mathematics

MDPI

Article

Rock Segmentation in the Navigation Vision of the Planetary Rovers

Boyu Kuang [1,*], Mariusz Wisniewski [1], Zeeshan A. Rana [1] and Yifan Zhao [2]

[1] Centre for Computational Engineering Sciences (CES), School of Aerospace, Transport and Manufacturing (SATM), Cranfield University, Bedfordshire MK43 0AL, UK; m.wisniewski@cranfield.ac.uk (M.W.); zeeshan.rana@cranfield.ac.uk (Z.A.R.)

[2] Centre for Life-Cycle Engineering and Management, School of Aerospace, Transport and Manufacturing (SATM), Cranfield University, Bedfordshire MK43 0AL, UK; yifan.zhao@cranfield.ac.uk

* Correspondence: neil.kuang@cranfield.ac.uk; Tel.: +44-7988-477406

Abstract: Visual navigation is an essential part of planetary rover autonomy. Rock segmentation emerged as an important interdisciplinary topic among image processing, robotics, and mathematical modeling. Rock segmentation is a challenging topic for rover autonomy because of the high computational consumption, real-time requirement, and annotation difficulty. This research proposes a rock segmentation framework and a rock segmentation network (NI-U-Net++) to aid with the visual navigation of rovers. The framework consists of two stages: the pre-training process and the transfer-training process. The pre-training process applies the synthetic algorithm to generate the synthetic images; then, it uses the generated images to pre-train NI-U-Net++. The synthetic algorithm increases the size of the image dataset and provides pixel-level masks—both of which are challenges with machine learning tasks. The pre-training process accomplishes the state-of-the-art compared with the related studies, which achieved an accuracy, intersection over union (IoU), Dice score, and root mean squared error (RMSE) of 99.41%, 0.8991, 0.9459, and 0.0775, respectively. The transfer-training process fine-tunes the pre-trained NI-U-Net++ using the real-life images, which achieved an accuracy, IoU, Dice score, and RMSE of 99.58%, 0.7476, 0.8556, and 0.0557, respectively. Finally, the transfer-trained NI-U-Net++ is integrated into a planetary rover navigation vision and achieves a real-time performance of 32.57 frames per second (or the inference time is 0.0307 s per frame). The framework only manually annotates about 8% (183 images) of the 2250 images in the navigation vision, which is a labor-saving solution for rock segmentation tasks. The proposed rock segmentation framework and NI-U-Net++ improve the performance of the state-of-the-art models. The synthetic algorithm improves the process of creating valid data for the challenge of rock segmentation. All source codes, datasets, and trained models of this research are openly available in Cranfield Online Research Data (CORD).

Keywords: image segmentation; remote sensing; terrain identification; data synthesis; transfer learning

Citation: Kuang, B.; Wisniewski, M.; Rana, Z.A.; Zhao, Y. Rock Segmentation in the Navigation Vision of the Planetary Rovers. *Mathematics* **2021**, *9*, 3048. https://doi.org/10.3390/math9233048

Academic Editors: Andrey Gorshenin, Mikhail Posypkin and Vladimir Titarev

Received: 17 September 2021
Accepted: 24 November 2021
Published: 27 November 2021

1. Introduction

Planetary rovers integrate various sensors and computing units, making the study an interdisciplinary research topic of subjects such as mathematics, human–robot interaction, and computer vision [1–3]. The *Spirit* rover endured the Martian winter, survived 1000 Martian days (*sols*), and traveled more than 6876 m, while the *Opportunity* rover traveled more than 9406 m [4]. However, the space environment poses challenges to the planetary rover operation [5]. The *Spirit* and *Opportunity* rovers experienced communication and function failures during their explorations [6,7]. To prevent this, automating onboard systems is essential for future planetary rovers [3,8]. This research focuses on the semantic terrain segmentation from the monocular navigation vision of the planetary rovers [8], which can provide support for the high-level planetary rover functionalities.

Semantic segmentation is an important research topic in computer vision [9]. Semantic segmentation can be achieved using either traditional computer vision or deep learning [10]. Traditional computer vision solutions utilize probabilistic models to predict pixels [11,12]. Deep learning-based solutions can be further classified into two categories: one-stage pipelines and two-stage pipelines [10]. One-stage pipelines provide End-to-End (E2E) [13] pixel-level predictions for each pixel [14,15]. Popular architectures include DeepLab [16], SSD [17], and U-Net [14]. Two-stage pipelines detect the bounding box of the target and then conduct pixel-level segmentations. Popular two-stage pipelines include RCNN [18], SDS [19], and Mask-RCNN [20].

Semantic segmentation plays an essential role in autonomous driving. Dewan et al. and Badrinarayanan et al. conducted multi-classification for each pixel (road, car, bicycle, column-pole, tree, and sky) [21,22]. Teichmann et al. committed to the road segmentation [23]. He et al. and Wu et al. focused on various traffic participants (vehicles and people) [20,24]. However, autonomous driving operates in a structured environment, while rover navigation, the focus of this research, operates in an unstructured environment. A structured environment refers to a scene with prior knowledge, while an unstructured environment refers to a scene without prior knowledge [25].

Rocks are typical semantic targets in planetary environments [26,27]. The jet propulsion laboratory (JPL) in the *National Aeronautics and Space Administration* (NASA) studied the terrain classification for the planetary rovers [6,28]. Rocks play a significant role in the planetary rovers' autonomy [26]. For example, the *Curiosity* Mars rover involves a generally flat plain with about 5% of the area covered by small (tens of cm size or smaller) rocks [26]. The *Spirit*, *Curiosity*, and *Opportunity* all occurred challenges because of rock-related terrain [6,7,29]. However, existing geometric hazard detection methods cannot detect all of the rocks [28].

The related studies on rock segmentation for planetary rovers can be divided into the following five categories. Table 1 summarizes the discussions in a tabular form, while their results have been summarized in Table 1 in the Appendix A.

Table 1. The summary of the related studies on rock segmentation for planetary rovers.

Category [1]	Explanation	Machine Learning-Based	Reference Index [2]
i	3D point cloud	No	[30–32]
ii	Edge-based method	No (except [33])	[4,5,33–36]
iii	Outstanding rocks	No	[5,37,38]
iv	Other non-machine learning studies	No	[32,39–41]
v	Machine learning studies	Yes	[8,27,28,35,42–44]

[1] "i", "ii", "iii", "iv", and "v" correspond to the same index of category in the context. [2] "Reference index" refers to the same citation index in References.

Category-i refers to the studies that use 3D point clouds [30–32]. The 3D point cloud is generally obtained through LIDAR or stereo cameras, which requires considerable computing resources and storage space. This research applies a less computing and lighter weight solution through 2D images and the monocular camera.

Category-ii refers to the studies that use texture and boundary-based image processing methods [4,5,33–36]. The *Rockster* [36] and *Rockfinder* [34] are popular software packages in this category. However, some image conditions (such as skylines, textures, backgrounds, and unclosed contours) can significantly affect their performance [4]. This research has better robustness on image conditions by applying the various brightness, contrast, and resolution to the input images.

Category-iii refers to the studies focusing on rock identification [5,37,38], while the rock segmentation is only a sub-session of the identification studies. However, this research focuses on pixel-level segmentation, which can achieve more accurate segmentation results.

Category-iv refers to all the rest of the studies using non-machine learning-based methods. Virginia et al. committed to using shadows to find rocks [39]. Li et al. built detailed topographic information of the area between two sites based on rock peak and

surface points [40]. Xiao et al. focus on reducing computational cost [32]. Yang and Zhang proposed a gradient-region constrained level set method [41]. In general, they applied artificial features, which usually require significant manual adjustments. This research uses learning-based features, which can intelligently learn the optimized feature from the image and annotations.

Category-v refers to the studies using machine learning methods. Dunlop et al. used a superpixel-based supervised learning method [35]. Ono et al. used Random Forest for terrain classification [28]. Ding Zhou et al. and Feng Zhou et al. focused on the mechanical properties corresponding to different terrain types [27,42]. Gao et al. reviewed the related results of monocular terrain segmentation [8]. Furlan et al. conducted a deeplabv3plus-based rock segmentation solution [43], and Chiodini et al. proposed a fully convolutional network-based rock segmentation solution [44]. Although their performance is much better than Category-i/ii/iii/iv, their training dataset is very small because the annotation costs significant time and effort. This research proposes a synthetic algorithm that can generate a large amount of data and corresponding annotations with very limited manual annotation.

Pixel-level rock segmentation is a challenging task. The shape of rocks in an unstructured planetary exploration environment is hard to predict [5]. Identifying the boundary of the rocks can be made difficult by the low resolution of the navigation camera and the blurred outlines between background and rocks. Furthermore, most rock segmentation datasets for the planetary rovers are confidential to the public or only in the form of images instead of video [7,45].

A solution based on generating synthetic data addresses these problems. Data synthesis produces pixel-level data annotation and image generation. Therefore, synthetic data can generate a large amount of images and corresponding annotations for the pre-training process [46]. Furthermore, the synthetic process is based on the practical video stream, which guarantees good transferability in the following transfer-training process. Then, the model can be transfer-trained to the convergence based on the prior knowledge from the pre-training process.

The contributions of this research include the following:

(i) This research proposed a synthetic algorithm and transfer learning-based framework, which provides a labor-saving solution for the rock segmentation in the navigation vision of the planetary rovers.
(ii) This research proposed a synthetic algorithm and a synthetic dataset, which aid the research into the rock segmentation in the navigation vision of the planetary rovers.
(iii) This research came up with an end-to-end (E2E) network (NI-U-Net++) for the pixel-level rock segmentation, which achieved state-of-the-art in the synthetic dataset.

All source codes, datasets, and trained models of this research are openly available in Cranfield Online Research Data (CORD) at https://doi.org/10.17862/cranfield.rd.16958728, accessed on 26 November 2021.

The article is arranged as follows. Section 2 depicts the proposed synthetic algorithm and rock segmentation network. Section 3 discusses the experimental results. Conclusions and future work are placed in Section 4.

2. Methods

The proposed rock segmentation framework is based on the transfer learning process (see Figure 1). Transfer learning is a typical solution for the data-limited situation [47,48]. The overall framework can be divided into the following.

(1) The framework can be divided into two processes. Figure 1 identifies the pre-training process and the transfer-training process with the blue and green frames, respectively. Rock segmentation in an unannotated scenario is significantly difficult, and the transfer learning strategy divides the learning process into two steps. Although the synthetic dataset can generate large amount of pixel-level annotated data, they inevitably have a significant difference from the real-life data. The real-life data represent the practical mission, while its annotation corresponds to an expensive

cost. Therefore, a cooperated solution between the synthetic data and real-life images becomes very promising. The pre-training process aims to achieve prior knowledge from a similar scene, and then, the transfer-training process fine-tunes the pre-trained weight to fit the real-life images.

(2) In the pre-training process:
   (a) The purple ellipse with "Annotation-1" refers to the first manual annotation, which aims to acquire the backgrounds and rock samples for the synthetic algorithm.
   (b) Then, the synthetic algorithm utilizes these backgrounds and rock samples to generate the synthetic dataset. The synthetic dataset contains 14,000 synthetic images and corresponding annotations.
   (c) The orange solid round frame refers to the proposed rock segmentation network (NI-U-Net++). The blue dash arrow refers to the pre-training, which aims to achieve prior knowledge from the synthetic dataset.
   (d) The pre-training process eventually accomplishes the pre-trained weights of the NI-U-Net++, and these pre-trained weights refer to the prior knowledge from the synthetic dataset.

(3) In the transfer-training process:
   (a) The purple ellipse with "Annotation-2" refers to the second manual annotation, which aims to produce some pixel-level annotations (see the green round frame with "Annotated visual dataset"). The "Annotated visual dataset" contains 183 real-life images and corresponding pixel-level annotations.
   (b) The green dash arrow refers to the transfer training, which aims to fine-tune the pre-trained weights to fit the "Annotated visual dataset".
   (c) (iii–iii) The transfer-training process comes up with the final weights of the NI-U-Net++.

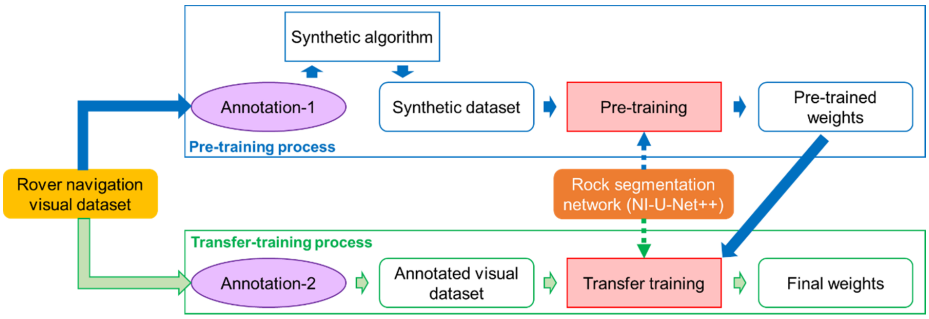

Figure 1. The pipeline of the proposed rock segmentation framework. The rover navigation visual dataset used in this research is the Katwijk beach planetary rover dataset [49], while it can be different in other scenarios. The synthetic dataset for the pre-training is not augmented, while the annotated visual dataset for the transfer training is applied augmentation to extend the dataset.

2.1. The Real-Life Visual Navigation Dataset for the Planetary Rovers

The visual navigation dataset of the planetary rovers used in this research is part1 and part2 of the *Katwijk* beach planetary rover dataset [49] from the European Space Agency (ESA) [50–53], which contains 2250 frames of the image. The *Katwijk* dataset is a professional open dataset for the navigation vision of the planetary rover research, and many studies use the *Katwijk* dataset as the planetary environment [44,52,54]. The *Katwijk* dataset is achieved at the site where is near the heavy-duty planetary rover (HDPR) platform project of the European Space and Technology Research Center [49].

The reasons for adopting the *Katwijk* dataset are as follows: (i) The focus of this research is to integrate a real-time and E2E rock segmentation framework into the navigation vision of planetary rovers. Thus, a navigation vision stream for evaluating the real-time performance is essential. (ii) The *Katwijk* dataset involves all relevant landmarks supported

in the research of Ono et al. [28]. (iii) Other datasets are not suitable for this research. For example, [54] involves some targets that are less likely to appear in planetary exploration (such as the tree, wall, and people). (iv) Other datasets (such as NASA raw images [55]) contain many different types of rock samples, introducing a more complex marginal probability distribution (this research utilizes the concept about task, domain, and marginal probability distribution from [56] as the fundaments). However, rock diversity (or even new rocks [38]) is not the focus of this research but an entirely new discipline.

2.2. The Synthetic Dataset

The proposed synthetic algorithm aims to generate a large amount of images and the corresponding pixel-level annotations with limited manual annotations. Although planetary exploration provides numerous visual data, they have barely been pixel-level annotated. Labor-saving annotation is a vital and usual challenge for planetary visual data. The target of the synthetic algorithm is to build a labor-saving solution to generate a large amount of images and corresponding pixel-level annotations for the pre-training process. Planetary explorations are expensive regarding labor, time, and resource, while the synthetic approach aims to minimize the associated costs. Although multi-labeler seems a promising solution for suppressing human errors, it will further increase the labor and time required. The proposed synthetic algorithm can generate pixel-level annotations while generating synthesized images. To maintain the labor-saving and annotation quality, the following four highlights are essential for designing the synthetic algorithm.

