AI-BASED PRE-PROCESSING FOR NAVIGATION ON SIGNIFICANTLY UNSTRUCTURED PLANETARY SURFACES

P. Suwinski*, A. Liesch*, B. Liu*, V. Chernykh*, K. Janschek*
* Institute of Automation, Faculty of Electrical and Computer Engineering, TUD Dresden University of Technology, 01187 Dresden, Germany

Abstract

The safe exploration of Small Celestial Bodies (SCBs) requires continuous human supervision and participation, which limits the number of missions that can be realized. If spacecraft were endowed with greater onboard autonomy, exploration could be carried out more efficiently, resulting in broader coverage and increased scientific output. By reducing the need for extensive human involvement, these missions could be conducted more frequently and with greater ease. This paper presents an AI-based, especially Convolutional Neural Network (CNN)-based, pre-processing for improving the navigation concept of the previously finished Astrone project, which introduces an exploration concept that suggests using autonomous hovering relocation maneuvers for exploration spacecraft on Small Solar System Bodies (SSSBs), such as comets and asteroids to enhance their mobility. Our improvements support autonomous navigation concepts by extracting relevant features from the spacecraft's irregular and unstructured environment, such as the comet 67P/Churyumov-Gerasimenko's surface, captured by a 65°×65° wide-angle 2D camera and a corresponding Light imaging, Detection And Ranging (LiDAR) sensor. This paper focuses on predicting Volume Center Points (VCPs) of different-sized stones in the spacecraft's camera frame as landmarks by CNNs. These VCPs are treated as key-points for local navigation, such as point cloud registration and place recognition. A simplified dataset for our specific scenario was generated to train and test different CNNs, such as the U-Net, DEFU-Net, U^2-Net, and the Deeplabv3+ with certain ResNet and ResNet-RS backbone net architectures. The CNNs were trained and tested for semantic segmentation tasks and VCP detection for different-sized stones. Further, the Deeplabv3+ with a ResNet-RS50 backbone net was extended by three separate Multi Perceptron Layers (MLPs) to predict the VCP coordinates independently. A Mean Absolute Error (MAE) within $0.837 \cdot \sigma_{radius}$ could achieved, where σ_{radius} is the stones' radius' standard deviation.

Keywords
Artificial Intelligence; Guidance, Navigation and Control; Space Exploration

NOMENCLATURE

Abbrevations

AI	Artificial Intelligence
ALLD	Artificial Lunar Landscape Dataset
ASPP	Atrous Spatial Pyramid Pooling
CNN	Convolutional Neural Network
DEM	Digital Elevation Map
GAN	Generative Adverisal Network
HITL	Human In The Loop
LiDAR	Light imaging, Detection And Ranging
mIOU	Mean Intersection Over Union
MLP	Multi-Layer Perceptron
ReLU	Rectified Linear Unit
SCB	Small Celestial Bodies
SELU	Scaled Exponential Linear Unit
SSSB	Small Solar System Body
VCP	Volume Center Point

1. INTRODUCTION

A growing scientific interest has recently evolved in exploring Small Celestial Bodies (SCBs) within our solar system, such as asteroids and comets. These objects achieved significant attention due to their intrinsic value and potential for resource utilization. Investigating these celestial bodies can provide profound insights into the origins and evolution of our solar system and enable the identification of valuable resources crucial for future space missions. Exploration missions to SCBs pose numerous challenges, considering the emerging trends in precise and autonomous mobility. Exploring Small Solar System Bodies (SSSBs) revolves around achieving autonomy in mission operations. Autonomous operation is crucial, particularly for long distances where real-time communication between the Earth and the spacecraft on the SSSB is impossible because of the high data transfer latency. Autonomy refers to the spacecraft's capability to perform tasks and make decisions without continuous human intervention. Previous SSSB exploration missions have emphasized developing autonomous capabilities, enabling spacecraft to navigate, land, and conduct scientific investigations without Human In The Loop (HITL) approaches.

These autonomous capabilities are achieved using optical relative navigation and radiometric absolute navigation techniques [1–4]. Radiometric tracking allows for precise and absolute tracking of the spacecraft's trajectory relative to the SSSB's global reference frame. Landmark-based

doi: 10.25967/610427

navigation relies on identifiable surface features of the SSSB to assist in spacecraft localization and orientation. While these technologies have proven successful, they still require the HITL. Landmark-based navigation necessitates extensive modeling efforts on the ground, while radiometric measurements may only sometimes be available. Additionally, the selection of a landing site requires thorough surface mapping. In contrast, a fully autonomous spacecraft would have the capability to land independently on the surface and subsequently explore the entire surface autonomously. By expanding the range of exploration in this manner, the overall scientific revenue of the mission would experience a significant increase. To support navigation tasks, in this contribution, an Artificial Intelligence (AI)-based pre-processing, primarily through Convolutional Neural Networks (CNNs), is introduced for improving the navigation concept of the previously finished Astrone [5] project in the context of the current Astrone KI [6] project. The CNN tasks split into semantic segmentation of different-sized stones on the SSSB's surface and detecting stone Volume Center Points (VCPs) as key-points for further navigation sub-tasks, e.g., point cloud registration and place recognition. In summary, the main contributions of this paper include the following:

- An overview about the generated dataset for the training and validation of CNNs in the context of the Astrone KI project,
- CNN architecture, training setup and validation results for semantic segmentation of different-sized stones, and
- CNN architecture, training setup and validation results for prediction of VCPs of different-sized stones.

2. RECENT WORK

Recent space missions have progressively integrated technologies to expand the boundaries of autonomy. For example, the Hayabusa-2 mission [7], launched in 2014, introduced upgraded navigation instruments and employed advanced characterization techniques such as radiometric tracking and autonomous descent. Similarly, the OSIRIS-REx mission [8], launched in 2016, utilized vision-based navigation for close-range operations and radiometric tracking while incorporating advanced exposure techniques and landmark tracking. The DART mission [9], launched in 2021, achieved kinetic impact deflection by utilizing fully autonomous navigation systems and avionics. Advancing the frontiers of autonomy in upcoming missions is of utmost importance as it approaches the limits of human intervention or ground-based control in SSSB missions. Autonomy becomes crucial to ensure optimal performance, especially when operating at considerable distances where communication delays pose significant challenges. As missions become more intricate, involving multiple probes or exploring multiple targets, the importance of autonomy is further underscored. Envisioned scenarios include multiple spacecraft independently landing and exploring asteroid fields with minimal human intervention. Replacing HITL approaches requires sophisticated AI-based machine-learning algorithms, which enable spacecraft to acquire real-time learning and adaptation capabilities to enhance their predictions. Improved sensor capabilities also play a crucial role, allowing spacecraft to acquire precise and reliable data for autonomous operations. This encompasses advancements in optical and Light imaging, Detection And Ranging (LiDAR) sensors and the integration of innovative sensing technologies. Furthermore, in the context of SSSB missions, the ability to make robust decisions and exhibit adaptive autonomy is of paramount importance. These aspects entail the development of algorithms and strategies capable of handling unexpected events, adapting to changing environments, and making intelligent decisions in complex and uncertain scenarios. By incorporating robustness and adaptability into the autonomy framework, spacecraft can effectively address the intricacies of missions and ensure successful operations.

