forms compared to long-ridge and whaleback yardangs. Since we focus on the most representative yardangs in this study, there is a need for further studies on the characterization of yardang features with variable complexities. Future work should also include a systematic analysis of the model optimization and comparative studies between deep learning and other methods. We hope that a largely automated and accurate yardang characterization based on the deep learning method will support future scientific studies on yardangs. The resulting

data should support geomorphometric and spatial distribution studies on yardangs and other aeolian landforms.

Author Contributions: Conceptualization, B.G. and N.C.; methodology, B.G.; software, B.G.; validation, B.G. and Y.X.; formal analysis, B.G. and C.Q.W.; investigation, B.G.; resources, N.C.; data curation, B.G.; writing—original draft preparation, B.G.; writing—review and editing, B.G., T.B., X.Y. and N.C.; visualization, B.G.; supervision, N.C. and T.B.; project administration, N.C.; funding acquisition, J.C., N.C., X.Y. and Z.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China under Grant No. 2016YFC1400903, the National Natural Science Foundation of China under Grant No. 41372205 and U1609202, the Fundamental Research Funds for the Central Universities under Grant No. 2019QNA3013, and the State Scientific Survey Project of China under Grant No. 2017FY101001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original very high spatial resolution remote sensing imagery was acquired from Google Earth. The annotated datasets and the source codes used in this study are available in these in-text citations: Gao et al. (2020) and are stored in Zenodo (<https://doi.org/10.5281/zenodo.4108325> (accessed on 31 December 2020)), which is an open access data repository.

Acknowledgments: We would like to thank Omid Ghorbanzadeh for constructive suggestions. We also thank Elizaveta Kaminov for assistance with the data annotation and English language proofreading. Thanks also to the anonymous reviewers for comments that significantly improved this manuscript.

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

References

1. McCauley, J.F.; Grolier, M.J.; Breed, C.S. *Yardangs of Peru and other Desert Regions*; US Geological Survey: Reston, VA, USA, 1977.
2. Lancaster, N. *Geomorphology of Desert Dunes*; Routledge: London, UK, 1995.
3. Goudie, A.S. Wind erosional landforms: Yardangs and pans. In *Aeolian Environments, Sediments and Landforms*; Goudie, A.S., Livingstone, I., Stokes, S., Eds.; John Wiley and Sons: Chichester, UK, 1999; pp. 167–180.
4. Alavi Panah, S.; Komaki, C.B.; Goorabi, A.; Matinfar, H. Characterizing land cover types and surface condition of yardang region in Lut desert (Iran) based upon Landsat satellite images. *World Appl. Sci. J.* **2007**, *2*, 212–228.
5. Ward, A.W. Yardangs on Mars: Evidence of recent wind erosion. *J. Geophys. Res. Solid Earth* **1979**, *84*, 8147–8166. [[CrossRef](#)]
6. Mandt, K.; Silva, S.d.; Zimbelman, J.; Wyrick, D. Distinct erosional progressions in the Medusae Fossae Formation, Mars, indicate contrasting environmental conditions. *Icarus* **2009**, *204*, 471–477. [[CrossRef](#)]
7. Wang, J.; Xiao, L.; Reiss, D.; Hiesinger, H.; Huang, J.; Xu, Y.; Zhao, J.; Xiao, Z.; Komatsu, G. Geological Features and Evolution of Yardangs in the Qaidam Basin, Tibetan Plateau (NW China): A Terrestrial Analogue for Mars. *J. Geophys. Res. Planets* **2018**, *123*, 2336–2364. [[CrossRef](#)]
8. Trego, K.D. Yardang Identification in Magellan Imagery of Venus. *Earth Moon Planets* **1992**, *58*, 289–290. [[CrossRef](#)]
9. Greeley, R.; Bender, K.; Thomas, P.E.; Schubert, G.; Limonadi, D.; Weitz, C.M. Wind-Related Features and Processes on Venus: Summary of Magellan Results. *Icarus* **1995**, *115*, 399–420. [[CrossRef](#)]
10. Paillou, P.; Radebaugh, J. Looking for Mega-Yardangs on Titan: A Comparative Planetology Approach. In Proceedings of the European Planetary Science Congress 2013, London, UK, 8–13 September 2013.
11. Laity, J.E. Landforms, Landscapes, and Processes of Aeolian Erosion. In *Geomorphology of Desert Environments*; Parsons, A.J., Abrahams, A.D., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 597–627. [[CrossRef](#)]
12. Al-Dousari, A.M.; Al-Elaj, M.; Al-Enezi, E.; Al-Shareeda, A. Origin and characteristics of yardangs in the Um Al-Rimam depressions (N Kuwait). *Geomorphology* **2009**, *104*, 93–104. [[CrossRef](#)]
13. Laity, J.E. Wind Erosion in Drylands. In *Arid Zone Geomorphology*; Thomas, D.S.G., Ed.; John Wiley & Sons: Chichester, UK, 2011; pp. 539–568. [[CrossRef](#)]
14. Pelletier, J.D.; Kapp, P.A.; Abell, J.; Field, J.P.; Williams, Z.C.; Dorsey, R.J. Controls on Yardang Development and Morphology: 1. Field Observations and Measurements at Ocotillo Wells, California. *J. Geophys. Res. Earth Surf.* **2018**, *123*, 694–722. [[CrossRef](#)]
15. Hedin, S.A. *Scientific Results of a Journey in Central Asia, 1899–1902*; Lithographic Institute of the General Staff of the Swedish Army: Stockholm, Sweden, 1907.
16. Blackwelder, E. Yardangs. *Gsa Bull.* **1934**, *45*, 159–166. [[CrossRef](#)]