(1) The synthetic algorithm also prepares data for the pre-training process. Therefore, the materials utilized in the synthetic algorithm should come from the real-life images.
(2) Another target is to generate images and annotations synchronously through the synthesis algorithm, thereby significantly reducing the cost of manual intervention.
(3) The target is to ensure the diversity of the synthetic dataset. The pre-training dataset can determine the robustness and generalization ability of the segmentation framework for the navigation visions. The data diversity introduced through morphology, brightness, and contrast transformations are significantly important to the above end.
(4) The embedded rock samples require further processing to simulate the visual comfortable images.

2.2.1. The Proposed Synthetic Algorithm

The synthetic algorithm uses image processing technology and the illumination intensity-based assumption. Equations (1)–(9) and Figure 2 depict the illumination intensity-based assumptions and the corresponding process based on the geometrics and mathematics. Figure 2a abstracts a typical navigation scenario of the planetary rovers using a sketch. The light source (I) can be approximated as the sun in the scenario. The angles between the rays i_1, i_2, and i_3 of the light and the horizontal ground (g) are θ_1, θ_2, and θ_3, respectively. When I is significantly far away from the ground g, this research considers that all light rays are parallel to each other, so the angles θ_1, θ_2, and θ_3 of i_1, i_2, and i_3 and the horizontal ground are equal (see Equation (1)).

$$\begin{cases} i_1 \parallel i_2 \parallel i_3, \\ \theta_1 = \theta_2 = \theta_3, \end{cases} \quad if\ I \to +\infty. \tag{1}$$

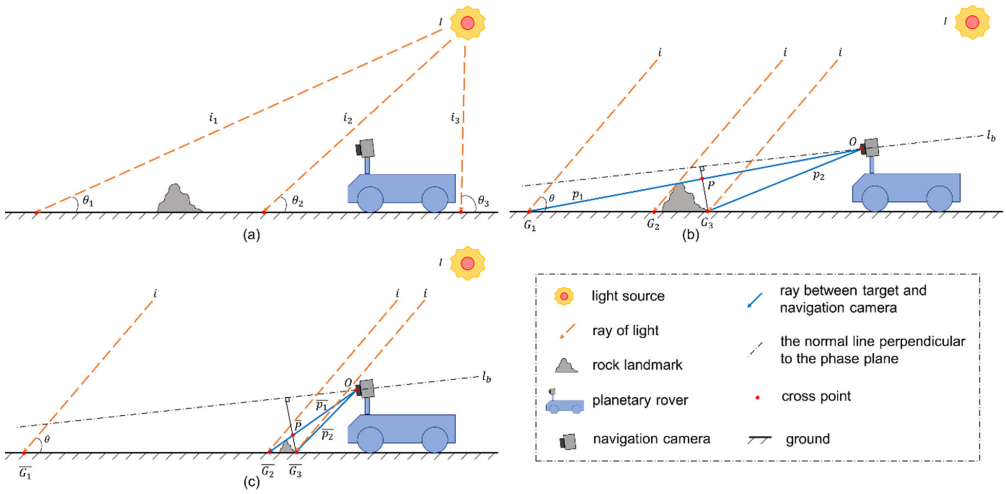

Figure 2. The sketches of the planetary rover navigations. (**a**) refers to the typical scenario of the planetary rover. (**b**) refers to the abstracted scenario through applying Equation (1) to (**a**). (**c**) refers to the abstracted scenario with a small and closer rock landmark compared to (**b**). G_1, G_2, G_3, P, p_1, and p_2 in the (**b**) scenario correspond to $\overline{G_1}$, $\overline{G_2}$, $\overline{G_3}$, $\overline{P}$, $\overline{p_1}$, and $\overline{p_2}$ in the (**c**) scenario, respectively.

Figure 2b shows the abstracted sketch of Figure 2a through applying Equation (1). The angles between all rays (i) and g all equal to θ. This research defines ρ to refer to the density of rays, which also refers to the illumination intensity (L) in the unit area on the ground. Therefore, the L on a specific ground area equals the multiplication between the area of the region (S) and ρ (see Equation (2)).

$$L = \rho * S \tag{2}$$

The solid blue lines (p_1 and p_2) in Figure 2b refer to the rock area captured by the navigation camera. The dashed line (l_b) refers to the normal line perpendicular to the phase plane. The solid black line segment (PG_3) refers to the corresponding rock on the image. Although the rock occupies the same image region as the ground G_1G_3, the L of the rock is different than the ground without rock because of the difference between G_2G_3 and G_1G_3. In Equation (3), $L_{G_1G_3}$ refers to the L in the G_1G_3 area, and PG_3 refers to the PG_3 area.

$$\rho = \frac{L_{G_1G_3}}{PG_3} \tag{3}$$

Notably, all images involved in this section refer to the grayscale images. Thus, PG_3 is a grayscale image. This research assumes the image grayscale value of the PG_3 area relates to two parameters, corresponding density (ρ) and the surface optical properties (P_{opt}) of the object (c_T).

i. The above discussion uses Equation (2) to achieve the desired illumination intensity, while ρ is difficult to obtain from a grayscale image. However, the known information is the corresponding image grayscale value (G_1G_3) and the area of PG_3. It is noteworthy that G_1G_3 and G_2G_3 appear in the same image region. This research assumes that the ratio ($\overline{\rho}$) between the sum grayscale in G_1G_3 and the area of PG_3 can approximate the value of ρ (see Equation (4)).

$$\rho \approx \overline{\rho} = \frac{L_{G_1G_3}}{PG_3} = \sum_{(x,y) \in T} \left[\frac{pixel_{img}(x, y)}{N_{pixel}} \right] \tag{4}$$

However, Figure 2c shows another scenario. A pronounced difference between G_1G_2 (Figure 2b) and $\overline{G_1G_2}$ (Figure 2c) comes from a smaller and closer rock landmark. Therefore, the difference ($\Delta\rho$) between ρ and $\overline{\rho}$ is located on G_1G_2 (equivalent to

$\overline{G_1G_2}$). It is noteworthy that ρ is the ratio between the sum grayscale of G_1G_3 and PG_3, whereas $\overline{\rho}$ is the ratio between the sum grayscale of G_2G_3 to PG_3 (see Equation (5)).

$$\Delta\rho = \overline{\rho} - \rho = \frac{L_{G_1G_3}}{PG_3} - \frac{L_{G_2G_3}}{PG_3} = \frac{L_{G_1G_3} - L_{G_2G_3}}{PG_3} \tag{5}$$

Substituting Equation (2) into Equation (5) can produce Equation (6), so $\Delta\rho$ is a value related to $L_{G_1G_2}$.

$$\Delta\rho = \overline{\rho} - \rho = \frac{\rho * G_1G_3 - \rho * G_2G_3}{PG_3} = \frac{\rho * G_1G_2}{PG_3} = \frac{L_{G_1G_2}}{PG_3} \tag{6}$$

ii. The optical properties of the object surface are complex (such as surface reflectance, refracting, and absorptivity), and they do not belong to the scope of this research. Here, we use a variable c_T to pack all factors related to optical properties. Equation (7) depicts the grayscale change caused through the optical properties.

$$P_{opt} = f_1(c_T) \tag{7}$$

Recalling the objective of the synthetic algorithm, Equation (7) can only correlate the optical properties and image grayscales implicitly. Thus, this research proposes Equation (8) to approach Equation (7) artificially. Equation (8) assumes that the grayscale distribution in the target region (rock in this research) is a function of the coordinates when ρ is constant. This research calculates the averaged grayscale value (img_{mean}) for the corresponding image area. Then, it subtracts the grayscale values (img) to img_{mean} to obtain a differential grayscale "image" (img_Δ), which is a statistical result only related to the coordinates.

$$P_{opt} \approx img_\Delta = img - img_{mean} \tag{8}$$

The synthetic algorithm corresponding to the rock-embedded area can be depicted using Equation (9):

$$\overline{L} = \overline{\rho} * img_\Delta - C. \tag{9}$$

The C refers to the constants used to correct the distance between ρ and $\overline{\rho}$. Recalling Equation (6), $\Delta\rho$ positively correlates to the $L_{G_1G_2}$. The practical area of $L_{G_1G_2}$ is a varying value that is dependent on the appearance of the target. Measuring $L_{G_1G_2}$ is challenging, but $L_{G_1G_2}$ positively correlates to img_{mean} (a brighter image causes a higher $L_{G_1G_2}$). Thus, this research assumes C is a constant that depends on img_{mean}. Table 2 depicts the values of C, while the detailed experiments for deciding C can be found in Appendix A.2 in the Appendix A. It is noteworthy that L, $\overline{L}$, and img_Δ all contain multiple values, which correspond to the coordinates.

Table 2. The constant C to correct $\overline{\rho}$ from ρ.

Conditions	Value [1]
$img_{mean} \leq 25$	0
$25 < img_{mean} \leq 50$	5
$50 < img_{mean} \leq 75$	10
$75 < img_{mean} \leq 100$	15
$100 < img_{mean} \leq 125$	20
$125 < img_{mean} \leq 150$	25
$150 < img_{mean} \leq 175$	30
$175 < img_{mean} \leq 225$	35
$200 < img_{mean} \leq 225$	40
$225 < img_{mean} \leq 250$	45
$250 < img_{mean}$	50

[1] The values correspond to the grayscale metric with 256 scales.

2.2.2. Implementation

Figure 3 and Algorithm 1 show the implementation process of the proposed synthetic algorithm.

i. This research randomly picks up 35 images from the *Katwijk* dataset for "Annotation-1". The number of 35 images is arbitrary; it needs to be large enough to get a sufficient dataset of rock annotations but not too large that it takes a long time to annotate the images. Furthermore, this research focuses on exploring a feasible framework so that the upper and lower limits of the image number in "Annotation-1" are not studied thoroughly.
ii. Then, the synthetic algorithm conducts the "Annotation-1" to these images (see Figure 3). The red mask refers to the rock sample, and the green masks refer to other rocks. It is noteworthy that each image only includes the largest rock in the rock samples. Before embedding into a new background, a morphological transformation is necessary to ensure the variant of the synthetic dataset. However, the enlarged morphological transformations can bring a significant resolution change if the rock sample is too small.
iii. The algorithm also utilizes the images in "Annotation-1" as backgrounds for the synthetic algorithm. The annotation rule for "Annotation-1" is: if a rock cannot be identified with the three to six times enlargement, this research decides to abandon it as a part of the background.
iv. The above three steps finish the data preparation for the synthetic algorithm. The *rocks* refer to rock samples, and *scenes* refer to backgrounds. Then, the algorithm conducts Algorithm 1 to generate the synthetic dataset.
v. Morphological transformations can increase the number and diversity of the synthetic dataset. The morphological transformation schemes for rock samples (aug_{rock}) come from the combinations using mirror, flatten, narrowing, and zooming. The morphological transformation schemes for backgrounds (aug_{scene}) further include the adjustments of brightness, contrast, and sharpness.
vi. Then, Algorithm 1 traverses each background with all aug_{scene} to achieve the morphologically transformed images ($scene_{aug}$ in Algorithm 1) (see row 3 and 4 in Algorithm 1). Meanwhile, the sky and ground segmentation model is applied to identify the ground pixels, and the rock samples are only embedded into the ground region. The sky and ground segmentation model comes from [57].
vii. For each $scene_{aug}$ the synthetic algorithm embeds a random number of $rock_{aug}$ (see row 11 in Algorithm 1).
viii. Each $rock_{aug}$ comes from a random selection from the rocks ($rock_{select}$). The algorithm also randomly selects a morphological transformation scheme (aug_r) from aug_{rock}. The algorithm conducts aug_r to $rock_{select}$, which results in a morphologically transformed rock ($rock_{aug}$) (see rows 9, 10, and 11 in Algorithm 1).
ix. The algorithm adopts Equation (8), Equation (4), Table 2, and Equation (9) to achieve img_Δ, $\overline{\rho}$, the correction constant for the corresponding $\overline{\rho}$ (C_{select}), and the grayscale values of the embedded rock ($rock_{replace}$) (see rows 12, 14, 15, and 16 in Algorithm 1). The further discussion of the values in Table 2 can be found in Appendix A.2.
x. Finally, the synthetic images that correspond to the $scene_{aug}$ are saved as the synthetic dataset.

Algorithm 1: Synthetic algorithm

Input: rock samples: $rock = [rock_1, rock_2, \ldots, rock_{35}]$
practical sense: $sense = [scene_1, scene_2, \ldots, scene_{35}]$
correction constant: $C = [C_1, C_2, \ldots, C_{11}]$
scene augmentation: $aug_{scene} = [s_1, s_2, \ldots, s_8]$
rock augmentation: $aug_{rock} = [r_1, r_2, \ldots, r_9]$
Output: Synthetic dataset: $img = [img_1, img_2, \ldots, img_n]$

for img_{idx} in $range(n)$ **do**
  for $scene$ in $scenes$ **do**
    for aug_s in aug_{scene} **do**
      $scene_{aug} = aug_s(scene)$
      for i in $range(50)$ **do**
        $N_{rocks} \leftarrow$ **a random integer between 5 and 20**;
        for j in $range(N_{rocks})$ **do**
          $(x_a, y_a) \leftarrow$ **random anchor point for rock**;
          $rock_{select} \leftarrow$ **random select in** $rocks$;
          $aug_r \leftarrow$ **random select in** aug_{rock};
          $rock_{aug} = aug_r(rock_{select})$
          $img_\Delta \leftarrow rock_{aug}$; Equation (8)
          $img_{mean} \leftarrow scene_{arg}$ & (x_a, y_a);
          $\bar{\rho} \leftarrow img_{mean}$; Equation (4)
          $C_{select} \leftarrow$ **find in** C; Table 2
          $rock_{replace} \leftarrow img_\Delta$ & $\bar{\rho}$ & C_{select}; Equation (9)
          $img_{idx} \leftarrow$ **embed** $rock_{replace}$ **in** $scene_{arg}$ **at** (x_a, y_a)
        end
      end
    end
  end
end

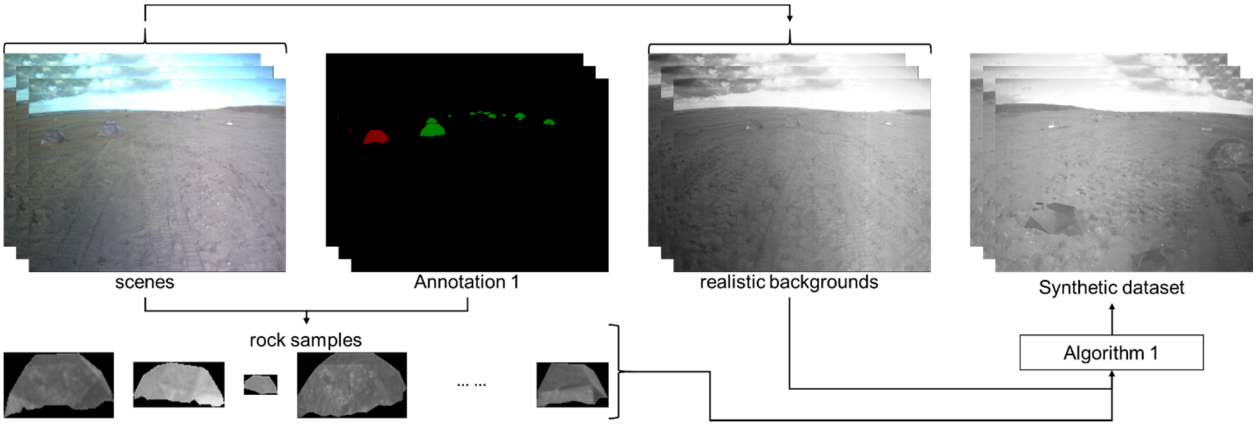

Figure 3. The preparation part in the implementation of the proposed synthetic algorithm. "Annotation 1" refers to the same "Annotation-1" as in Figure 1. The red and green pixels in "Annotation 1" refer to the rock samples and other rocks, respectively.