Information about the planetary environment on SSSBs from imaging sensors such as mono cameras and LiDAR sensors is essential for the perception and navigation of mobile exploration robots on planetary surfaces. Basic navigation tasks include detecting and sensing objects and terrain features under adverse conditions in significantly unstructured environments. In this context, semantic classification of the amorphous and irregular terrain surface plays a crucial role. The recently finished Astrone project presented by Martin et al. (2021) presents an exploration concept that suggests using autonomous hovering relocation maneuvers for exploration spacecraft on SSSBs, such as comets and asteroids, to enhance mobility [5]. According to this previous Astrone project's exploration concept, image and LiDAR sensor data are acquired from an aerial position [6]. Only high-resolution 1024×1024 grayscale 2D image data will be available, fused with low-resolution 128×128 LiDAR 3D point cloud data from a flash-LiDAR sensor to produce highly accurate 3D terrain maps can be used for navigation and path planning on SSSBs [10]. In addition, an autonomous flash-LiDAR-based inertial navigation system for a spacecraft near the SSSB surface has been successfully developed and tested by Liu et al. (2021) [11].

3. CONCEPT OF THE AI PRE-PROCESSING

Our enhanced Astrone navigation concept, Astrone KI, includes additional AI algorithms. The previous Astrone and the Astrone KI navigation concepts are shown in Fig. 1 and 2. This paper focuses on the AI pre-processing block. The AI-Mapping system was introduced by Liesch et al. (2023) [10]. The estimation of positions, velocities, and attitudes of a hovering spacecraft on SSSBs includes considerable uncertainties. To reduce estimation uncertainties, in this contribution, an AI pre-processing for observations of the relative position and attitude of the spacecraft is introduced. The Astrone KI navigation concept focuses on the collaborative use of AI-based feature extraction and a discriminator with analytical non-AI (conventional) algorithms as part of local navigation on the SSSB's surface.

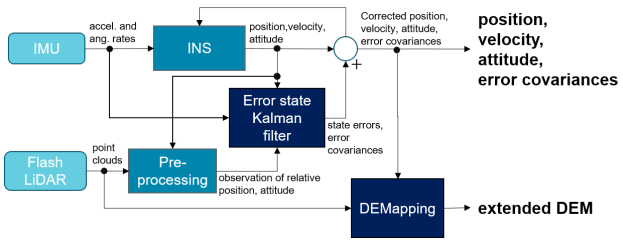


FIG 1. Navigation concept of the Astrone project [5, 6].

The previous Astrone navigation concept applied analytical models to process point clouds from the SSSB's

unstructured surface for local navigation tasks. In the Astrone KI navigation concept, these analytical models were replaced through AI-based approaches for complex regression tasks in 2D images and corresponding 3D distance data (depth images), especially classifications.

The feature extraction contains a CNN that processes early fused coherent high-resolution grayscale 2D image and low-resolution flash-LiDAR data in the camera frame as input data. To early-fuse these data, the patch-based Generative Adverisal Network (GAN) approach from Liesch et al. (2023) [12] is applied for rescaling the low-resolution depth image to the resolution of the corresponding grayscale 2D image. To evaluate the VCP prediction approach, different CNNs are evaluated using a CNN input image resolution of 256×256 instead of 1024×1024 to reduce computational cost. Suitable CNNs are adapted for semantic segmentation tasks in our specific irregular and unstructured environment. Certain CNN variations were trained to extract VCPs of small stones, big stones, and rocks for further use as key-points in discriminator and matching algorithms. Due to our approach lacking satisfying datasets for aerospace for supervised training of CNNs, a simplified simulated synthetic dataset through the approach of Schmipf et al. (2022) [13] was developed. The dataset is inspired by the Rosetta orbiter mission's generated images of the 67P/Churyumov-Gerasimenko comet's surface [14].

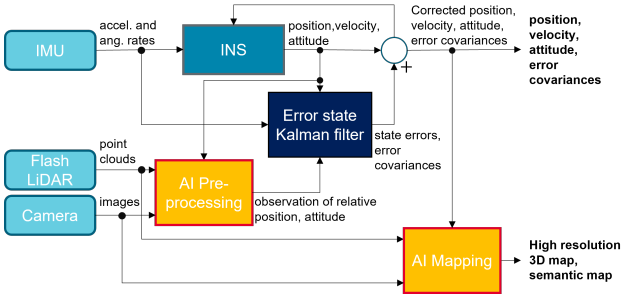


FIG 2. Navigation concept including AI-algorithms of the Astrone KI project.

3.1. Dataset

To generate grayscale 2D data and corresponding 3D distance data provided by a simulated LiDAR sensor, the approach presented by Schimpf et al. (2022) [13] was improved. Schimpf et al. (2022) introduced the generation of a Digital Elevation Map (DEM) by overlying layers, where each layer contains different environmental information, such as rocks, hills, craters, slopes, and surface texture, as shown in Fig. 3.

While this approach focuses on generating a DEM and local navigation in a DEM perspective, an additional rendering for navigation in the camera perspective by the software Blender [15] was applied to render grayscale 2D and corresponding depth images in the spacecraft's camera frame. The rendered grayscale 2D image data include shadows and a Lambertian scattering reflectance model. Different shadows on the DEM's surface were applied through a moving sun at $\pm 45°$ in $15°$ steps in the simulated spacecraft's yaw-axis, where $0°$ is in the camera's back. For semantic segmentation, ground truth 2D images for supervised learning-based training of the CNNs were generated. Fig. 4a - 4f presents a grayscale 2D image and its corresponding data. The VCPs of stones highlighted in blue and red in Fig. 4f were generated through a coordinate transformation of the VCPs in the DEM's global coordinate frame to the spacecraft's camera frame. In the Astrone KI project, a high-fidelity 2D and 3D data generator software CamSim [16] developed by Astos Solutions is applied. CamSim simulates a high-resolution grayscale 2D camera and a flash-LiDAR sensor to render data from an SSSB's surface, inspired by the 67P/Churyumov-Gerasimenko comet's surface. Fig. 4h and 4g presents 2D image and depth image output data rendered by CamSim.

Further, the focus is on comparing CNN performances on our generated dataset and the Artificial Lunar Landscape Dataset (ALLD) for semantic segmentation tasks on a simulated lunar surface with the same classes as in our generated dataset. Additionally, the ALLD is cleaned from artifacts for usage in a single-class multi-label scenario. The ALLD provides similar environment constraints as our dataset, except for the significantly unstructured SCB surface, applied light scattering models, and missing depth images to their corresponding grayscale 2D images.

3.2. Semantic Segmentation

Different CNNs from comparable semantic segmentation approaches of obstacles and objects for object detection tasks related to our navigation concept were applied. Based on the widely used encoder-decoder structure of the U-Net [17] from Ronneberger et al. (2015), Zhou et al. (2018) developed the U-Net++ [18] by adding nested and dense skip connections to the original U-Net. By redesigning these skip connections to full-scale skip connections, Huang et al. (2020) reduced the complexity of the U-Net++ in their developed U-Net3+ [19]. They achieved a potentially more accurate position-aware and boundary-enhanced segmentation map as output. Qin et al. (2020) introduced a highly complex CNN structure with nested U-Nets in a typical U-Net structure [20]. This architecture uses the advantages of a U-Net with improvements of additional implemented residual U-Net blocks to combine multi-scale and local features. Compared to the original U-Net, this results in potentially high-resolution semantic segmentation results without a significantly increased memory and computation afford. Zhang et al. (2021) realized another approach to improve the U-Net by extracting additional features in their developed Dual Encoder Fusion U-Net (DEFU-Net) [21]. This approach results in a highly increased resolution of semantic segmentation results, but a high computation afford.