17. Breed, C.S.; McCauley, J.F.; Whitney, M.I. Wind erosion forms. *Arid Zone Geomorphol.* **1989**, *284–307*.
18. Goudie, A.S. Mega-Yardangs: A Global Analysis. *Geogr. Compass* **2007**, *1*, 65–81. [[CrossRef](#)]
19. Kapp, P.; Pelletier, J.D.; Rohrmann, A.; Heermance, R.; Russell, J.; Ding, L. Wind erosion in the Qaidam basin, central Asia: Implications for tectonics, paleoclimate, and the source of the Loess Plateau. *Gsa Today* **2011**, *21*, 4–10. [[CrossRef](#)]
20. Li, J.; Dong, Z.; Qian, G.; Zhang, Z.; Luo, W.; Lu, J.; Wang, M. Yardangs in the Qaidam Basin, northwestern China: Distribution and morphology. *Aeolian Res.* **2016**, *20*, 89–99. [[CrossRef](#)]
21. Hu, C.; Chen, N.; Kapp, P.; Chen, J.; Xiao, A.; Zhao, Y. Yardang geometries in the Qaidam Basin and their controlling factors. *Geomorphology* **2017**, *299*, 142–151. [[CrossRef](#)]
22. De Silva, S.L.; Bailey, J.E.; Mandt, K.E.; Viramonte, J.M. Yardangs in terrestrial ignimbrites: Synergistic remote and field observations on Earth with applications to Mars. *Planet. Space Sci.* **2010**, *58*, 459–471. [[CrossRef](#)]
23. Xiao, X.; Wang, J.; Huang, J.; Ye, B. A new approach to study terrestrial yardang geomorphology based on high-resolution data acquired by unmanned aerial vehicles (UAVs): A showcase of whaleback yardangs in Qaidam Basin, NW China. *Earth Planet Phys* **2018**, *2*, 398–405. [[CrossRef](#)]
24. Ehsani, A.H.; Quiel, F. Application of Self Organizing Map and SRTM data to characterize yardangs in the Lut desert, Iran. *Remote Sens. Environ.* **2008**, *112*, 3284–3294. [[CrossRef](#)]
25. Zhao, Y.; Chen, N.; Chen, J.; Hu, C. Automatic extraction of yardangs using Landsat 8 and UAV images: A case study in the Qaidam Basin, China. *Aeolian Res.* **2018**, *33*, 53–61. [[CrossRef](#)]
26. Yuan, W.; Zhang, W.; Lai, Z.; Zhang, J. Extraction of Yardang Characteristics Using Object-Based Image Analysis and Canny Edge Detection Methods. *Remote Sens. Basel* **2020**, *12*, 726. [[CrossRef](#)]
27. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
28. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
29. Ball, J.; Anderson, D.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [[CrossRef](#)]
30. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; 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](#)]
31. Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognit.* **2017**, *61*, 539–556. [[CrossRef](#)]
32. Li, W.; Hsu, C.-Y. Automated terrain feature identification from remote sensing imagery: A deep learning approach. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 637–660. [[CrossRef](#)]
33. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
34. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Fully convolutional neural networks for remote sensing image classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5071–5074.
35. Palafox, L.F.; Hamilton, C.W.; Scheidt, S.P.; Alvarez, A.M. Automated detection of geological landforms on Mars using Convolutional Neural Networks. *Comput. Geosci.* **2017**, *101*, 48–56. [[CrossRef](#)] [[PubMed](#)]
36. Huang, L.; Luo, J.; Lin, Z.; Niu, F.; Liu, L. Using deep learning to map retrogressive thaw slumps in the Beiluhe region (Tibetan Plateau) from CubeSat images. *Remote Sens. Environ.* **2020**, *237*, 111534. [[CrossRef](#)]
37. Li, S.; Xiong, L.; Tang, G.; Strobl, J. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology* **2020**, *354*, 107045. [[CrossRef](#)]
38. Romera-Paredes, B.; Torr, P.H.S. Recurrent Instance Segmentation. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 312–329.
39. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969.
40. Li, J.; Wei, Y.; Liang, X.; Dong, J.; Xu, T.; Feng, J.; Yan, S. Attentive Contexts for Object Detection. *IEEE Trans. Multimed.* **2017**, *19*, 944–954. [[CrossRef](#)]
41. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 7–12 December 2015; pp. 91–99.
42. Wen, Q.; Jiang, K.; Wang, W.; Liu, Q.; Guo, Q.; Li, L.; Wang, P. Automatic Building Extraction from Google Earth Images under Complex Backgrounds Based on Deep Instance Segmentation Network. *Sensors* **2019**, *19*, 333. [[CrossRef](#)]
43. Zhao, P.; Gao, H.; Zhang, Y.; Li, H.; Yang, R. An Aircraft Detection Method Based on Improved Mask R-CNN in Remotely Sensed Imagery. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1370–1373.
44. You, Y.; Cao, J.; Zhang, Y.; Liu, F.; Zhou, W. Nearshore Ship Detection on High-Resolution Remote Sensing Image via Scene-Mask R-CNN. *IEEE Access* **2019**, *7*, 128431–128444. [[CrossRef](#)]
45. Zhang, W.; Witharana, C.; Liljedahl, A.K.; Kanevskiy, M. Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery. *Remote Sens. Basel* **2018**, *10*, 1487. [[CrossRef](#)]