It is noteworthy that the proposed synthetic algorithm is a typical incremental method through embedding new rock samples into the original image, which inevitably adds many large and obvious rocks. Thus, the synthetic algorithm may lead the quantitative metrics to a better result in the pre-training process than the transfer-training process. In addition,

the metrics adopted in this research include the accuracy, intersection over union (IoU), and Dice score.

2.3. Proposed Rock Segmentation Network

This section discusses the modified rock segmentation network (named the NI-U-Net++). Figure 4 depicts the proposed NI-U-Net++, which is a modified U-Net++ [15] through modifying the overall architecture and integrating some modified micro-networks. It is noteworthy that this research has been inspired by the U-Net++ [15] and NIN [58].

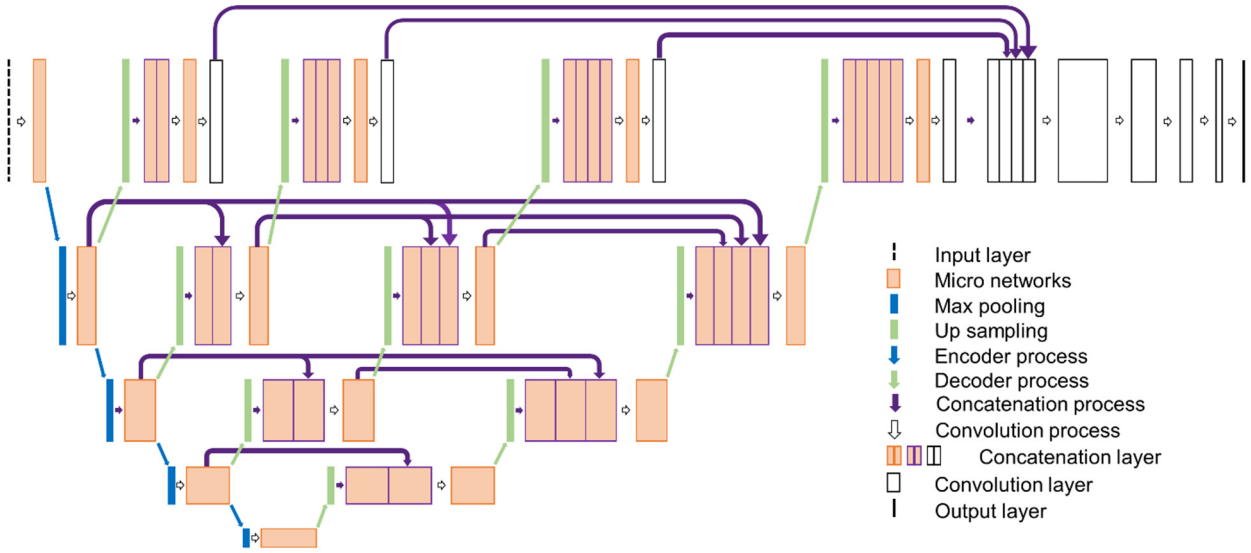

Figure 4. The proposed rock segmentation network (NI-U-Net++).

The U-Net network uses the encoder–decoder configuration and concatenation layer to configure a deep network [14,59], which provides an efficient and effective structure for feature extraction and backpropagation. U-Net++ is an updated U-Net, which adopts a new encoder–decoder network with a series of nested, dense skip pathways. U-Net++ further applies deep supervision to avoid the skips of the shallow sub-U-Nets [15].

The proposed NI-U-Net++ adopts a similar overall structure of the encoder–decoder design and deep supervision as the U-Net++. The blue and green solid arrows in Figure 4 refer to the encoder and decoder part, respectively. The encoder part in the NI-U-Net++ has four scale reductions (see the four blue arrows in Figure 4). Deep supervision is implemented using the concatenate and convolutional layers (see the purple arrows and white blocks at the top of Figure 4). Moreover, the purple arrows refer to the "highway" using the concatenate layers to connect the front and back layers. These highways can pass the backpropagation gradients in the front layers, thereby avoiding the gradient vanishing.

The NIN refers to the micro-block of networks assembled in another neural network. The 1×1 convolutions play an essential role in NIN. The 1×1 convolutions have low computational consumption, while they can integrate cross-channel information. Furthermore, 1×1 convolutions can transform the number of channels without changing the tensor scale [58].

This research proposes a modified NIN network as a micro-network to integrate into the NI-U-Net++. Figure 5 depicts the structure of the proposed micro-network, which is the orange squares in Figure 4. The channel of the micro-network input tensor is N channels, and the first 3×3 convolution decreases it to $N/4$ channels. Then, the following two 1×1 convolutions can be understood as the fully connected layers along the channel axis. Finally, another 3×3 convolution restores the tensor channels to N channels. Thus, the micro-network ensures the proposed NI-U-Net++ can adopt a deep structure with a small computational graph.

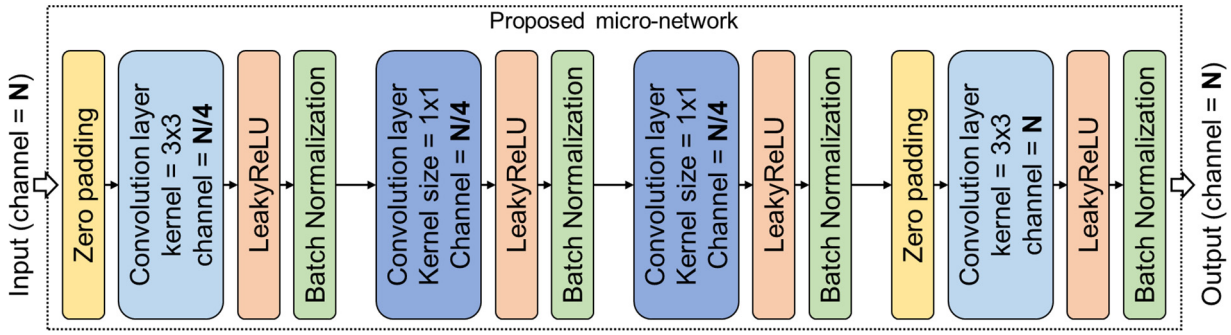

Figure 5. The layer configuration of the proposed micro-network. The yellow, light blue, orange, green, and dark blue squares refer to the zero-padding layer, convolution layer with 3 × 3 kernel size, LeakyReLU activation layer, and batch normalization layer, respectively.

There are three highlights in the proposed NI-U-Net++ rock segmentation network. (1) NI-U-Net++ does not determine the image scale-change in NI-U-Net++. NI-U-Net++ provides more flexible freedom of the scale-change than U-Net, and the task can automatically find the optimal scale. The scale-change refers to the optimal number of continuous downsampling operations before the decoder. (2) The strategy of deep supervision is adopted in the proposed NI-U-Net++. Zhou et al. mentioned that the shallow sub-U-Nets might be disconnected when the deep supervision is not activated [15]. To this end, deep supervision can provide the backpropagation to any sub-U-Nets. (3) The micro-network establishes the cross-channel data relevance in each scale of the segmentation network. (The further pairwise comparisons between proposed NI-U-Net++ and related studies can be found in Appendix A.4.)

2.4. The Pre-Training Process

The pre-training process aims to provide efficient prior knowledge for rock segmentation. The pre-training process divides the synthetic dataset into a training, validation, and testing set according to the ratios of 80%, 10%, and 10%. The hyperparameters are listed in Table 3. The number of epochs is set to 50 epochs, the batch size is set to 5 samples per batch, the learning rate is set to 0.00005, the optimizer adopts the Adam, and the binary cross-entropy loss is chosen as the loss function. The pixels of the rocks are annotated using value one, and the background pixels use value zero. Furthermore, the pre-training process uses six usual metrics to compare the proposed NI-U-Net++ to the related studies. The six metrics are accuracy, intersection over union (IoU), Dice score, root mean squared error (RMSE), and receiver operating characteristic curve (ROC). The related studies correspond to U-Net [14], U-Net++ [15], NI-U-Net [57], Furlan2019 [50], and Chiodini2020 [44].

Table 3. The hyperparameters of the pre-training process.

Hyperparameter	Setting	Hyperparameter	Setting
Epoch	50 epochs	Batch size	5 sample per batch
Learning rate	0.00005	Optimizer	Adam
Loss function	Binary cross-entropy	Training set ratio	80% of the synthetic dataset
Validation set ratio	10% of the synthetic dataset	Testing set ratio	10% of the synthetic dataset
Evaluation metrics	Accuracy, intersection over union (IoU), Dice score, root mean squared error (RMSE), and receiver operating characteristic curve (ROC)		

The chosen evaluation metrics come from the following reasons. (i) Loss function decides the learning gradient, and it is the specific factor for fitting conditions, converges, and the learning process. (ii) Accuracy refers to a very intuited indicator for knowing performance. (iii) IoU is a prevalent and influential metric in semantic segmentation studies, but it is also based on the confusion matrix as the accuracy. (iv) Dice score is a similar

metric. Thus, this research puts the Dice score in the Appendix A as additional results. (v) ROC indicates the sensitivity for different thresholds of positive and negative prediction.

It is noteworthy that the training, validation, and testing sets are saved to local storage to prevent the potential uncertainty from the dataset shuffle. Thus, any synthetic dataset mentioned in this study refers to the same data distribution. Some details of the related studies have been discussed as following:

(i) U-Net is proposed by Ronneberger et al. [14], which is a very popular one-stage image segmentation network [60,61]. The applied U-Net references the high-starred implementations on GitHub [62,63]. The encoder of U-Net contains four downsampling layers, the decoder contains four upsampling layers, and the activation uses the "ReLU" function. The size of each convolution kernel is 3×3.

(ii) U-Net++ is proposed by Zhou et al. [15] in 2018, which is an undated U-shaped network based on the U-Net. The applied U-Net++ references the high-stared implementation on GitHub [64]. The applied U-Net++ contains four downsampling layers, and the deep supervision has been activated.

(iii) The NI-U-Net [57] shares the same architecture as the sky and ground segmentation network used in Section 2.2. NI-U-Net only contains a single U-shaped encoder-decoder design, and the micro-networks have also been applied.

(iv) Furlan et al. proposed a deeplabv3plus-based rock segmentation solution in 2019, and the implementation of Furlan2019 referenced the study in [43].

(v) Chiodini et al. proposed a fully convolutional network-based rock segmentation solution in 2020; the implementation of Chiodini2020 referenced the study in [44].

2.5. The Transfer-Training Process

The aim of the transfer-training process is to fine-tune the NI-U-Net++ from the "Pre-trained weights" to the "Final weights" for the real-life images (see Figure 1). The "Annotated visual dataset" is divided into training, validation, and testing sets according to the ratio of 80%:10%:10% (similar to the pre-training process). The hyperparameters have been depicted in Table 4: the number of epochs is set to 50 epochs, the batch size is set to 5 samples per batch, the learning rate is set to 0.00005, the optimizer uses the Adam, and the loss function uses the binary cross-entropy. The evaluation also uses the three popular metrics, accuracy, IoU, and Dice score.

Table 4. The hyperparameters of the transfer-training process.

Hyperparameter	Setting	Hyperparameter	Setting
Epoch	50 epochs	Batch size	5 sample per batch
Learning rate	0.00005	Optimizer	Adam
Loss function	Binary cross-entropy	Training set ratio	80% of the synthetic dataset
Validation set ratio	10% of the synthetic dataset	Testing set ratio	10% of the synthetic dataset
Evaluation metrics	Accuracy, intersection over union (IoU), Dice score, root mean squared error (RMSE), and receiver operating characteristic curve (ROC)		

The data for the transfer-training process comes from "Annotation-2" in Figure 1. "Annotation-2" can be composed of the following four steps.

(1) "Annotation-2" randomly re-selects 150 images from the *Katwijk* dataset.

(2) "Annotation-2" performs pixel-level annotations on these images.

(3) The images annotated in "Annotation-1" can also be used for the transfer-training process, so "Annotation-2" merges 35 images in "Annotation-1" with the 150 images. It is noteworthy that there are two duplicate images, so the final number of images for "Annotation-2" is 183 images (The 183 images are only about 8% of the *Katwijk* dataset)

(4) "Annotation-2" uses data augmentation to simulate possible situations for the planetary rover operations. For example, rotations simulate the pose changes, brightness changes simulate changes in illumination conditions, and contrast changes simulate

changes in imaging conditions. The data augmentation eventually achieves about 4000 images.

3. Results and Discussion

All experiments in this research were conducted on the same data, hardware, and software. This research saved the random-shuffled dataset to ensure the repeatability for any experiment. The CPU, GPU, and memory size are Core i7-7700, NVIDIA RTX1080, and 32 GB. The deep learning platform, GPU parallel computing support, programming language, and operating system are TensorFlow 2.1, CUDA 10.1, Python 3.6, and Ubuntu 18.04.

3.1. The Results of the Proposed Synthetic Algorithm

The proposed synthetic algorithm (Section 2.2) generates 14,000 synthetic images as the synthetic dataset using the rock samples and real-life backgrounds from Section 2.2. Figure 6a visualizes an example in the synthetic dataset. The grayscale distributions between the rock samples and the real-life backgrounds can be significantly different. For example, directly embedding a rock sample extracted from a dark region to a bright region of the real-life background is not visually comfortable. The solid blue frames in Figure 6a refer to the embedded rocks, and the green dashed frames refer to the original rocks. The grayscale distributions of the embedded rocks are visually comfortable. Furthermore, Figure 6 illustrates some complex cases that usually appear in the practical planetary explorations (such as occlusion, unclosed outline, far and small target, etc.). These complex cases can significantly enforce the robustness and generalization-ability of the synthetic dataset. Figure 6b refers to the corresponding annotation of Figure 6a, which is the synthetic image.