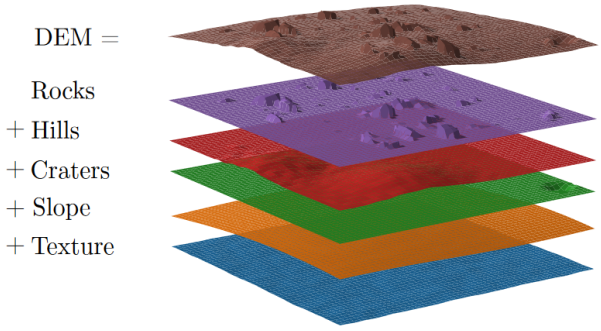
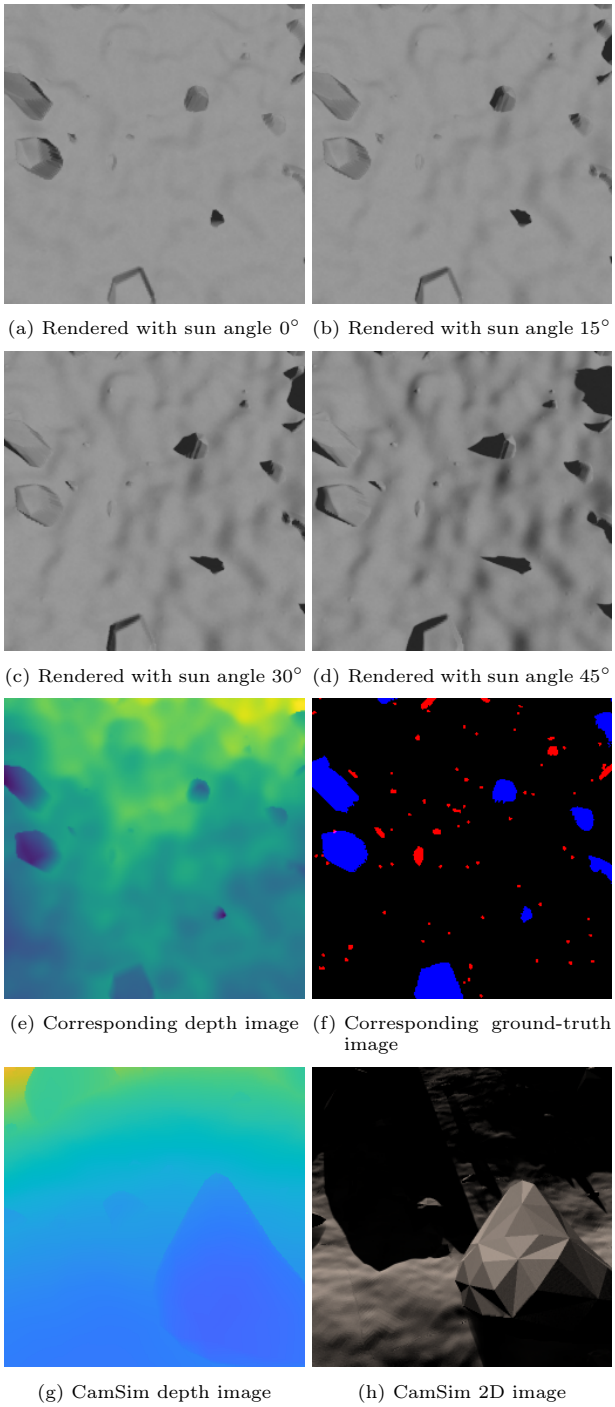


FIG 3. DEM generator by Schimpf et al. (2022) [13].

Another approach to improve the U-Net by adding attention gates, the Attention U-Net (AttU-Net) [22] by Oktay et al. (2018), and its combination with the Residual Recurrent U-Net (R2U-Net) [23] by Alom et al. (2018), the R2AttU-Net, promises potentially good results for our semantic segmentation task. Further, Rundo et al. (2019) added Squeeze-and-Excitation (SE) blocks after each encoder layer of the U-Net and developed the resulting SEU-Net [24]. Roy et al. (2018) added concurrent spatial SE blocks on the decoder site to the SEU-Net and developed the scSEU-Net [25].



(a) Rendered with sun angle 0° (b) Rendered with sun angle 15°

(c) Rendered with sun angle 30° (d) Rendered with sun angle 45°

(e) Corresponding depth image (f) Corresponding ground-truth image

(g) CamSim depth image (h) CamSim 2D image

FIG 4. Rendered grayscale 2D images with different lightning conditions, corresponding qualitative depth image, corresponding ground-truth image, and CamSim 2D and qualitative depth image output data.

The Deeplabv3+ [26] utilizes a CNN as a backbone net to extract hierarchical features in different scales to capture global and local context information from the pixel of the Deeplabv3+'s input data. An Atrous Spatial Pyramid Pooling (ASPP) enables the Deeplabv3+ to perceive fine-graded and larger scales to enhance the segmentation accuracy. The ASPP is based on capturing context information at varying receptive field sizes through atrous dilated convolutions with multiple rates. As the backbone net for feature extraction, the focus is on ResNet [27] and ResNet-RS-based [28] architectures, which include identity mappings based on residual connections. ResNet-RS architectures extend vanilla ResNets by adding Random Skip connections and residual connections to aggregate current feature maps with a randomly selected subset of previous feature maps. The random skip connections enhance CNN's learning capabilities and the learned representations' diversity.

For semantic segmentation tasks in 2D images and corresponding depth images of the unstructured SSSB's surface in our generated dataset, ResNet50, ResNet50-RS, ResNet101, ResNet101-RS, ResNet152, and ResNet152-RS as backbone nets for the Deeplabv3+ were focused. The semantic segmentation's pixel-level classification follows a single-label multi-class classification approach, where only one class can be assigned to one pixel of the CNN's predicted output 2D image, as formulated in Eq. 1.

$$\mathbf{1}_{\mathbf{A}}(\mathbf{x}) = \begin{cases} 1, & if\ \mathbf{x} \in \mathbf{A} \\ 0, & if\ \mathbf{x} \notin \mathbf{A} \end{cases} \tag{1}$$

The outputs of all chosen CNNs were fitted to a three-class classifier. The CNNs and their input vector $\mathbf{x}$ treats their output vector $\mathbf{y}$ as a probability distribution based on a softmax function. Following the notation of Papernot et al. (2016), $\mathbf{F}_1, ..., \mathbf{F}_n$ refers to CNN layers [29]. $\mathbf{F}(\mathbf{x})$ represents the CNN, including the softmax function, with $\mathbf{x}$ as the CNN's input vector. The output of the CNN without the softmax function is defined as $\mathbf{Z}(\mathbf{x})$ and represents the logits. The output vector $\mathbf{y}$ is defined by Eq. 2 with the input vector $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$ and the output vector $\mathbf{y} \in \mathbb{R}^{H \times W \times 3}$, where H is height, W is width, and C is the input image data channel.

$$\begin{aligned} \mathbf{F} &= softmax \circ \mathbf{F}_n \circ \mathbf{F}_{n-1} \circ ... \circ \mathbf{F}_1 \\ \mathbf{y} &\coloneqq \mathbf{F}(\mathbf{x}) = softmax(\mathbf{Z}(\mathbf{x})) \end{aligned} \tag{2}$$

3.3. Volume Center Point prediction

For predictions of VCPs of big and small stones, the structure of a CNN were adapted by changing its output layer architecture while implicitly using the encoder-decoder structure for semantic segmentation tasks. Instead of a pixelwise probability distribution that is focused on semantic segmentation tasks, the output of the CNN $\mathbf{y}_{VCP}$ is a list of VCPs of different-sized stones in the spacecraft's camera frame $\mathbf{y}_{VCP} \in \mathbb{R}^{k \times VCP}$ where k is the number of VCPs in the current rendered scene's camera frame where each $VCP \in (x, y, z)^T$. The coordinates in the camera frame were defined as a cartesian left-handed system with the origin in the camera's focal point. Following this definition, x and y can be positive and negative, while z is always positive [30].