46. Chen, Z.; Scott, T.R.; Bearman, S.; Anand, H.; Scott, C.; Arrowsmith, J.R.; Das, J. Geomorphological Analysis Using Unpiloted Aircraft Systems, Structure from Motion, and Deep Learning. Available online: <https://arxiv.org/abs/1909.12874> (accessed on 20 December 2020).
47. Maxwell, A.E.; Pourmohammadi, P.; Poyner, J.D. Mapping the Topographic Features of Mining-Related Valley Fills Using Mask R-CNN Deep Learning and Digital Elevation Data. *Remote Sens. Basel* **2020**, *12*, 547. [[CrossRef](#)]
48. Rohrmann, A.; Heermance, R.; Kapp, P.; Cai, F. Wind as the primary driver of erosion in the Qaidam Basin, China. *Earth Planet. Sci. Lett.* **2013**, *374*, 1–10. [[CrossRef](#)]
49. Han, W.; Ma, Z.; Lai, Z.; Appel, E.; Fang, X.; Yu, L. Wind erosion on the north-eastern Tibetan Plateau: Constraints from OSL and U-Th dating of playa salt crust in the Qaidam Basin. *Earth Surf. Process. Landf.* **2014**, *39*, 779–789. [[CrossRef](#)]
50. Wada, K. labelme: Image Polygonal Annotation with Python. Available online: <https://github.com/wkentaro/labelme> (accessed on 28 April 2020).
51. Hensman, P.; Masko, D. The impact of imbalanced training data for convolutional neural networks. Degree Project in Computer Science; KTH Royal Institute of Technology: Stockholm, Sweden, 2015.
52. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
53. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
54. 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 20 May 2020).
55. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens. Basel* **2015**, *7*, 14680–14707. [[CrossRef](#)]
56. Penatti, O.A.; Nogueira, K.; Dos Santos, J.A. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 44–51.
57. Stewart, E.L.; Wiesnerhanks, T.; Kaczmar, N.; Dechant, C.; Wu, H.; Lipson, H.; Nelson, R.J.; Gore, M.A. Quantitative Phenotyping of Northern Leaf Blight in UAV Images Using Deep Learning. *Remote Sens. Basel* **2019**, *11*, 2209. [[CrossRef](#)]
58. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014, Proceedings of the 2014 European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; Volume 5, pp. 740–755.
59. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. Available online: <https://arxiv.org/abs/1712.04621> (accessed on 24 July 2020).
60. Yu, X.; Wu, X.; Luo, C.; Ren, P. Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. *Gisci Remote Sens* **2017**, *54*, 741–758. [[CrossRef](#)]
61. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sens. Basel* **2019**, *11*, 196. [[CrossRef](#)]
62. Jung, A.B.; Wada, K.; Crall, J.; Tanaka, S.; Graving, J.; Reinders, C.; Yadav, S.; Banerjee, J.; Vecsei, G.; Kraft, A.; et al. Imgaug. Available online: <https://github.com/aleju/imgaug> (accessed on 1 February 2020).
63. Chen, L.; Barron, J.T.; Papandreou, G.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4545–4554.
64. Su, H.; Wei, S.; Yan, M.; Wang, C.; Shi, J.; Zhang, X. Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1454–1457.
65. Trier, Ø.D.; Cowley, D.C.; Waldeland, A.U. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeol. Prospect.* **2019**, *26*, 165–175. [[CrossRef](#)]
66. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1349–1362. [[CrossRef](#)]
67. Hong, D.; Yokoya, N.; Xia, G.-S.; Chanussot, J.; Zhu, X.X. X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data. *Isprs. J. Photogramm.* **2020**, *167*, 12–23. [[CrossRef](#)] [[PubMed](#)]
68. Wang, S.; Chen, W.; Xie, S.M.; Azzari, G.; Lobell, D.B. Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery. *Remote Sens. Basel* **2020**, *12*, 207. [[CrossRef](#)]
69. Blaschke, T. Object based image analysis for remote sensing. *Isprs. J. Photogramm.* **2010**, *65*, 2–16. [[CrossRef](#)]