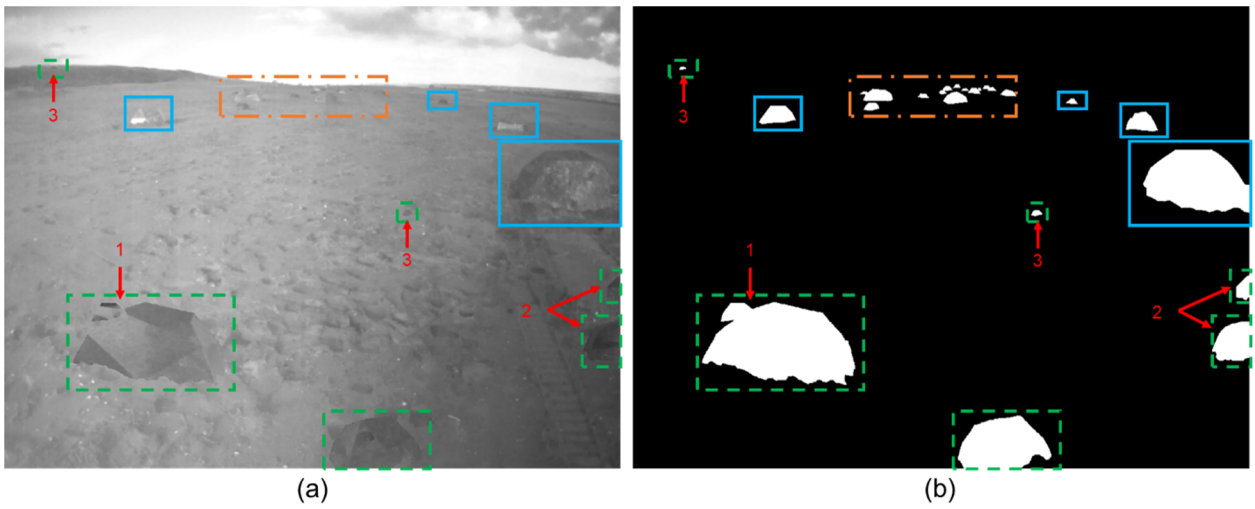

Figure 6. A typical example in the synthetic dataset. (**a**,**b**) refer to the synthetic image (from the proposed synthetic algorithm) and the simultaneously generated annotation. Blue solid frames, green dash frames, and orange dot-dash frames refer to original rocks, embedded synthetic rocks, and un-highlighted rocks, respectively. The reason for the un-highlighted rocks is their small and dense distribution, which can cause a bad visualization. "1", "2", and "3" highlight the complex cases of occlusion, unclosed outline, far, and small target, respectively. (**b**) uses white pixels to refer to the rocks, while black pixels refer to the background.

The target of the synthetic algorithm is to simulate real-life images as much as possible when generating the synthetic data. The difference between synthetic and real-life images comes from different imaging sources. Figure A1 in the Appendix A shows that the synthetic algorithm without well-optimization can cause an apparent visual difference. The materials used in the synthetic algorithm are all derived from real-life images to ensure visual comfort (such as rock samples and backgrounds). Furthermore, the synthetic

algorithm further optimizes visual comfort through the illumination intensity assumption. It is noteworthy that using synthetic data aims to assist rock segmentation in real-life images. Therefore, this research utilizes the results in real-life images to verify the capacity of the proposed synthetic algorithm (see Figure A2 and the demo video in the supplementary material).

3.2. The Results of the Pre-Training Process

This section compares the proposed NI-U-Net++ with five related studies. Table 5 describes the quantitative comparisons of the pre-training process. Figure 7 depicts the loss and accuracy curves of the training and validation sets for the proposed NI-U-Net++. Figures A4–A6 describe the loss and accuracy curves of U-Net, U-Net++, NI-U-Net, Furlan2019, and Chiodini2020, respectively. Dice scores have been described in Table A2 in the Appendix A. Figure 8 compares the ROC curve of the proposed NI-U-Net++ with the advanced studies from Furlan2019 [43] and Chiodini2020 [44].

Table 5. The results of the pre-training process.

Network	Loss			Accuracy			IoU			RMSE		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
U-Net	0.0360	0.0393	0.0397	99.13%	98.96%	98.95%	0.8446	0.8248	0.8255	0.1027	0.1039	0.1043
U-Net++	0.0121	0.0211	0.0209	99.56%	99.28%	99.28%	0.9182	0.8769	0.8783	0.0668	0.0744	0.0743
NI-U-Net	0.0102	0.0281	0.0280	99.63%	99.25%	99.24%	0.9313	0.8715	0.8720	0.0665	0.0775	0.0776
Furlan2019	0.0273	0.0307	0.0308	99.04%	98.86%	98.86%	0.9125	0.9005	0.9001	0.0912	0.0924	0.0926
Chiodini2020	0.0108	0.1724	0.1692	99.38%	97.98%	98.00%	0.9423	0.8299	0.8330	0.1298	0.1336	0.1328
NI-U-Net++	0.0117	0.0175	0.0173	99.58%	99.40%	99.41%	0.9209	0.8972	0.8991	0.0665	0.0775	0.0775

The "loss" refers to the binary cross-entropy used for training. The "accuracy", "IoU", and "RMSE" refer to the adopted evaluation metrics. The "train", "valid", and "test" refer to the results from the training, validation, and testing sets. U-Net, U-Net++, NI-U-Net, Furlan2019, and Chiodini2020 refer to the related studies [14,15,43,44,57], respectively. NI-U-Net++ refers to the proposed network. Gray shadings indicate the lowest loss, highest accuracy, highest IoU, and lowest RMSE.

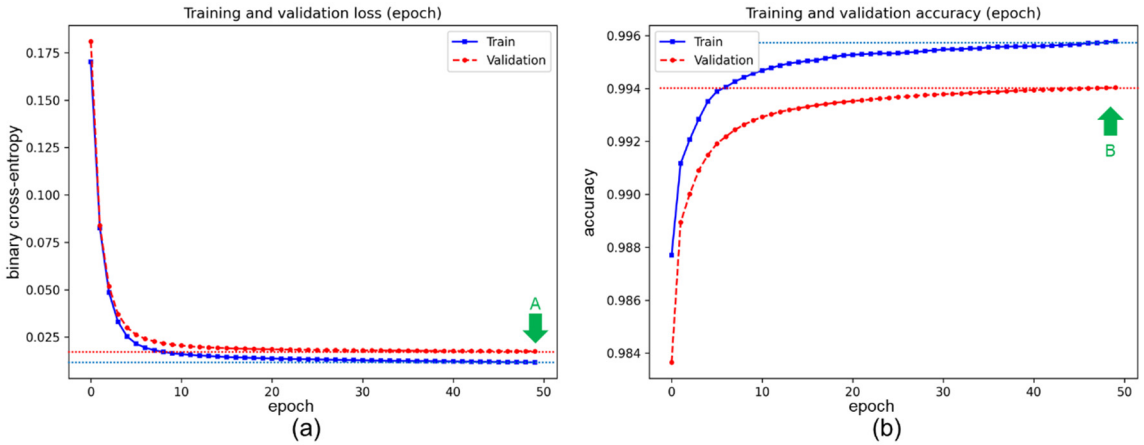

Figure 7. The loss and accuracy curves of NI-U-Net++ using the synthetic dataset. The green "A" and "B" correspond to the two highlights mentioned in the content of the NI-U-Net++ curves. (**a**) refers to the epoch-wised loss curves in the training and validation sets. (**b**) refers to the epoch-wised accuracy curves in the training and validation sets. The horizontal dash lines refer to the references of final converge status.

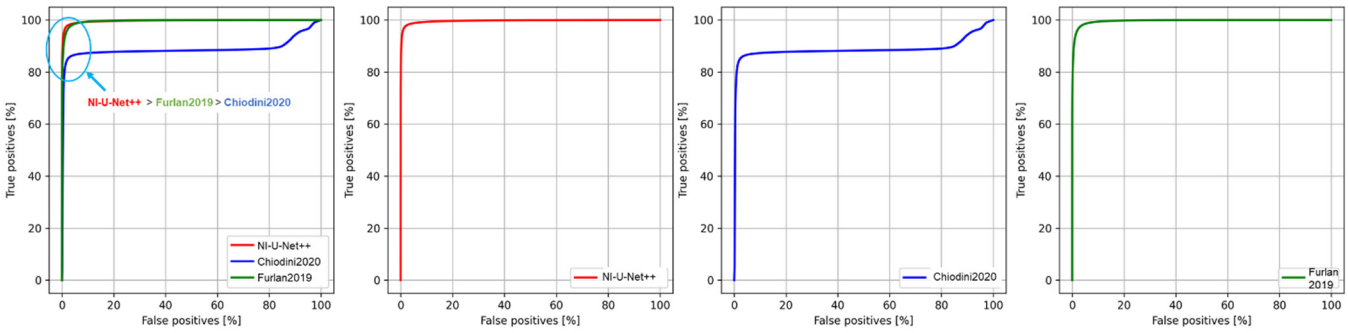

Figure 8. The ROC curves of the proposed NI-U-Net++, Furlan2019 [43], and Chiodini2020 [44].

The gray shadings in Tables 5 and A2 highlight the best results in each column. NI-U-Net and NI-U-Net++ show better performances than the U-Net and U-Net++ with a lower "loss" value and higher "accuracy", "IoU", and "Dice" values. This suggests that the proposed micro-network helps to improve the performance of rock segmentation. Moreover, Figures 6 and A6 both appear to have a more rapid initial learning speed compared to Figures A4 and A5. Thus, the proposed micro-network can accelerate the learning efficiency.

The NI-U-Net achieves the highest training accuracy (see Table 5). Arrow "A" in Figure A6a highlights a U-shaped rise that appears on the validation loss curve, while the training loss curve keeps decreasing. These indicate that overfitting occurs for NI-U-Net. This can explain that NI-U-Net achieved the lowest training loss and highest training accuracy, but the validation and testing loss and accuracy were poorer than others. The green arrow "B" in Figure A6b indicates that NI-U-Net produces the largest distance in accuracy between the training and validation sets. NI-U-Net is a modified U-Net using the micro-network. Compared with U-Net, all the results of NI-U-Net achieve improvements. Therefore, the proposed micro-network can also suppress the overfitting level.

Table 5 and arrow "A" in Figure A4a find that U-Net achieves a higher loss and lower accuracy. Thus, U-Net has the highest level of underfitting. Arrow "B" in Figure A4b highlights that accuracy curves keep flat at the first two epochs. This indicates that the learning process of U-Net is difficult. This comes from a fixed and high encoder ratio. The down-sampling operations in the encoder can cause significant information loss, especially for small targets.

U-Net++ also appears to have overfitting in Table 5. The U-Net++ training curves and the horizontal reference lines depict that the training curves keep learning throughout the pre-training process, while the validation curves come to the convergence (see the arrows "A" in Figure A5a and "B" in Figure A5b).

Table 5 shows that the proposed NI-U-Net++ achieves the lowest validation loss, lowest validation and testing loss, highest validation and testing accuracy, as well as lowest RMSE. The curves in Figure 6a,b appear to be promising learning trends. In the initial stage of training, it drops rapidly and then slowly converges. The arrows "A" and "B" in Figure 6a,b indicate that NI-U-Net++ stays stable on both the training and validation sets, and the overfitting level is low. The outstanding evaluation results indicate that the risk of underfitting is also low. NI-U-Net++ achieved the best pre-training results by improving the overall configuration and introducing the micro-network.

This research further applied two advanced related studies as the comparisons. (i) The "Chiodini2020" in Table 5 and Figure A7 indicates the results using Chiodini et al. [44]. The proposed NI-U-Net++ suppresses all qualitative results of Chiodini2020. Moreover, Chiodini2020 appears to have significant overfitting and unstable conditions on the validation set. (ii) The "Furlan2019" in Table 5 and Figure A8 indicates the results using Furlan et al. [43]. Furlan et al. applied a fully convolutional network (FCN)-based rock segmentation solution. Although Furlan2019 achieves higher IoU than the proposed

NI-U-Net++, it is only 0.1–0.3% higher than the proposed NI-U-Net. Furthermore, the proposed NI-U-Net++ achieves significantly better results in loss, accuracy, and RMSE.

Figure 9 depicts the visualizations of NI-U-Net++ from the pre-training process. Table A3 indicates the quantitative results of using different numbers of synthetic images, and the further discussion can be found in Appendix A.5.

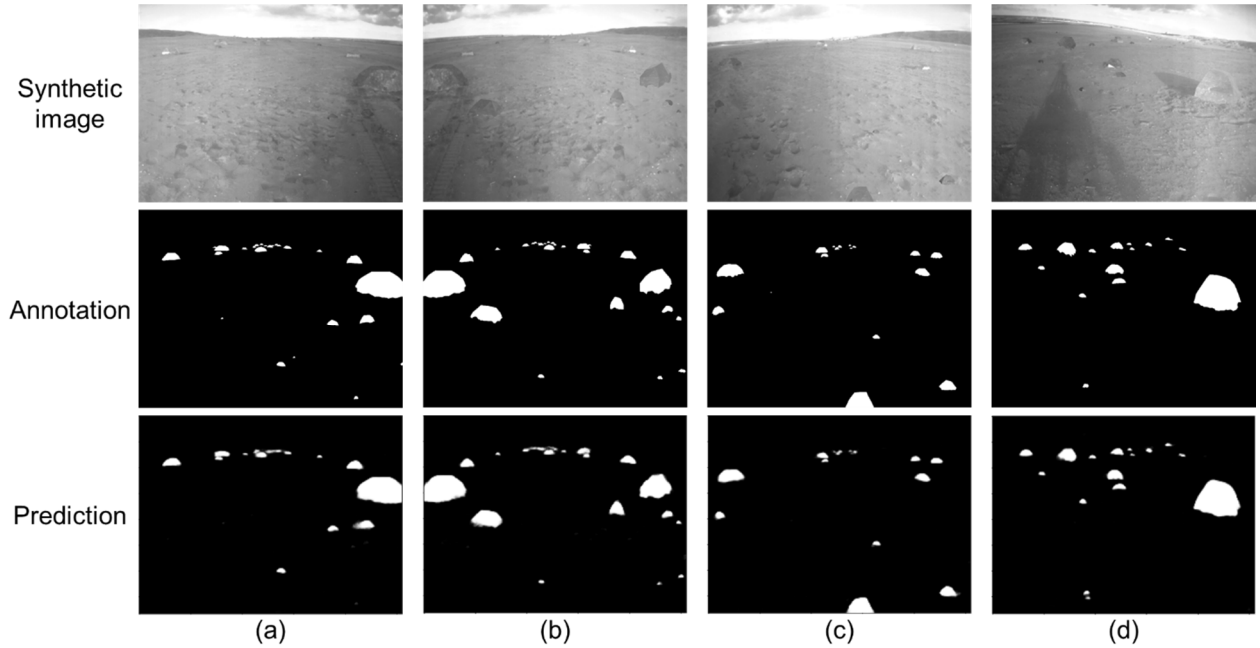

Figure 9. The example results of the NI-U-Net++ rock segmentation network in the pre-training process. "Synthetic image", "Annotation", and "Prediction" refer to the synthetic input images, the simultaneously generated ground truth annotations, and the predictions from the pre-trained NI-U-Net++, respectively. (**a**–**d**) correspond to four examples.