The CNN output layer's softmax function as described in Eq. 2 of the CNN for semantic segmentation tasks is replaced through a Multi-Layer Perceptron (MLP) architecture to generate a list of VCPs by regression. The MLP applies, in multiple instances, a regression algorithm on extracted stone surface contours to predict the stones' VCPs:

1) Normalize the CNN's output logits $\mathbf{Z}(\mathbf{x})$ in their data channel to reduce computation effort [31] and reshape the resulting output through flattening,
2) Apply three independent three-layer depth MLPs to process the regressions for each VCP independently in x, y, and z for each direction, and
3) Stack the separately predicted cartesian x, y, and z coordinates to achieve the expected dimension of single VCPs where each $VCP \in (x, y, z)^T$.

The CNN output $\mathbf{y}_{VCP}$ for predicting VCPs of different-sized stones is described in Eq. 3, where $\mathbf{Z}(\mathbf{x})$ are the CNN's ($\mathbf{F}_{VCP}$) logits, MLP_x, MLP_y, and MLP_z the applied MLPs in the corresponding x, y, and z direction, and σ is the applied MLP's output's activation function.

$$\mathbf{y}_{VCP} \coloneqq \mathbf{F}_{VCP}(\mathbf{x}) = \begin{pmatrix} \sigma_{SELU}(MLP_x(\mathbf{Z}(\mathbf{x}))) \\ \sigma_{SELU}(MLP_y(\mathbf{Z}(\mathbf{x}))) \\ \sigma_{ReLU}(MLP_z(\mathbf{Z}(\mathbf{x}))) \end{pmatrix} \tag{3}$$

A ReLU activation function was chosen to predict the VCPs' position in the z-direction because their values can only be positive where $z \in \mathbb{R}^+$. For predicting VCPs' positions in the x- and y direction, a Scaled Exponential Linear Unit (SELU) activation function as a modification of the ReLU was applied, defined by Eq. 4a and 4b with $\lambda \approx 1.05$ and $\alpha \approx 1.67$ as derived scalar quantities [32].

$$\sigma_{ReLU}(x) = \begin{cases} x, & if\ x > 0 \\ 0, & if\ x \leq 0 \end{cases} \tag{4a}$$

$$\sigma_{SELU}(x) = \begin{cases} \lambda x, & if\ x > 0 \\ \lambda\alpha(e^x - 1), & if\ x \leq 0 \end{cases} \tag{4b}$$

4. TEST RESULTS

In the specific scenario, a three-level test campaign focuses on getting a satisfying CNN structure for semantic segmentation tasks and VCP predictions. The CNN architectures are compared by the metrics defined in Eq. 5a - 5d, and the final VCP prediction is validated through the Mean Absolute Error (MAE) metric defined in Eq. 5e.

$$precision = \frac{100}{n} \cdot \sum_{k=1}^{n} \frac{TP_k}{TP_k + FP_k} \tag{5a}$$

$$recall = \frac{100}{n} \cdot \sum_{k=1}^{n} \frac{TP_k}{TP_k + FN_k} \tag{5b}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5c}$$

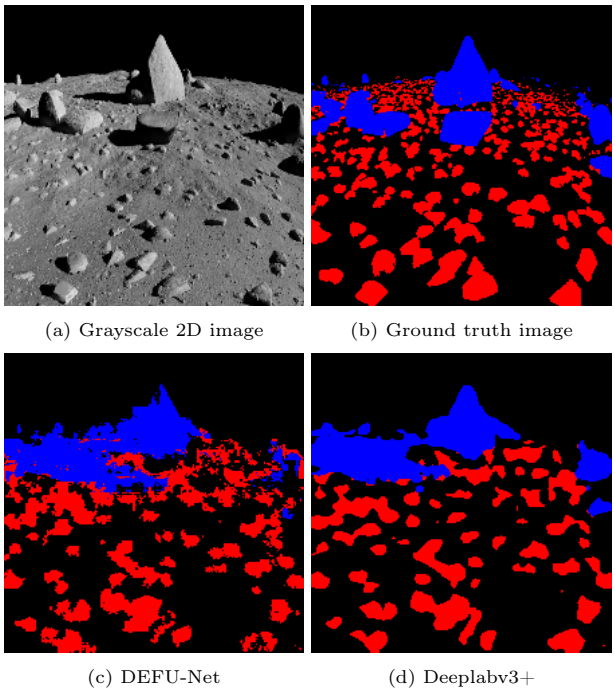
$$mIOU = \frac{100}{n} \cdot \sum_{k=1}^{n} \frac{|\mathbf{y}_{k,true} \cap \mathbf{y}_{k,pred}|}{|\mathbf{y}_{k,true} \cup \mathbf{y}_{k,pred}|} \tag{5d}$$

$$MAE = \frac{1}{n} \cdot \sum_{k=1}^{n} |\mathbf{y}_{k,true} - \mathbf{y}_{k,pred}| \tag{5e}$$

Because our scenario and datasets differ significantly from those used to develop the CNNs, different CNNs for semantic segmentation tasks were trained and tested on the ALLD to assess their performance on comparable data. The ALLD includes similar training and test data, which is expected for our use case, but with the limitation of missing grayscale 2D images' corresponding depth images, as described in Section 3.1. Tab. 1 presents the test results of different CNNs trained on the ALLD. The Deeplabv3+ with a ResNet50 backbone net and the DEFU-Net show the best results. Respecting the computational effort for inference of the trained CNNs, the Deeplabv3+ outperforms the DEFU-Net in addition to its higher metric values, e.g., mIOU. Fig. 5 presents a grayscale 2D input image, the corresponding ground truth, and the DEFU-Net's and Deeplabv3+'s visualized predictions.

TAB 1. Validation of semantic segmentation on the Artificial Lunar Landscape Dataset (ALLD), with backbone nets in round brackets

CNN	Prec.	Rec.	F1	mIOU
U-Net	91.6	89.1	90.3	44.8
U-Net3+	91.3	87.7	89.4	49.9
U^2-Net	84.6	66.6	74.5	33.9
DEFU-Net	92.4	89.3	90.8	60.7
R2U-Net	86.8	76.4	81.3	40.8
AttU-Net	83.8	75.7	79.5	35
R2AttU-Net	86.7	83	84.9	39.6
SEU-Net	86.9	35.1	50	37.8
scSEU-Net	89.2	39.6	54.9	39
MAP-Net	87.7	53.8	66.7	34.4
Deeplabv3+ (ResNet50)	93.5	92.2	92.9	67.2



(a) Grayscale 2D image (b) Ground truth image

(c) DEFU-Net (d) Deeplabv3+

FIG 5. A grayscale 2D image (a) and corresponding ground truth image (b) of the ALLD, and visualized predictions of the CNNs DEFU-Net (c), and Deeplabv3+ (d) with highlighted classes big stone (blue), small stone (red), and ground (black).

For further training and tests of CNNs on our generated dataset, the Deeplabv3+ with different backbone nets is focused. Due to deep neural network architectures prone

to overfitting with increasing depth, the layer depth of the backbone ResNet and ResNet-RS architectures are limited to 50, 101, and 152. Firstly, the different Deeplabv3+ variants were trained and tested with grayscale 2D image data to test these CNN variants' ability to predict stone contours without distance information between the spacecraft's camera and stones. Additionally, these CNNs were trained and tested with early fused [33, 34] grayscale 2D and depth image input data. For training and testing, the grayscale 2D image lightning conditions were varied in both scenarios from ±45° in 15° steps.