3.3. The Results of the Transfer-Training Process

The results of transfer learning are presented in Table 6. Figure 10 depicts the loss, accuracy, IoU, and Dice score curves on training and validation sets. Each curve comes to convergence with a smooth and stable trend. Thus, the model does not appear to be overfitting. Although transfer learning only used 183 images from the *Katwijk* dataset, the proposed synthetic algorithm and the transfer learning strategy accomplish a significantly low loss value, high accuracy, high IoU, and high Dice score. It is noteworthy that the used navigation vision has about 2500 images from the *Katwijk* dataset, and the transfer learning only uses about 8%. Furthermore, the good results of the metrics indicate that the NI-U-Net++ does not appear to be underfitting either.

Table 6. Result of transfer-training using the proposed NI-U-Net++.

Loss			**Accuracy**			**IOU**			**Dice**			**RMSE**		
Train	**Valid**	**Test**	**Train**	**Valid**	**Test**	**Train**	**Valid**	**Test**	**Train**	**Valid**	**Test**	**Train**	**Valid**	**Test**
0.0061	0.0183	0.0151	99.74%	99.54%	99.58%	0.8841	0.7223	0.7476	0.9384	0.8387	0.8556	0.0499	0.0594	0.0557

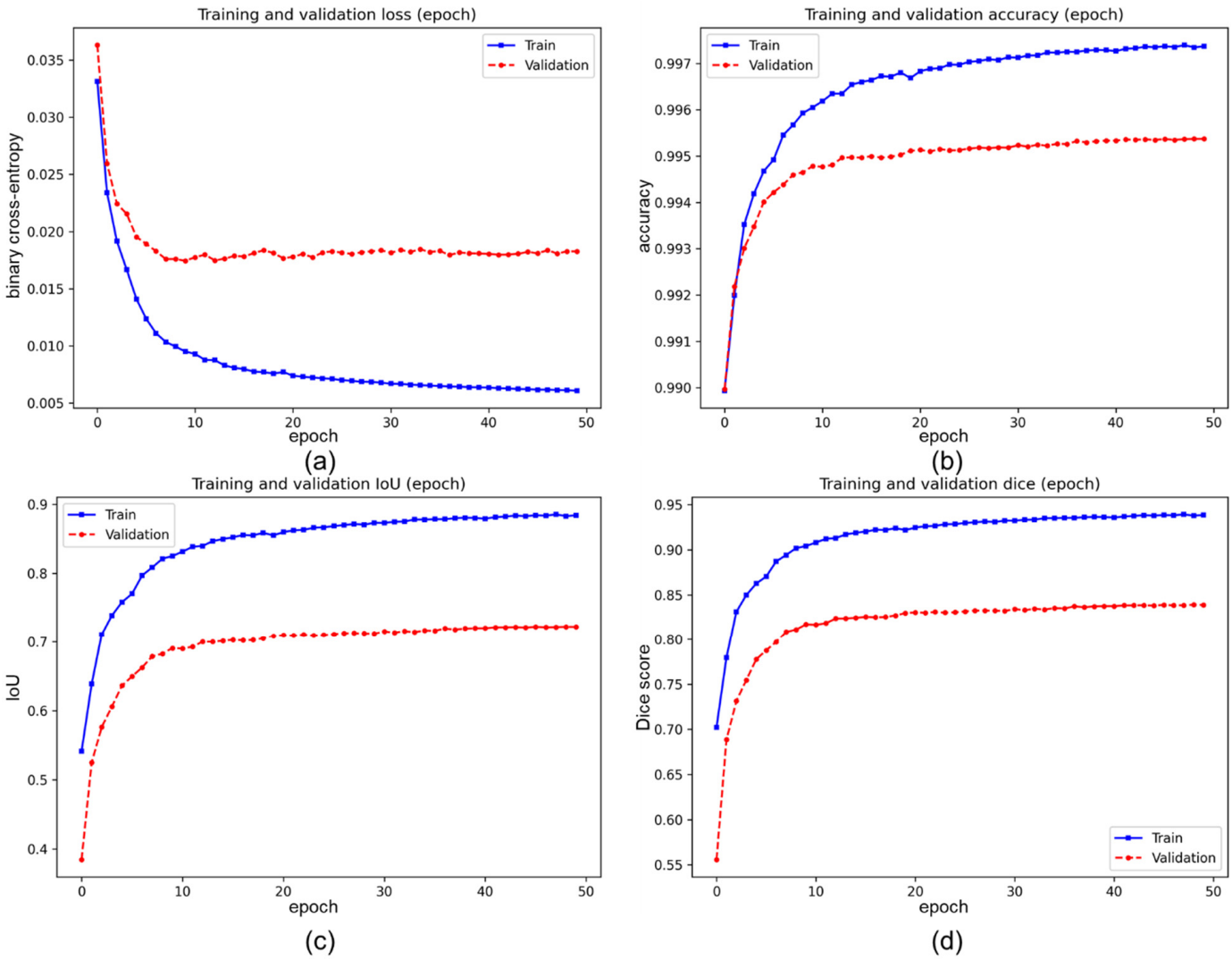

Figure 10. The transfer training curves using the "Annotation-2" dataset. (**a**–**d**) refer to the transfer training curves of the loss, accuracy, IoU, and Dice score, respectively. The blue solid curves and red dash curves refer to the training records from the training and validation sets.

Table 6 provides the quantitative records of the transfer learning process (the green frame in Figure 1). The loss value in Table 6 has reached a small magnitude, and the accuracy reaches a high level. Although IoU is not as high as in the pre-training process (see Table 5), it is already a considerably high value in the image segmentation topic (see the IoU level in [16,65,66]). The performance of the Dice score and RMSE are good. Figure 11 indicates the qualitative results of transfer learning. Figure 11 also involves the results using the pre-trained model. The pre-trained model can achieve highly similar predictions, which justified the help using the transfer learning strategy. The Supplementary Video S1 depicts the integration of the transfer learning achievement into the navigation vision of the planetary rovers. Compared with the frame rate of the original navigation vision (8 FPS), the processing speed of the proposed NI-U-Net++ is 32.57 FPS (or the inference time is 0.0307 s per frame), which is 4.071 times the frame rate in the original video. The details of the inference time can be found in Appendix A.6 in the Appendix A, which shows that the real-time performance of the proposed NI-U-Net++ appears excellent on the tested device.

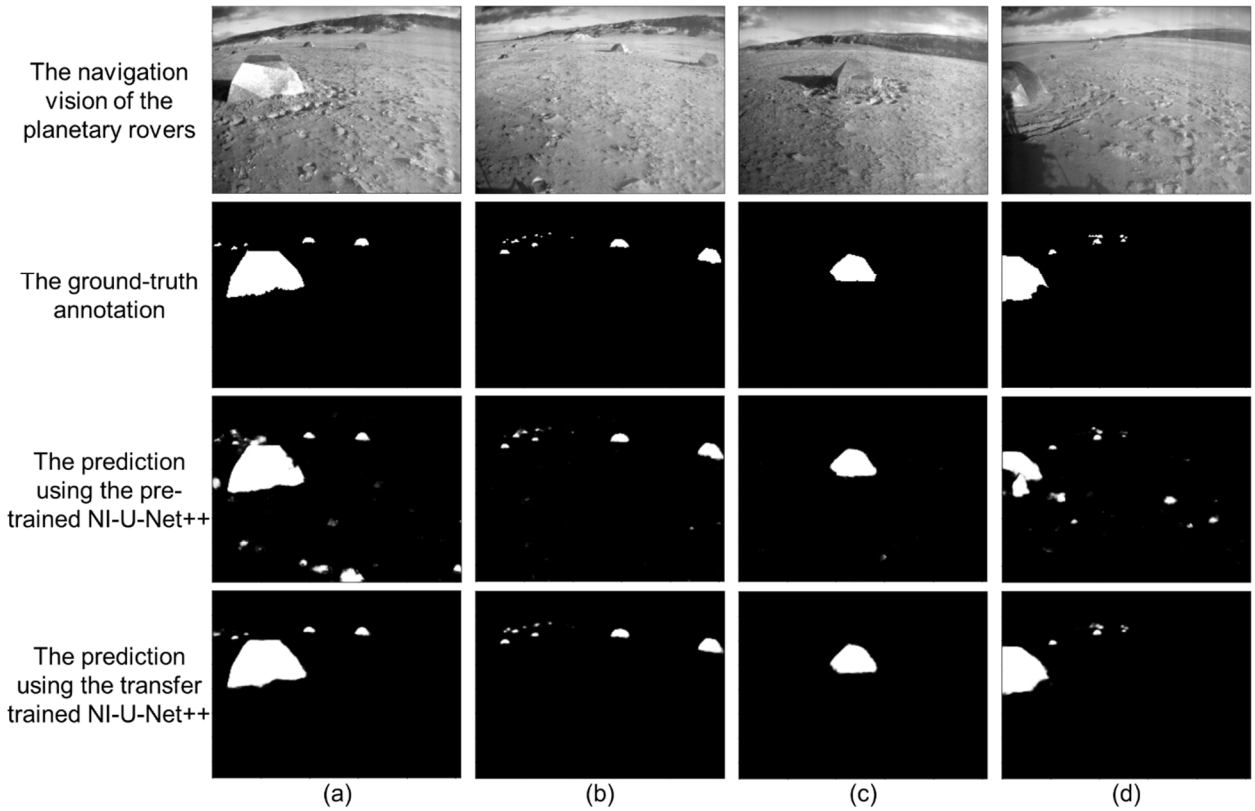

Figure 11. The visualized results of the proposed NI-U-Net++ in the transfer-training process. The navigation vision of the planetary rovers refers to the images from the Katwijk dataset. (**a**–**d**) refer to four variant selected images.

Notably, the quantitative results of the metrics between the transfer learning and the pre-training are not directly comparable. Compared with pre-training, the result of transfer learning is low (such as IoU in Table 5). As discussed in Section 2.2, the synthetic dataset is essentially generated using the incremental approach. The synthetic algorithm sets the scaling to be 0.6 to 1.0. The evaluation metrics (accuracy, IoU, and Dice) are all based on the statistical results of pixels. The embedded synthetic rock samples can be divided into two categories: the clustered pixels (that are easy to determine) and the edge pixels (that are not easy to determine). As the target size increases, the clustered pixels pull the overall metrics to a high level. Moreover, many situations do not appear in the pre-training dataset (such as significant changes in pose, brightness, illumination, sharpness, etc.), which enlarges the marginal probability distribution of the transfer-training process.

4. Conclusions and Future Works

This research proposed a rock segmentation framework for the navigation vision of the planetary rovers using the synthetic algorithm and transfer learning. This framework provided an end-to-end rock segmentation solution for the future planetary rover autonomy. Furthermore, the proposed synthetic algorithm provided a new idea for handling the challenge of the lack of pixel-level semantic annotations in the planetary explorations. The synthetic dataset also provided a valid dataset and benchmark for the related research. The proposed NI-U-Net++ achieved the best results (see Section 3.2) in all three popular metrics compared to the state-of-the-art (the accuracy, IoU, Dice score, and RMSE are 99.41%, 0.8991, 0.9459, and 0.0075, respectively). Moreover, both the pre-training and transfer-training processes achieved outstanding training curves and results (the accuracy, IoU, Dice score, and RMSE are 99.58%, 0.7476, 0.8556, and 0.0557, respectively), which proved the assumptions (of the proposed synthetic algorithm) in Section 2.2.

The proposed framework made a significant step in the semantic segmentation of unstructured planetary explorations. As a cheap and extensive sensor, the monocular

camera generates a large amount of data for planetary rover navigation. The proposed framework can efficiently conduct a semantic analysis for the planetary rover. These rocks can be integrated into the visual navigation system to further assist various advanced functions, such as path planning, localization, scene matching, etc.

The future works include transfering the proposed framework to the onboard device. The proposed framework uses the normal TensorFlow library, while only TensorFlow lite can operate on the onboard device. The potential action may also include the network slimming to fit the specific onboard device. Furthermore, the proposed NI-U-Net++ requires optimizations for the targeted system, hardware, and software.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/math9233048/s1, Video S1: The demo video of the proposed rock segmentation solution.

Author Contributions: Conceptualization, B.K.; methodology, B.K.; software, B.K. and M.W.; validation, B.K.; investigation, B.K., Z.A.R. and Y.Z.; resources, Z.A.R.; writing—original draft preparation, B.K.; writing—review and editing, B.K., M.W., Z.A.R. and Y.Z.; visualization, B.K., M.W., Z.A.R. and Y.Z.; supervision, Z.A.R. and Y.Z.; project administration, Z.A.R.; funding acquisition, Z.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: The Future Aviation Security Solutions (FASS) Programme, a joint Department for Transport and Home Office initiative with support from Connected Places Catapult (CPC) part funded this research.

Data Availability Statement: The proposed dataset is openly available in Cranfield Online Research Data (CORD) at https://doi.org/10.17862/cranfield.rd.16958728. The Katwijk beach planetary rover dataset is available at https://robotics.estec.esa.int/datasets/katwijk-beach-11-2015/.

Acknowledgments: All simulations have been carried out using the HPC facility (HILDA) at Digital Aviation Research and Technology Centre (DARTeC), Cranfield.

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

Nomenclature

X, Y, $\overline{X}$, $\overline{Y}$	Formal parameters.
I	Light source.
$i_1 - i_3$, i	Rays of light.
g	Horizontal ground.
θ_1, θ_2, θ_3, θ	Angles between the corresponding ray of the Light and the horizontal ground.
ρ	The density of rays in the unit ground area.
L	Light intensity.
S	Area of a ground region.
p_1, p_2	The boundary rays between object and camera.
l_b	The normal line perpendicular to the phase plane.
XY	The area between X and Y cross-points.
$G_1 - G_3$, $\overline{G_1} - \overline{G_3}$	Cross points on the ground.
P	Cross point between p_1 and PG_3.
O	The origin of image plane.
L_{XY}	The light intensity in the area between X and Y cross-points in the sketch.
P_{opt}	Abstracted value of the optical properties.
c_T	A variable to pack all factors related to optical properties.
$\overline{\rho}$	An approximate value of ρ.
T	The target (rocks in this research) in the sketch.
(x, y)	Coordinate.
$pixel_{img}(x, y)$	Grayscale value at coordinate (x, y).
N_{pixel}	The number of pixels in a specific region.
$\Delta\rho$	The difference between ρ and $\overline{\rho}$.

$\overline{XY}$	The area between $\overline{X}$ and $\overline{Y}$ cross-points.
f_1	An implicit function to correlate the optical properties and image grayscales.
img_{mean}	The averaged grayscale value for the corresponding image area.
img	A set of the grayscale values for the corresponding image area.
img_{Δ}	A set of the differential values between img and img_{mean} (only related to the coordinates).
C	The constants to correct $\overline{\rho}$ from ρ.
$\overline{L}$	An approximation of L using the proposed synthetic algorithm.

Appendix A

Appendix A.1 Further Details of the Related Studies

Table 1. The detailed results of the related studies of Table 1.

Reference	Category in Table 1 [1]	Results [2]
[30]	i	Only the qualitative segmentation results.
[31]	i	Only the qualitative segmentation results.
[32]	i	Only the qualitative segmentation results.
[4]	ii	Fit error = 1.504~114.934.
[5]	ii and iii	Only the qualitative segmentation results.
[33]	ii	Only the qualitative segmentation results.
[34]	ii	(1) Average precision = 89% (the center matching method); (2) Average precision = 87% (the overlap method); (3) Average precision $\geq$ 90% = 83 images.
[35]	ii and v	(1) Standard deviation of recall = 0.2–0.3; (2) Standard deviation of precision = 0.2–0.3; (3) Recall and precision are modestly improved.
[36]	ii	Only the qualitative segmentation results.
[37]	iii	Only the qualitative segmentation results.
[38]	iii	(1) RMS error (X) = 0.22–0.93 (HiRISE pixel), RMS error (Y) = 0.22–0.97 (HiRISE pixel); (2) RMS error (X) = 0.23–0.70 (HiRISE pixel), RMS error (Y) = 0.23–0.89 (HiRISE pixel).
[32]	iv	CPU time (seconds): 0.2214–0.7484 (MAD); 0.1966–0.6955 (LMedsq); 0.5994–2.2033 (IKOSE); 0.2931–0.9633 (PDIMSE); 0.0380–0.1238 (RANSAC); 106.4747–236.2487 (RECON).
[39]	iv	(1) Processing time: 2–3 s for 256 * 256 images; 20–45 s for 640 * 480 images (2) Only the qualitative segmentation results.
[40]	iv	Medium rock match is successful (up to 26 m).
[41]	iv	The proposed method is robust and efficient for small- and large-scale rock detection.
[8]	v	A survey for terrain classification (including rock segmentation).
[27]	v	(1) Pixel-wise accuracy = 99.69% (background); 97.89% (sand); 89.33% (rock); 96.33% (gravel); 89.73% (bedrock). (2) Mean intersection-over-union (mIoU) = 0.9459.
[28]	v	(1) Accuracy = 76.2% (derivable terrain comprising sand, bedrock, and loose rock); (2) Accuracy = 89.2% (embedded pointy rocks).
[42]	v	mIoU = 0.93; recall = 96%; frame rate = 116 frame per second
[43]	v	F-score = 78.5%
[44]	v	Accuracy = 90~96%; IoU = 0.21~0.58.

[1] "i", "ii", "iii", "iv", and "v" in column "Category in Table 1" correspond to the same category index in Table 1. [2] The "Results" column only provides a statistic summary among the results of related studies. The exact values have no comparability either between each other or with this research. The reason comes from the different research focus, applied data, and experimental environments. The valid quantitative comparisons can be found in Section 3.2.

A.2. The Experiments of the Values in Table 2

Figure A1 illustrates the examples using different settings for the constant value C. It is noteworthy that C in Table 2 aims to correct the difference from the " $\rho \approx \overline{\rho}$" in Equation (4). According to the assumption in Section 2.2.1, this difference is small but sensitive to rock property. The goal of C is to optimize the visual comfort mentioned in Section 3.1. Therefore, C is affected by two settings, range and scale.

The range of the pixel value is between 0 and 256. (i) In Table 2, the range of C is set from 0 to 50. Thus, the maximum adjustment is less than 20% of the pixel value. Figure A1b shows a synthetic example using Table 2, and the white pixels in Figure A1a highlight the rocks. Figure A1a,b depict that Table 2 is a visually comfortable setting. Figure A1c,d respectively increase the range of C by three times and five times, and the embedded rock samples are very unreal to the background. Figure A1e reduces the range of C to about half of Figure A1b. The inserted rock is too bright compared to the background. Therefore, the range setting of C in Table 2 is in a reasonable range. (ii) C in Table 2 is divided into 11 scales according to the corresponding conditions (img_{mean}). A higher img_mean corresponds to a higher C. Figure A1f doubles the scale-setting to 21 scales, but there is no significant change compared to Figure A1b. Therefore, the classification method in Table 2 is also reasonable.

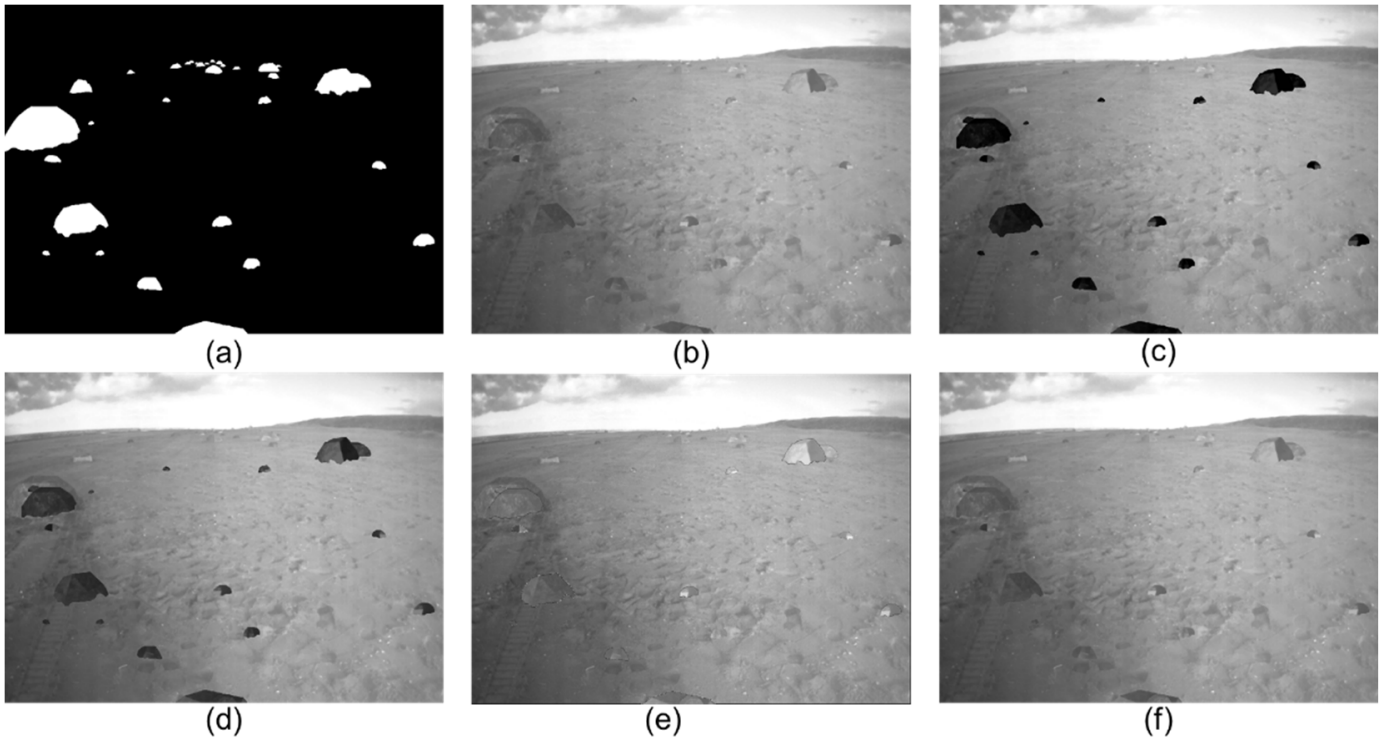

Figure A1. The visualized results of the experiments for the constant C in Table 2. (**a**) refers to the synthetic annotation, while (**b**–**f**) corresponds to the synthetic images through different settings of C. (**b**) applies the same setting as Table 2. (**c**) keeps the grade setting but increases the range of the C, the maximum C is set to 250. (**d**) refers to the results of only increasing maximum C to 150. (**e**) refers to the results of decreasing maximum C to 20. (**f**) keeps the maximum C as the same as (**b**), while the grade setting applies 21 grades.

Appendix A.3 Qualitative Examples of the Proposed Synthetic Dataset

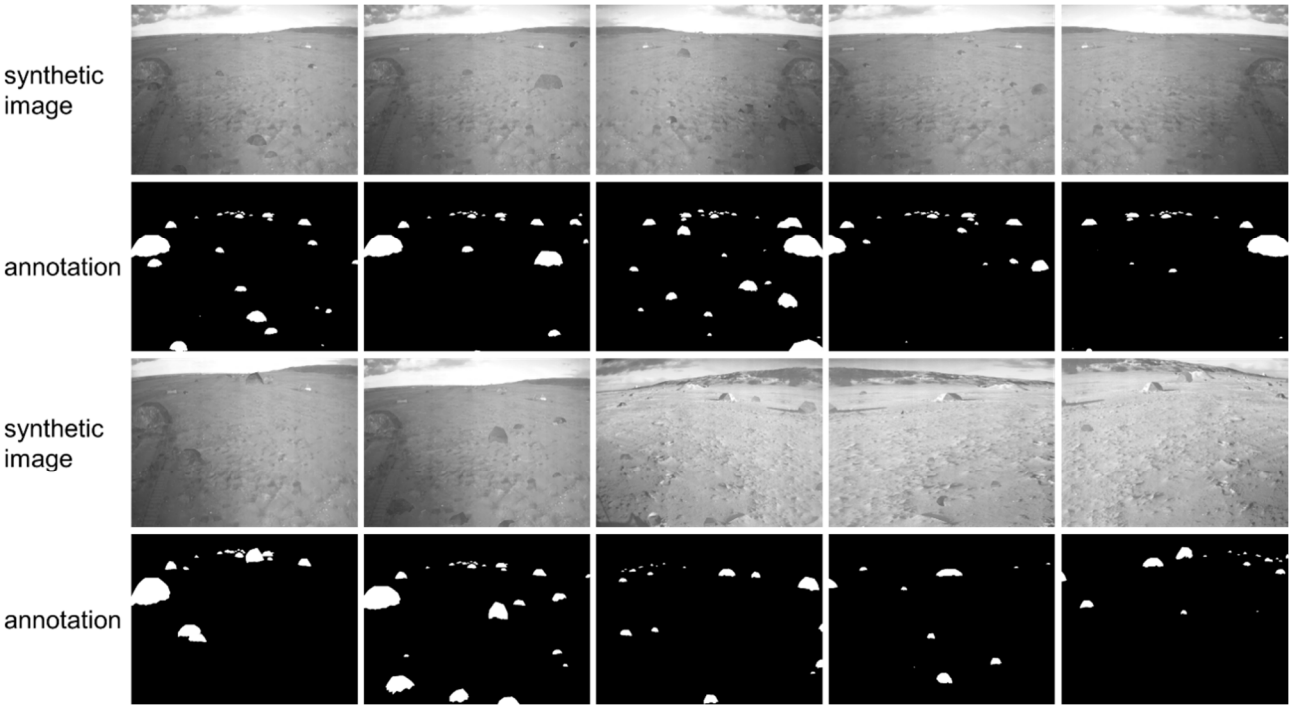

Figure A2. Some examples from the synthetic dataset. "Synthetic image" and "annotation" refer to the synthetic images and corresponding annotations, respectively. Annotations use white and black pixels to represent the rock and background pixels.

Appendix A.4 Pairwise Comparisons between Proposed NI-U-Net++ and Related Studies

This research uses Figures 3 and A3 in Section 2.3 to discuss the pairwise comparison between NI-U-Net++ and related studies in Table 5. Figure A3 uses NI-U-Net++ as background, but some highlights have been added for further comparison. The red arrows refer to the sub-U-Nets; each of them has a complete encoder–decoder process. Here defines a concept of compression ratio, which is the ratio between the input and output size (height or weight) of the encoder. "Sub-U-Net No.1" has the highest compression ratio, while "Sub-U-Net No.4" has the lowest compression ratio. The blue dash frame highlights the deep supervision mentioned in Section 2.3, and the orange frames refer to the micro-networks.

i. NI-U-Net++ with U-Net [14,63]:
   a. U-Net only has the "Sub-U-Net No. 1". Therefore, the compression ratio is constant at a high level.
   b. U-Net does not have deep supervision design.
   c. U-Net utilizes the 3×3 convolution layers and "Relu" activation instead of the micro-network in Figures 3, 4 and A3.

ii. NI-U-Net++ with U-Net++ [15]:
   a. U-Net++ also has four sub-U-Nets as in the NI-U-Net++.
   b. U-Net++ also has the deep supervision as in the NI-U-Net++.
   c. However, the U-Net++ applies the 3×3 convolution layer and "Relu" activation as in U-Net instead of the micro-network in NI-U-Net++.

iii. NI-U-Net++ with NI-U-Net [57]:
   a. NI-U-Net only has the "Sub-U-Net No. 1". Therefore, the compression ratio is constant at a high level.
   b. NI-U-Net has not deep supervision design.
   c. NI-U-Net utilizes the same micro-network as in NI-U-Net++.

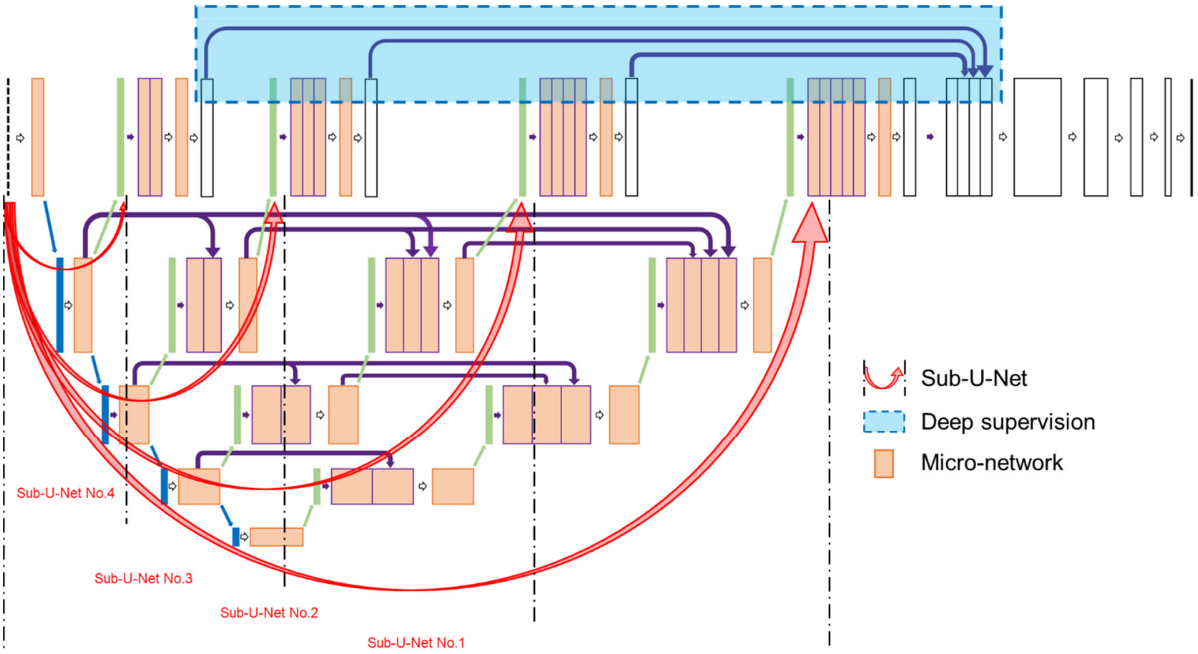

Figure A3. The pairwise comparisons for U-Net, U-Net++, NI-U-Net, and proposed NI-U-Net++.