Tab. 2 and 3 present the validation results for both scenarios. The mIOU shows a mean improvement of 4.067, and the best Deeplabv3+ variant with a ResNet50-RS backbone net an increase of the mIOU of 5.24. As expected, the deeper variants of the ResNet and ResNet-RS architectures show less performance on the test data than the 50-layer deep variants. Fig. 6 and 7 present visualized predictions of the tested Deeplabv3+ versions for the input and corresponding ground truth data, presented in Fig. 4. Appendix A Fig. 8 includes an overview of visualized Deeplabv3+'s predictions with a ResNet50-RS backbone net for different illumination conditions. Further, Appendix A Fig. 9 presents the absolute-difference images (misclassifications) of the visualized CNN predictions (Fig. 8) and the expected ground-truth.

Based on the tests of the Deeplabv3+ variants, the Deeplabv3+ with the ResNet50-RS backbone net was trained and tested to predict VCPs of different-sized stones in camera frame with early-fused grayscale 2D and depth images as the CNN's input data. The CNN was trained to predict VCPs of stones with an average radius larger than 10 cm for training. An MAE of 17.65 cm was achieved, where the whole training and test dataset contains stones with a mean radius of $r_{mean} = 30.84\,\text{cm}$, a median of all radii of $r_{median} = 23.6\,\text{cm}$, and a stone radius' standard deviation of $\sigma_{radius} = 21.09\,\text{cm}$. In summary, a VCP prediction within $0.837 \cdot \sigma_{radius}$ was achieved.

A Ryzen 9 5950x CPU and an NVIDIA RTX 3090 GPU were utilized for all training and testing. The ALLD contains 9,766 images, where 8,000 images were used for training and 1,766 for testing. The generated semantic segmentation and VCP prediction dataset was split into 76,500 grayscale 2D and depth images for training and 8,500 for testing. Furthermore, as a loss function, a categorical cross-entropy was used for semantic segmentation and a logarithmic, hyperbolic cosine for VCP prediction. An Adam optimizer with its $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1$ parameters was applied for each setup.

TAB 2. Validation of semantic segmentation on the simplified simulated dataset with 2D image input data, with backbone nets in round brackets

CNN	Prec.	Rec.	F1	mIOU
Deeplabv3+ (ResNet50)	97.2	97.16	97.18	69.05
Deeplabv3+ (ResNet101)	97.05	97.01	97.03	68.20
Deeplabv3+ (ResNet152)	96.97	96.94	96.96	67.84
Deeplabv3+ (ResNet50-RS)	97.48	97.45	97.47	71.58
Deeplabv3+ (ResNet101-RS)	97.36	97.32	97.34	70.85
Deeplabv3+ (ResNet152-RS)	97.35	97.3	97.32	70.35

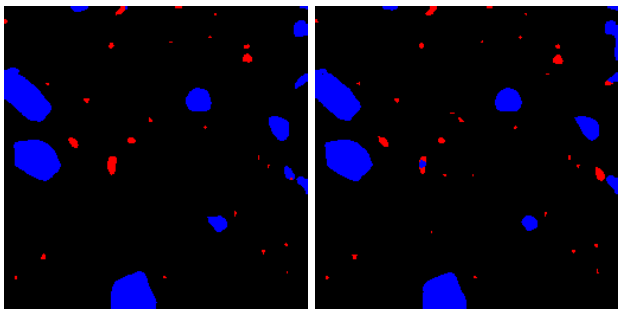
TAB 3. Validation of semantic segmentation on the simplified simulated dataset with 2D image corresponding depth image input data, with backbone nets in round brackets

CNN	Prec.	Rec.	F1	mIOU
Deeplabv3+ (ResNet50)	97.94	97.91	97.92	72.8
Deeplabv3+ (ResNet101)	97.96	97.91	97.93	72.72
Deeplabv3+ (ResNet152)	97.97	97.93	97.95	72.5
Deeplabv3+ (ResNet50-RS)	98.28	98.27	98.27	76.2
Deeplabv3+ (ResNet101-RS)	98.23	98.21	98.22	75.5
Deeplabv3+ (ResNet152-RS)	98.26	98.24	98.25	75.59

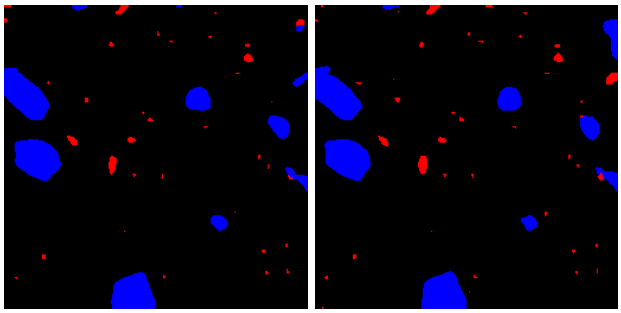
5. CONCLUSION AND OUTLOOK

An AI-based improvement for pre-processing environmental data as part of the improvement of the recently finished Astrone project was introduced. Firstly, the performance of different CNN architectures for semantic segmentation tasks was evaluated on the ALLD, which has similar environmental conditions but includes only 2D images and lacks corresponding depth images. Further, a simplified dataset was generated based on the DEM generation approach presented by Schimpf et al. (2022) [13]. Custom-generated simplified DEMs were rendered with Blender software to generate grayscale 2D images in varying illumination conditions between ±45° in 15° steps. Further, their corresponding depth and ground-truth images for semantic segmentation tasks were also rendered in the spacecraft's camera frame. The CNN Deeplabv3+, including different backbone nets, which has shown the most promising ALLD results, was trained and tested on our generated dataset. To determine the effect of processing additional depth images instead of only grayscale 2D images, different Deeplabv3+ variants were trained for both scenarios. A mIOU mean improvement of 4.067 and a mIOU improvement of 5.24 for the Deeplabv3+ with a ResNet50-RS backbone net was achieved, which has shown the best results in both scenarios.

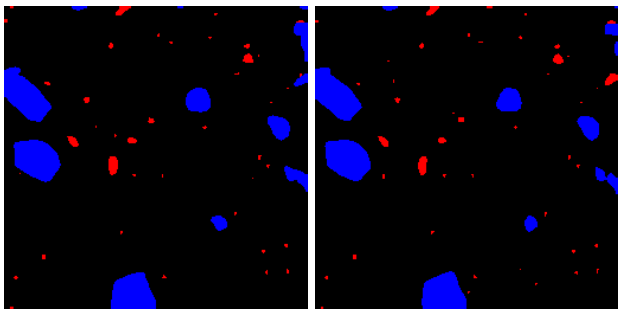
Considering these results, the Deeplabv3+ output layer's softmax function was replaced by an MLP with three independent stages. The stages independently compute the position of VCPs in x, y, and z direction, where x and y can be positive and negative, and z can only be positive, based on the defined left-handed cartesian coordinate system. For the MLP stages, for prediction of the VCP positions in the x and y direction, a SELU, and the z direction, a ReLU activation function was applied. An MAE of 17.65 cm was achieved, where the stones' standard deviation is $\sigma_{radius} = 21.09\,\text{cm}$.



(a) Deeplabv3+ (ResNet50) (b) Deeplabv3+ (ResNet101)

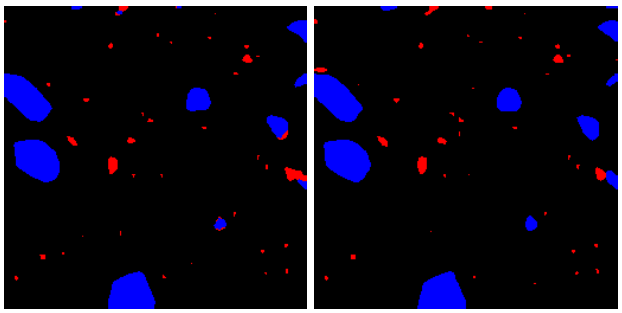


(c) Deeplabv3+ (ResNet152) (d) Deeplabv3+ (ResNet50-RS)*

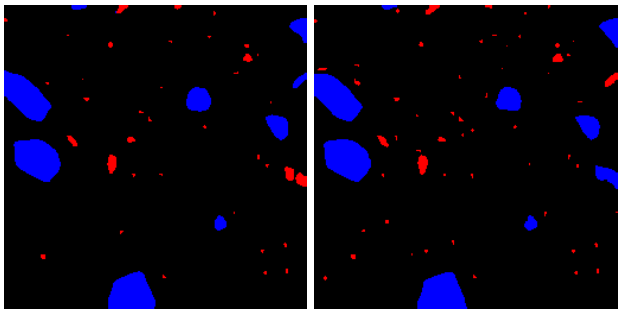


(e) Deeplabv3+ (ResNet101-RS) (f) Deeplabv3+ (ResNet152-RS)

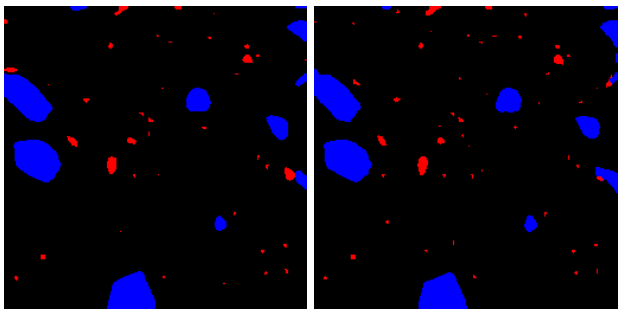
FIG 6. Visualized Deeplabv3+'s predictions with certain backbone nets in round brackets for semantic segmentation of big stone obstacles (blue), small stones (red) and ground (black) with the CNN's input data as shown in Fig. 4a without corresponding depth image data (* best).



(a) Deeplabv3+ (ResNet50) (b) Deeplabv3+ (ResNet101)



(c) Deeplabv3+ (ResNet152) (d) Deeplabv3+ (ResNet50-RS)*



(e) Deeplabv3+ (ResNet101-RS) (f) Deeplabv3+ (ResNet152-RS)

FIG 7. Visualized Deeplabv3+'s predictions with certain backbone nets in round brackets for semantic segmentation of big stone obstacles (blue), small stones (red) and ground (black) with the CNN's input data as shown in Fig. 4a and 4e (* best).

Our achieved VCP prediction with a MAE within $0.837 \cdot \sigma_{radius}$ can be further improved by minimizing the current artificial landscape dataset limitations, e.g., restructuring the landscape and stone generation process to achieve more realistic different-sized stones, applying Perlin noise to the DEM's and stone's surface, and increasing the DEM and render resolutions. Further, the MAE can be reduced by further developments of the CNN structure, e.g., with Vision Transformer as backbone nets.

6. ACKNOWLEDGEMENT

The project "Astrone KI - AI support for high surface mobility of planetary research platforms (agile, autonomous, robust)." has received funding from the German Federal Ministry for Economic Affairs and Energy (BMWi) under funding number "50 RA 2130C" supervised by the German Space Agency (DLR, Deutsches Zentrum für Luft- und Raumfahrt).

Contact address:
patrick.suwinski@tu-dresden.de

References

[1] J Veverka, B Farquhar, M Robinson, P Thomas, S Murchie, A Harch, PG Antreasian, SR Chesley, JK Miller, WM Owen Jr, et al. The landing of the near-shoemaker spacecraft on asteroid 433 eros. Nature, 413(6854):390–393, 2001.

[2] Karl-Heinz Glassmeier, Hermann Boehnhardt, Detlef Koschny, Ekkehard Kührt, and Ingo Richter. The rosetta mission: flying towards the origin of the solar system. Space Science Reviews, 128:1–21, 2007.

[3] CT Russell and CA Raymond. The dawn mission to vesta and ceres. The dawn mission to minor planets 4 Vesta and 1 Ceres, pages 3–23, 2012.

[4] Makoto Yoshikawa, Junichiro Kawaguchi, Akira Fujiwara, and Akira Tsuchiyama. Hayabusa sample return mission. Asteroids IV, 1(397-418):1, 2015.

[5] M Martin, F Belien, B Liu, J Olucak, F Schimpf, T Brüggemann, A Falke, V Sadovski, V Chernykh, K Janschek, et al. Astrone–gnc for enhanced surface mobility on small solar system bodies. In 11 th Inter-

national ESA Conference on Guidance, Navigation & Control Systems, 2021.

[6] M Martin, E Caroselli, J Olucak, S Busi, W Fichter, B Liu, A Liesch, P Suwinski, V Chernykh, and K Janschek. Pioneering the small bodies frontiers: The key enabling technologies for autonomous precise mobilitys. In 12 th International Conference on Guidance, Navigation & Control Systems (GNC), 2023.

[7] Go Ono, Fuyuto Terui, Naoko Ogawa, Shota Kikuchi, Yuya Mimasu, Kent Yoshikawa, Hitoshi Ikeda, Yuto Takei, Seiji Yasuda, Kota Matsushima, et al. GNC strategies and flight results of Hayabusa2 first touchdown operation. Acta Astronautica, 174:131–147, 2020.

[8] DS Lauretta, SS Balram-Knutson, E Beshore, WV Boynton, C Drouet d'Aubigny, DN DellaGiustina, HL Enos, DR Golish, CW Hergenrother, ES Howell, et al. OSIRIS-REx: sample return from asteroid (101955) Bennu. Space Science Reviews, 212:925–984, 2017.

[9] Patrick Michel, Michael Kueppers, Holger Sierks, Ian Carnelli, Andy F Cheng, Karim Mellab, Mikael Granvik, Antti Kestilä, Tomas Kohout, Karri Muinonen, et al. European component of the AIDA mission to a binary asteroid: Characterization and interpretation of the impact of the DART mission. Advances in Space Research, 62(8):2261–2272, 2018.

[10] Alexander Liesch, Patrick Suwinski, Bangshang Liu, Valerij Chernykh, and Klaus Janschek. Ai-based mapping for navigation on significantly unstructured planetary surfaces. Deutscher Luft- und Raumfahrt Kongress, Stuttgart, Germany, Sep 19-21, 2023.

[11] Bangshang Liu, Vasko Sazdovski, Valeriy Chernykh, and Klaus Janschek. Flash lidar aided-inertial navigation on surfaces of small solar system bodies using error state kalman filtering. In AAS/AIAA Astrodynamics Specialist Conference, 08-12 2021. Paper AAS 21-752.

[12] A Liesch, P. Suwinski, B. Liu, V. Chernykh, and K. Janschek. Ai-based lidar / camera data fusion to enable high-resolution 3d surface reconstruction for autonomous asteroid exploration mission. AAS/AIAA Astrodynamics Specialist Conference, Big Sky, MT, Aug 13-17, 2023. Paper AAS-422.

[13] Fabian Schimpf, Jan Olucak, and Walter Fichter. Robust landing site detection for flight over small solar system bodies. In AIAA SCITECH 2022 Forum, page 0955, 2022.

[14] MGGT Taylor, N Altobelli, BJ Buratti, and M Choukroun. The rosetta mission orbiter science overview: the comet phase. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2097):20160262, 2017.

[15] Blender Online Community. Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[16] J Eggert, S Weikert, and I Kossev. Imaging sensor emulation and dynamics simulation for pil/hil. 2023.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[18] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support, pages 3–11. Springer, 2018.