Appendix A.5 Additional Results of the Pre-Training Process

Table A2. The Dice score of U-Net, U-Net++, NI-U-Net, and NI-U-Net++.

Networks	Dice Score		
	Train	Valid	Test
U-Net	0.9158	0.9040	0.9044
U-Net++	0.9574	0.9344	0.9352
NI-U-Net	0.9644	0.9313	0.9316
NI-U-Net++	0.9588	0.9458	0.9469

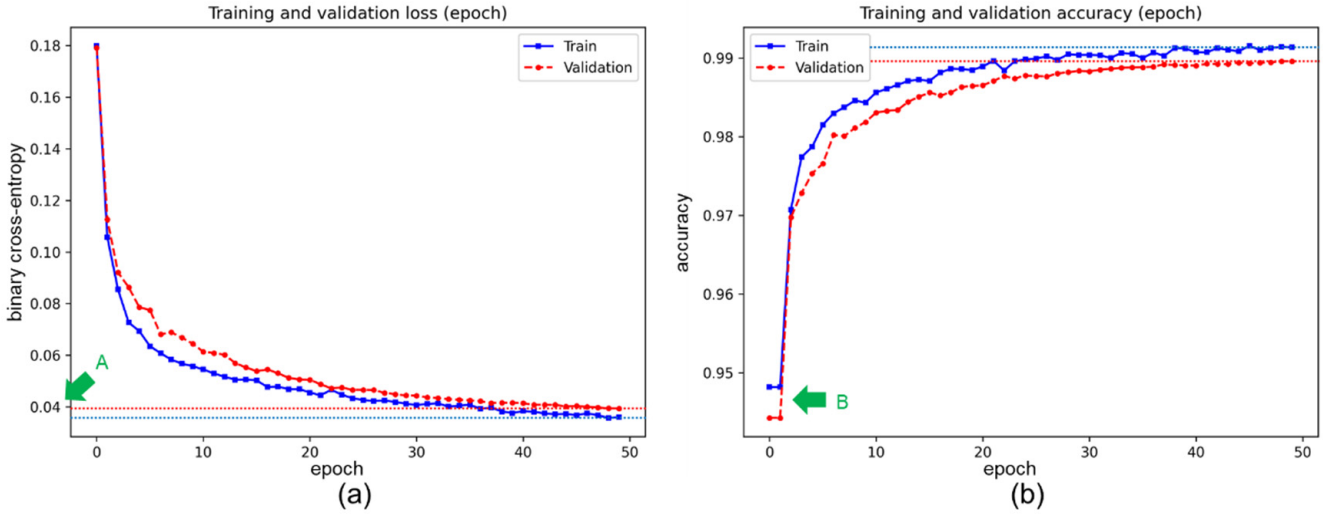

Figure A4. The loss and accuracy curves of U-Net [14] using the synthetic dataset. The green "A" and "B" correspond to the two highlights mentioned in Section 3.2. (**a**) Refers to the epoch-wised loss curves in the training and validation sets. (**b**) Refers to the epoch-wised accuracy curves in the training and validation sets. The horizontal dash lines refer to the references of final converge status.

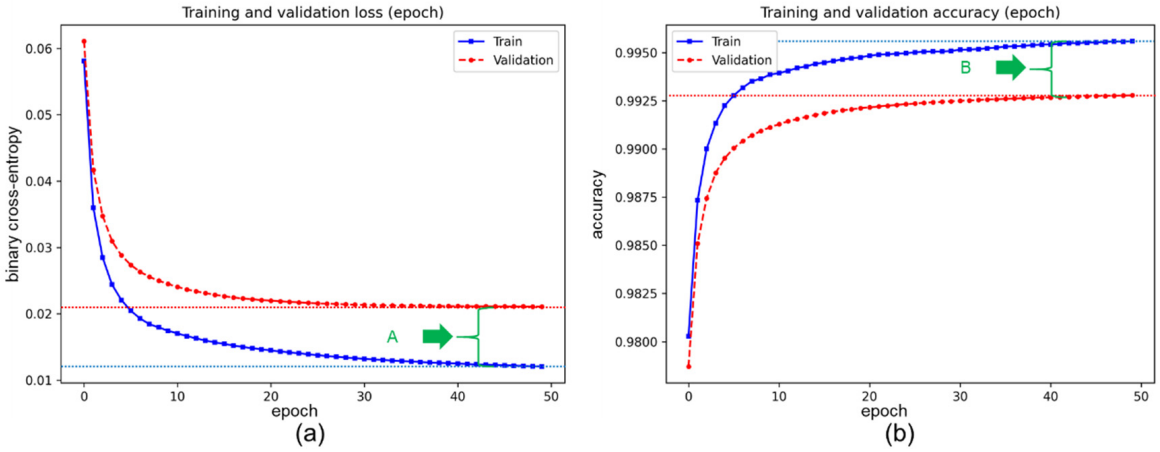

Figure A5. The loss and accuracy curves of U-Net++ [15] using the synthetic dataset. The green "A" and "B" correspond to the two highlights mentioned in Section 3.2. (**a**) Refers to the epoch-wised loss curves in the training and validation sets. (**b**) Refers to the epoch-wised accuracy curves in the training and validation sets. The horizontal dash lines refer to the references of final converge status.

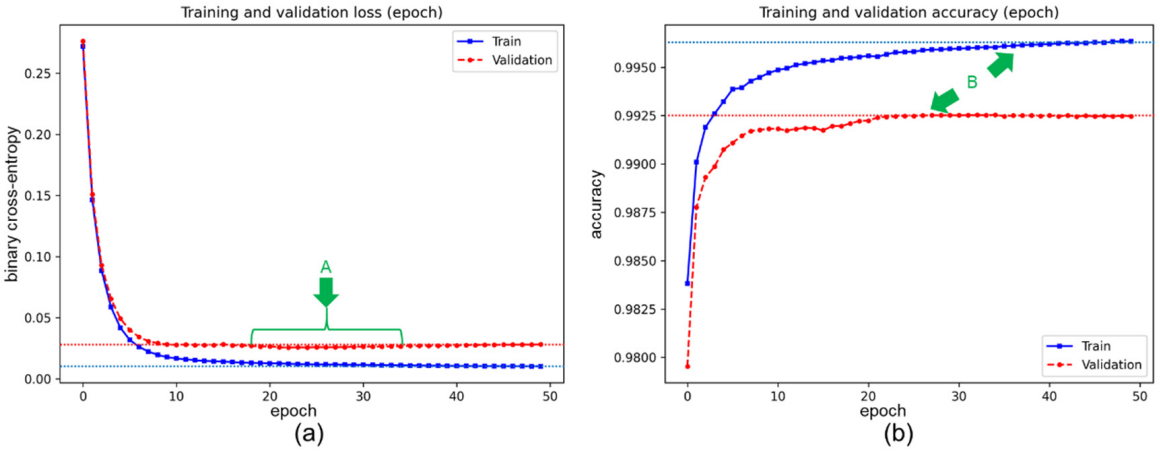

Figure A6. The loss and accuracy curves of NI-U-Net [57] using the synthetic dataset. The green "A" and "B" correspond to the two highlights mentioned in Section 3.2. (**a**) Refers to the epoch-wised loss curves in the training and validation sets. (**b**) Refers to the epoch-wised accuracy curves in the training and validation sets. The horizontal dash lines refer to the references of final converge status.

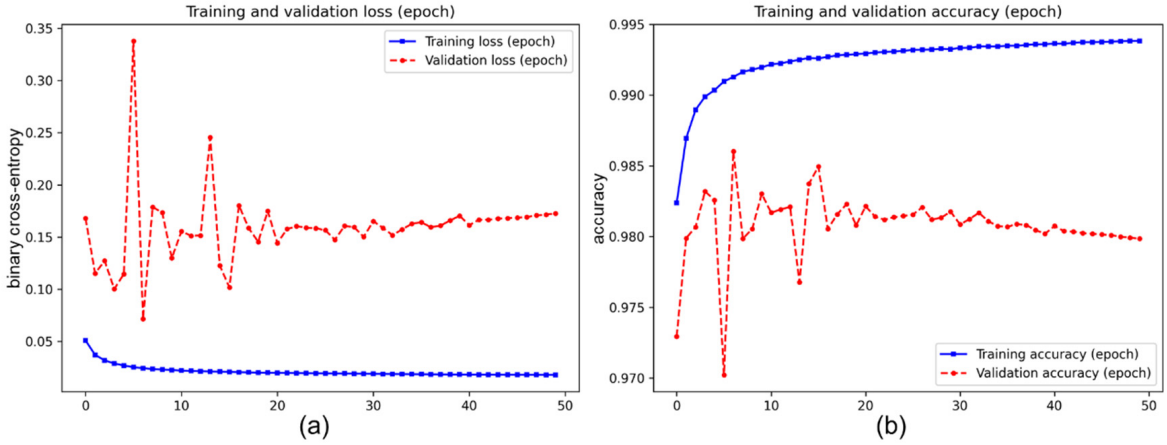

Figure A7. The loss and accuracy curves of Chiodini2020 [44] using the synthetic dataset. (**a**) Refers to the epoch-wised loss curves in the training and validation sets. (**b**) Refers to the epoch-wised accuracy curves in the training and validation sets.

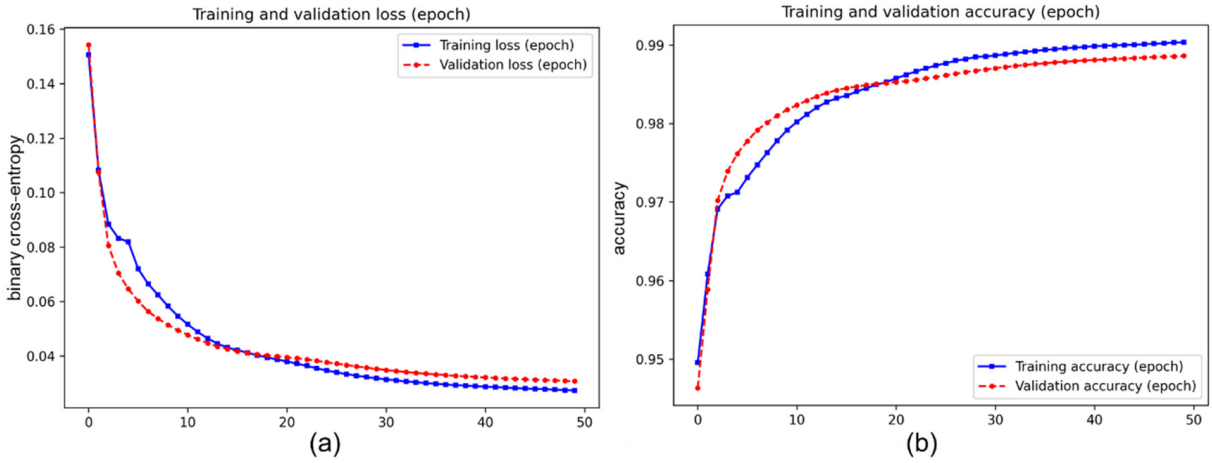

Figure A8. The loss and accuracy curves of Furlan2019 [43] using the synthetic dataset. (**a**) Refers to the epoch-wised loss curves in the training and validation sets. (**b**) Refers to the epoch-wised accuracy curves in the training and validation sets.

Table A3 refers to the results of training NI-U-Net++ using different numbers of synthetic images. This research chooses about 50% (7000 images) and 10% (1000 images) as two experiment settings to evaluate the impact when the number of synthetic images decreases. All results decrease when the images number decreases. The synthetic algorithm aims to generate a large amount of valid data, so applying all available data is more fitted to the target of this research.

Table A3. The quantitative results of NI-U-Net++ tested using a different number of synthetic images.

Number [1] (Images)	Loss			Accuracy			IoU			Dice Score		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
7000	0.0137	0.0189	0.0199	99.49%	99.39%	99.37%	0.9164	0.8919	0.8876	0.9564	0.9429	0.9405
1000	0.0273	0.0618	0.0580	99.53%	99.02%	99.03%	0.9175	0.8374	0.8389	0.9570	0.9115	0.9124

[1] "Number" refers to the number of synthetic images used in corresponding experiment.

Appendix A.6 Additional Results of the Transfer-Training Process

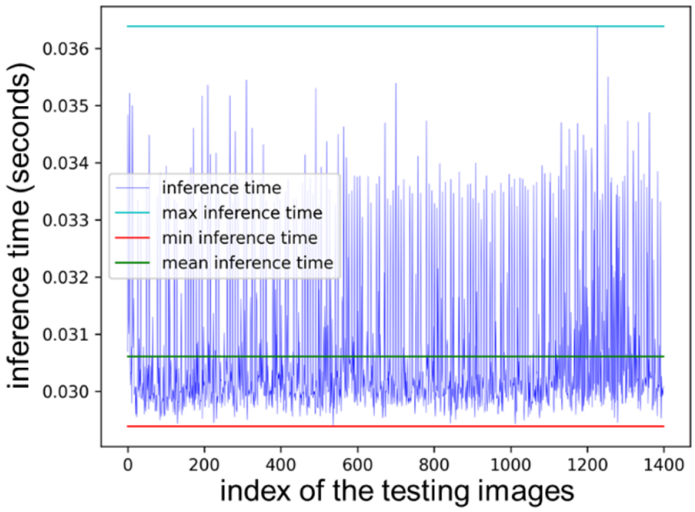

Figure A9. The inference time record. The max, min, and mean inference time is 0.0364, 0.0307, and 0.0294 s.