[19] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1055–1059. IEEE, 2020.

[20] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. Pattern recognition, 106:107404, 2020.

[21] Lipei Zhang, Aozhi Liu, Jing Xiao, and Paul Taylor. Dual encoder fusion u-net (defu-net) for cross-manufacturer chest x-ray segmentation. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 9333–9339, 2021. DOI: 10.1109/ICPR48806.2021.9412718.

[22] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018. DOI: 10.48550/ARXIV.1804.03999, https://arxiv.org/abs/1804.03999.

[23] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, 2018. DOI: 10.48550/ARXIV.1802.06955.

[24] Leonardo Rundo, Changhee Han, Yudai Nagano, Jin Zhang, Ryuichiro Hataya, Carmelo Militello, Andrea Tangherloni, Marco S Nobile, Claudio Ferretti, Daniela Besozzi, et al. Use-net: Incorporating squeeze-and-excitation blocks into u-net for prostate zonal segmentation of multi-institutional mri datasets. Neurocomputing, 365:31–43, 2019.

[25] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks. In International conference on medical image computing and computer-assisted intervention, pages 421–429. Springer, 2018. DOI: https://doi.org/10.1007/978-3-030-00928-1_48.

[26] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), pages 801–818, 2018.

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[28] Irwan Bello, William Fedus, Xianzhi Du, Ekin Dogus Cubuk, Aravind Srinivas, Tsung-Yi Lin, Jonathon Shlens, and Barret Zoph. Revisiting resnets: Improved training and scaling strategies. Advances in Neural Information Processing Systems, 34:22614–22627, 2021.

[29] Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, and Ananthram Swami. Distillation as a defense to adversarial perturbations against deep neural networks. In 2016 IEEE symposium on security and privacy (SP), pages 582–597. IEEE, 2016.

[30] Zhengyou Zhang. A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence, 22(11):1330–1334, 2000.

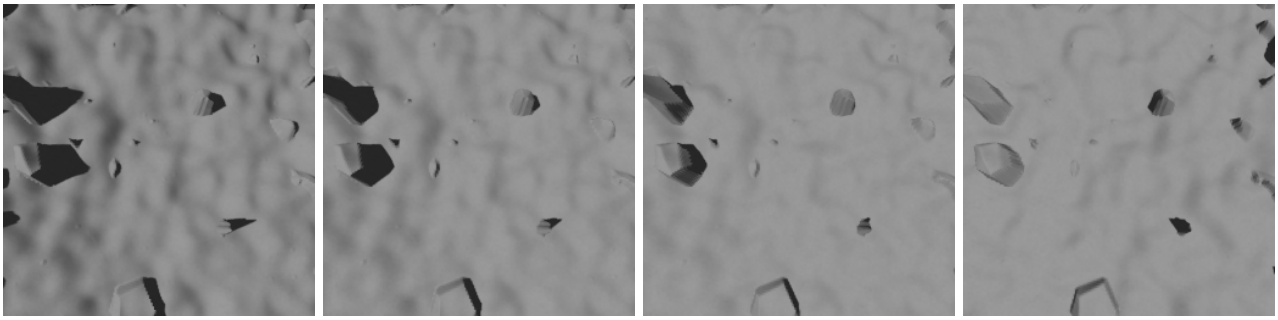
[31] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.

[32] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. CoRR, abs/1706.02515, 2017.

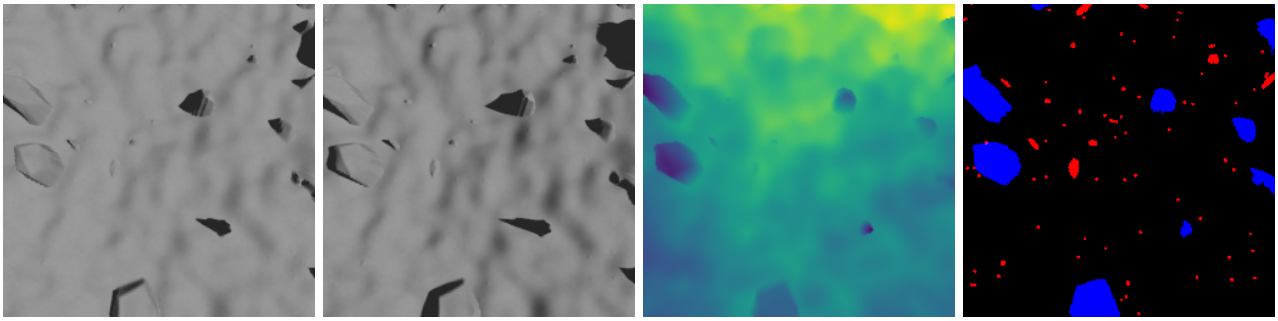
[33] Zexiao Xie, Xiaoxuan Yu, Xiang Gao, Kunqian Li, and Shuhan Shen. Recent advances in conventional and deep learning-based depth completion: A survey. IEEE Transactions on Neural Networks and Learning Systems, pages 1–21, 2022. DOI: 10.1109/TNNLS.2022.3201534.

[34] Junjie Hu, Chenyu Bao, Mete Ozay, Chenyou Fan, Qing Gao, Honghai Liu, and Tin Lun Lam. Deep depth completion from extremely sparse data: A survey, 2022.

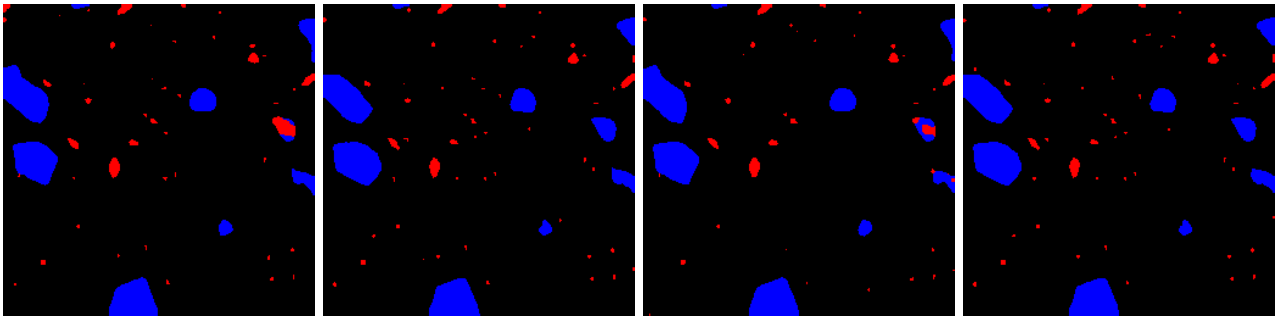
A. APPENDIX



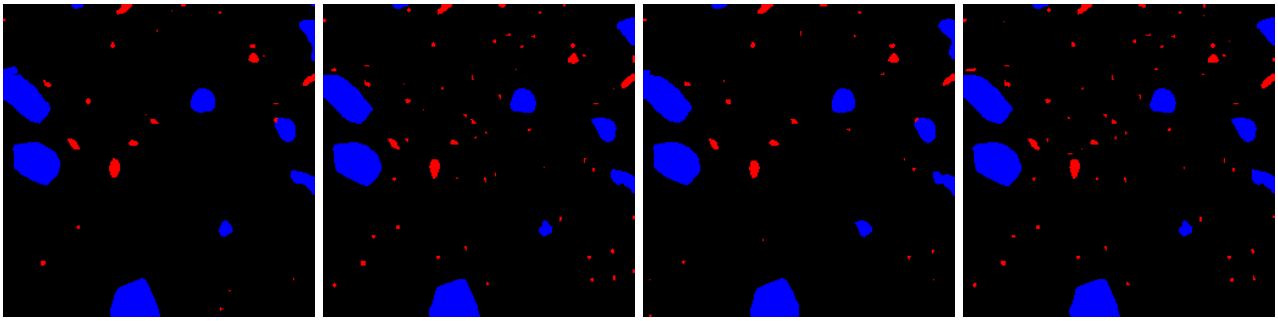
(a) Grayscale 2D image (−45°) (b) Grayscale 2D image (−30°) (c) Grayscale 2D image (−15°) (d) Grayscale 2D image (+15°)



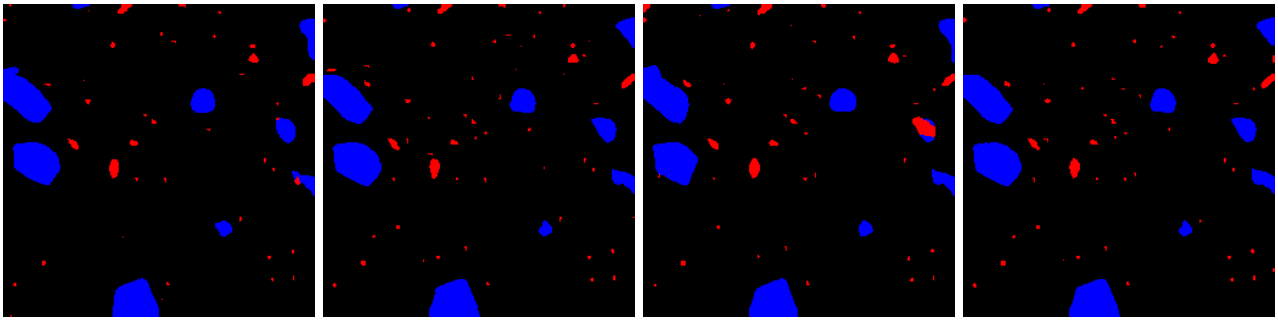
(e) Grayscale 2D image (+30°) (f) Grayscale 2D image (+45°) (g) Qualitative depth image (h) Ground truth image



(i) 2D image only (−45°) (j) 2D and depth image (−45°) (k) 2D image only (−30°) (l) 2D and depth image (−30°)

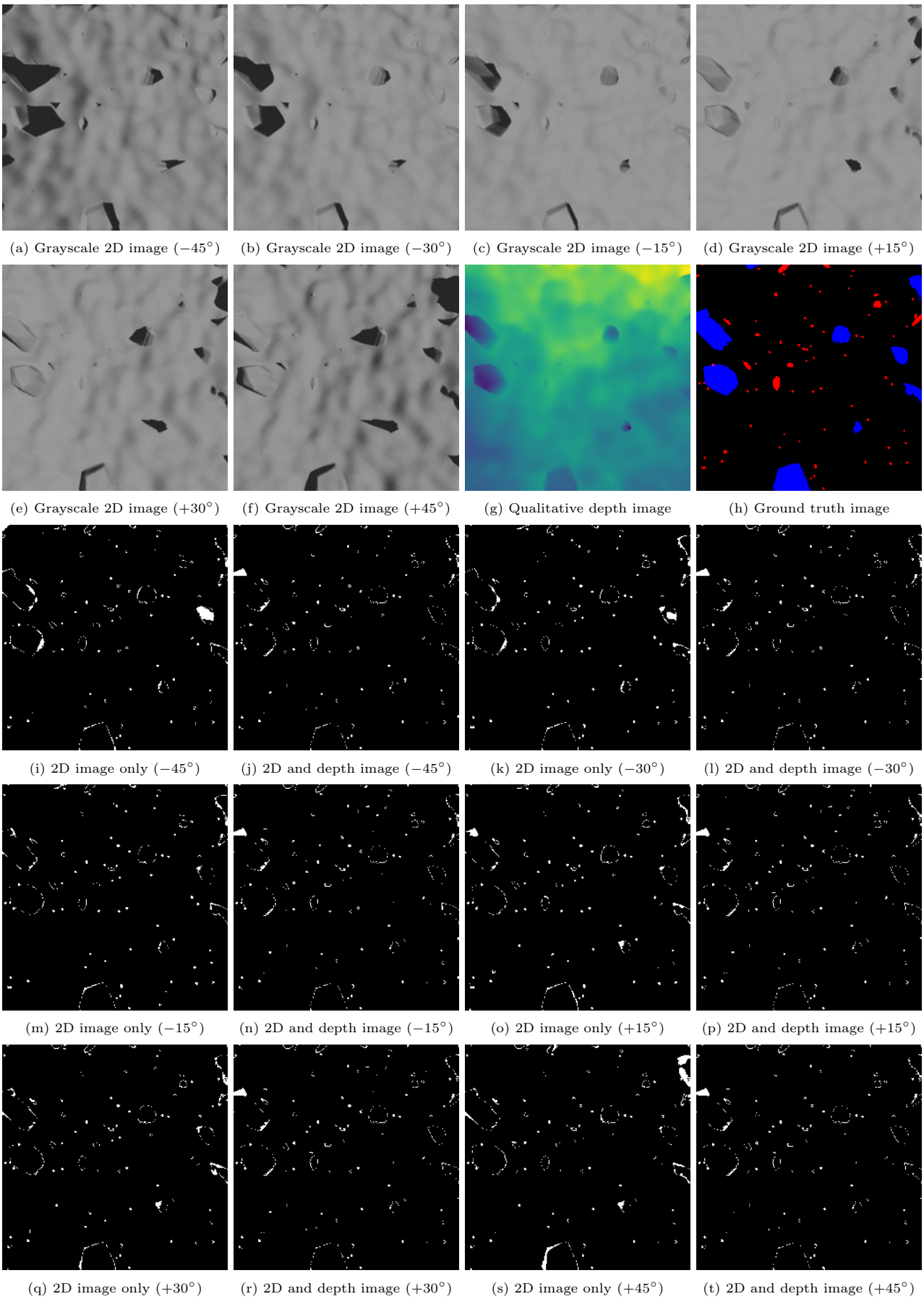


(m) 2D image only (−15°) (n) 2D and depth image (−15°) (o) 2D image only (+15°) (p) 2D and depth image (+15°)



(q) 2D image only (+30°) (r) 2D and depth image (+30°) (s) 2D image only (+45°) (t) 2D and depth image (+45°)

FIG 8. Deeplabv3+'s input data under varying sun azimuth angles in round brackets as shown in (a) - (f), the corresponding depth image (g), and ground-truth image (h), and visualized Deeplabv3+'s predictions with the ResNet50-RS as backbone net (i) - (t).



(a) Grayscale 2D image (−45°) (b) Grayscale 2D image (−30°) (c) Grayscale 2D image (−15°) (d) Grayscale 2D image (+15°)

(e) Grayscale 2D image (+30°) (f) Grayscale 2D image (+45°) (g) Qualitative depth image (h) Ground truth image

(i) 2D image only (−45°) (j) 2D and depth image (−45°) (k) 2D image only (−30°) (l) 2D and depth image (−30°)

(m) 2D image only (−15°) (n) 2D and depth image (−15°) (o) 2D image only (+15°) (p) 2D and depth image (+15°)

(q) 2D image only (+30°) (r) 2D and depth image (+30°) (s) 2D image only (+45°) (t) 2D and depth image (+45°)

FIG 9. Deeplabv3+'s input data under varying sun azimuth angles in round brackets as shown in (a) - (f), the corresponding depth image (g), and ground-truth image (h), and white highlighted misclassifications of the visualized Deeplabv3+'s predictions with the ResNet50-RS as backbone net (i) - (t).