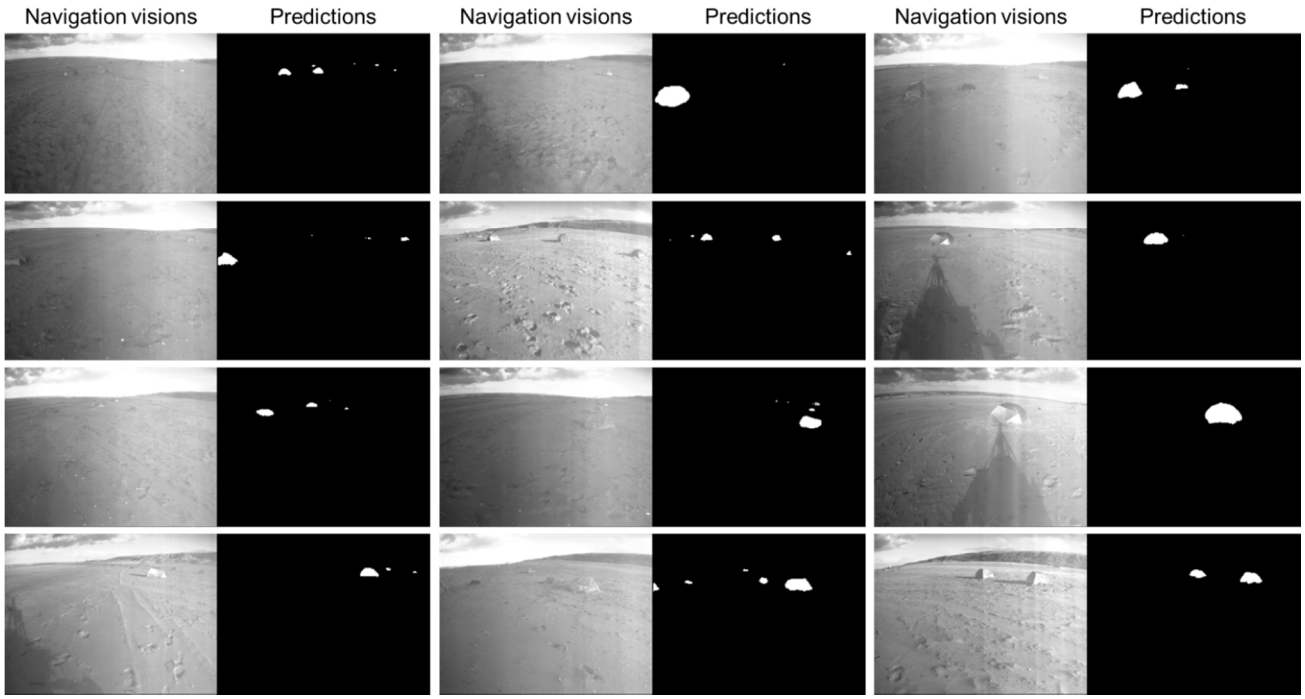

Figure A10. Some examples of the real-life rover vision and corresponding predictions.

References

1. Privitera, C.M.; Stark, L.W. Human-vision-based selection of image processing algorithms for planetary exploration. *IEEE Trans. Image Process.* **2003**, *12*, 917–923. [CrossRef] [PubMed]
2. Kim, W.S.; Diaz-Calderon, A.; Peters, S.F.; Carsten, J.L.; Leger, C. Onboard centralized frame tree database for intelligent space operations of the Mars Science Laboratory rover. *IEEE Trans. Cybern.* **2014**, *44*, 2109–2121. [CrossRef]
3. Gao, Y.; Chien, S. Review on space robotics: Toward top-level science through space exploration. *Sci. Robot.* **2017**, *2*, eaan5074. [CrossRef]
4. Castano, R.; Estlin, T.; Gaines, D.; Chouinard, C.; Bornstein, B.; Anderson, R.C.; Burl, M.; Thompson, D.; Castano, A.; Judd, M. Onboard autonomous rover science. In Proceedings of the 2007 IEEE Aerospace Conference, Big Sky, MT, USA, 3–10 March 2007; pp. 1–13.
5. Estlin, T.A.; Bornstein, B.J.; Gaines, D.M.; Anderson, R.C.; Thompson, D.R.; Burl, M.; Castaño, R.; Judd, M. AEGIS automated science targeting for the MER opportunity rover. *ACM Trans. Intell. Syst. Technol.* **2012**, *3*, 1–19. [CrossRef]
6. Otsu, K.; Ono, M.; Fuchs, T.J.; Baldwin, I.; Kubota, T. Autonomous terrain classification with co- and self-training approach. *IEEE Robot. Autom. Lett.* **2016**, *1*, 814–819. [CrossRef]
7. Swan, R.M.; Atha, D.; Leopold, H.A.; Gildner, M.; Oij, S.; Chiu, C.; Ono, M. AI4MARS: A dataset for terrain-aware autonomous driving on Mars. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021.
8. Gao, Y.; Spiteri, C.; Pham, M.-T.; Al-Milli, S. A survey on recent object detection techniques useful for monocular vision-based planetary terrain classification. *Robot. Auton. Syst.* **2014**, *62*, 151–167. [CrossRef]
9. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, 1–22. [CrossRef]
10. Liu, D.; Bober, M.; Kittler, J. Visual semantic information pursuit: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1404–1422. [CrossRef]
11. Zoller, T.; Buhmann, J.M. Robust image segmentation using resampling and shape constraints. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1147–1164. [CrossRef]
12. Alpert, S.; Galun, M.; Brandt, A.; Basri, R. Image segmentation by probabilistic bottom-up aggregation and cue integration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 315–327. [CrossRef]
13. Saltzer, J.H.; Reed, D.P.; Clark, D.D. End-to-end arguments in system design. *ACM Trans. Comput. Syst.* **1984**, *2*, 277–288. [CrossRef]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9351, pp. 234–241. ISBN 9783319245737.

15. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11045, pp. 3–11. [CrossRef]
16. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv* **2014**, arXiv:1412.7062.
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot multibox detector. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9905, pp. 21–37. ISBN 9783319464473.
18. Gupta, S.; Girshick, R.; Arbeláez, P.; Malik, J. Learning rich features FROM RGB-D images for object detection and segmentation. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8695, pp. 345–360. ISBN 9783319105833.
19. Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. Simultaneous detection and segmentation. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8695, pp. 297–312. ISBN 9783319105833.
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 386–397. [CrossRef]
21. Dewan, A.; Oliveira, G.L.; Burgard, W. Deep semantic classification for 3D LiDAR data. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; Volume 2017, pp. 3544–3549.
22. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
23. Teichmann, M.; Weber, M.; Zollner, M.; Cipolla, R.; Urtasun, R. MultiNet: Real-time joint semantic reasoning for autonomous driving. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Suzhou, China, 26–30 June 2018; Volume 2018, pp. 1013–1020.
24. Wu, B.; Wan, A.; Yue, X.; Keutzer, K. SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 1887–1893.
25. Busquets, D.; Sierra, C.; López de Màntaras, R. A multiagent approach to qualitative landmark-based navigation. *Auton. Robots* **2003**, *15*, 129–154. [CrossRef]
26. Kovács, G.; Kunii, Y.; Maeda, T.; Hashimoto, H. Saliency and spatial information-based landmark selection for mobile robot navigation in natural environments. *Adv. Robot.* **2019**, *33*, 520–535. [CrossRef]
27. Zhou, R.; Ding, L.; Gao, H.; Feng, W.; Deng, Z.; Li, N. Mapping for planetary rovers from terramechanics perspective. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 1869–1874.
28. Ono, M.; Fuchs, T.J.; Steffy, A.; Maimone, M.; Jeng, Y. Risk-aware planetary rover operation: Autonomous terrain classification and path planning. In Proceedings of the 2015 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2015; pp. 1–10.
29. Zhou, F.; Arvidson, R.E.; Bennett, K.; Trease, B.; Lindemann, R.; Bellutta, P.; Iagnemma, K.; Senatore, C. Simulations of Mars rover traverses. *J. Field Robot.* **2014**, *31*, 141–160. [CrossRef]
30. Pedersen, L. Science target assessment for Mars rover instrument deployment. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and System, Lausanne, Switzerland, 30 September–4 October 2002; Volume 1, pp. 817–822.
31. Di, K.; Yue, Z.; Liu, Z.; Wang, S. Automated rock detection and shape analysis from mars rover imagery and 3D point cloud data. *J. Earth Sci.* **2013**, *24*, 125–135. [CrossRef]
32. Xiao, X.; Cui, H.; Tian, Y. Robust plane fitting algorithm for landing hazard detection. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 2864–2875. [CrossRef]
33. Dunlop, H.; Thompson, D.R.; Wettergreen, D. Multi-scale features for detection and segmentation of rocks in Mars images. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–7.
34. Castano, R.; Judd, M.; Estlin, T.; Anderson, R.C.; Gaines, D.; Castano, A.; Bornstein, B.; Stough, T.; Wagstaff, K. Current results from a rover science data analysis system. In Proceedings of the 2005 IEEE Aerospace Conference, Big Sky, MT, USA, 5–12 March 2005; Volume 2005, pp. 356–365.
35. Castano, R.; Mann, T.; Mjolsness, E. Texture analysis for Mars rover images. In *Applications of Digital Image Processing XXII*; Tescher, A.G., Ed.; Society of Photo-optical Instrumentation Engineers: Bellingham, WA, USA, 1999; Volume 3808, pp. 162–173.
36. Burl, M.C.; Thompson, D.R.; DeGranville, C.; Bornstein, B.J. Rockster: Onboard rock segmentation through edge regrouping. *J. Aerosp. Inf. Syst.* **2016**, *13*, 329–342. [CrossRef]
37. Castafio, R.; Anderson, R.C.; Estlin, T.; DeCoste, D.; Fisher, F.; Gaines, D.; Mazzoni, D.; Judd, M. Rover traverse science for increased mission science return. In Proceedings of the 2003 IEEE Aerospace Conference Proceedings, Big Sky, MT, USA, 8–15 March 2003; Volume 8, pp. 8_3629–8_3636. Available online: https://ieeexplore.ieee.org/document/1235546 (accessed on 26 November 2021).
38. Di, K.; Liu, Z.; Yue, Z. Mars rover localization based on feature matching between ground and orbital imagery. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 781–791. [CrossRef]
39. Gulick, V.C.; Morris, R.L.; Ruzon, M.A.; Roush, T.L. Autonomous image analyses during the 1999 Marsokhod rover field test. *J. Geophys. Res. Planets* **2001**, *106*, 7745–7763. [CrossRef]

40. Li, R.; Di, K.; Howard, A.B.; Matthies, L.; Wang, J.; Agarwal, S. Rock modeling and matching for autonomous long-range Mars rover localization. *J. Field Robot.* **2007**, *24*, 187–203. [CrossRef]
41. Yang, J.; Kang, Z. A gradient-region constrained level set method for autonomous rock detection from Mars rover image. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch.* **2019**, *42*, 1479–1485. [CrossRef]
42. Zhou, R.; Feng, W.; Yang, H.; Gao, H.; Li, N.; Deng, Z.; Ding, L. Predicting terrain mechanical properties in sight for planetary rovers with semantic clues. *arXiv* **2020**, arXiv:2011.01872.
43. Furlán, F.; Rubio, E.; Sossa, H.; Ponce, V. Rock detection in a Mars-like environment using a CNN. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11524, pp. 149–158. [CrossRef]
44. Chiodini, S.; Torresin, L.; Pertile, M.; Debei, S. Evaluation of 3D CNN semantic mapping for rover navigation. In Proceedings of the 2020 IEEE International Workshop on Metrology for AeroSpace, Pisa, Italy, 22 June–5 July 2020; pp. 32–36. [CrossRef]
45. Pessia, R. Artificial Lunar Landscape Dataset. Available online: https://www.kaggle.com/romainpessia/artificial-lunar-rocky-landscape-dataset (accessed on 22 June 2021).
46. Bonechi, S.; Bianchini, M.; Scarselli, F.; Andreini, P. Weak supervision for generating pixel–level annotations in scene text segmentation. *Pattern Recognit. Lett.* **2020**, *138*, 1–7. [CrossRef]
47. Nalepa, J.; Myller, M.; Kawulok, M. Transfer learning for segmenting dimensionally reduced hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1228–1232. [CrossRef]
48. Li, J.; Zhang, L.; Wu, Z.; Ling, Z.; Cao, X.; Guo, K.; Yan, F. Autonomous Martian rock image classification based on transfer deep learning methods. *Earth Sci. Inform.* **2020**, *13*, 951–963. [CrossRef]
49. Hewitt, R.A.; Boukas, E.; Azkarate, M.; Pagnamenta, M.; Marshall, J.A.; Gasteratos, A.; Visentin, G. The Katwijk beach planetary rover dataset. *Int. J. Robot. Res.* **2018**, *37*, 3–12. [CrossRef]
50. Sánchez-Ibánez, J.R.; Pérez-del-Pulgar, C.J.; Azkarate, M.; Gerdes, L.; García-Cerezo, A. Dynamic path planning for reconfigurable rovers using a multi-layered grid. *Eng. Appl. Artif. Intell.* **2019**, *86*, 32–42. [CrossRef]
51. Gerdes, L.; Azkarate, M.; Sánchez-Ibáñez, J.R.; Joudrier, L.; Perez-del-Pulgar, C.J. Efficient autonomous navigation for planetary rovers with limited resources. *J. Field Robot.* **2020**, *37*, 1153–1170. [CrossRef]
52. Furlán, F.; Rubio, E.; Sossa, H.; Ponce, V. CNN based detectors on planetary environments: A performance evaluation. *Front. Neurorobot.* **2020**, *14*, 1–9. [CrossRef]
53. Meyer, L.; Smíšek, M.; Fontan Villacampa, A.; Oliva Maza, L.; Medina, D.; Schuster, M.J.; Steidle, F.; Vayugundla, M.; Müller, M.G.; Rebele, B.; et al. The MADMAX data set for visual-inertial rover navigation on Mars. *J. Field Robot.* **2021**, *38*, 833–853. [CrossRef]
54. Lamarre, O.; Limoyo, O.; Marić, F.; Kelly, J. The Canadian planetary emulation terrain energy-aware rover navigation dataset. *Int. J. Robot. Res.* **2020**, *39*, 641–650. [CrossRef]
55. NASA. NASA Science Mars Exploration Program. Available online: https://mars.nasa.gov/mars2020/multimedia/raw-images/ (accessed on 29 May 2021).
56. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
57. Kuang, B.; Rana, Z.A.; Zhao, Y. Sky and ground segmentation in the navigation visions of the planetary rovers. *Sensors* **2021**, *21*, 6996. [CrossRef] [PubMed]
58. Lin, M.; Chen, Q.; Yan, S. Network in Network. In Proceedings of the 2nd International Conference on Learning Representations ICLR 2014, Banff, AB, Canada, 14–16 April 2014; pp. 1–10.
59. Gurita, A.; Mocanu, I.G. Image segmentation using encoder-decoder with deformable convolutions. *Sensors* **2021**, *21*, 1570. [CrossRef]
60. Marcinkiewicz, M.; Nalepa, J.; Lorenzo, P.R.; Dudzik, W.; Mrukwa, G. Segmenting brain tumors from MRI using cascaded multi-modal U-nets. In *International MICCAI Brainleison Workshop*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 13–24.
61. Tarasiewicz, T.; Nalepa, J.; Kawulok, M. Skinny: A lightweight U-net for skin detection and segmentation. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 2386–2390.
62. Zhixuhao. Unet. Available online: https://github.com/zhixuhao/unet (accessed on 23 July 2021).
63. Mulesial. Pytorch-UNet. Available online: https://github.com/milesial/Pytorch-UNet (accessed on 23 July 2021).
64. 4uiiurz1. Pytorch-Nested-Unet. Available online: https://github.com/4uiiurz1/pytorch-nested-unet (accessed on 26 November 2021).
65. Lin, C.H.; Kong, C.; Lucey, S. Learning efficient point cloud generation for dense 3D object reconstruction. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 7114–7121.
66. Zuo, J.; Xu, G.; Fu, K.; Sun, X.; Sun, H. Aircraft type recognition based on segmentation with deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 282–286. [CrossRef]