

117 A Supplementary Material

118 A.1 Legacy Architecture: GAN-Based System

119 Before adopting a diffusion-based pipeline, DESPINA was implemented using a GAN architecture—specifically Pix2PixHD trained
 120 on segmented lunar terrain images. While this approach produced
 121 coarse outputs that roughly matched terrain structure, it failed to
 122 preserve fine detail and often introduced artifacts. The GAN system
 123 also required manually crafted label maps, which limited scalability
 124 and generalization. These limitations motivated a transition to
 125 a diffusion-based architecture, which demonstrated superior texture
 126 realism and structural coherence when paired with depth and
 127 soft-edge constraints.
 128

130 A.2 Dataset Preparation

132 To establish a reliable dataset for DESPINA’s architecture, we exclusively utilized Apollo mission imagery, which is publicly available
 133 and readily accessible. The decision to rely solely on Apollo data,
 134 rather than incorporating additional datasets like those from more
 135 recent missions like Chang’e 5, was driven by the need for image
 136 consistency in format and structure. Although other sources, such
 137 as Chang’e 5 and Yutu rover images, as seen in, offered potentially
 138 valuable data, they introduced significant challenges: either
 139 the datasets were fragmented and lacked cohesive organization,
 140 or access was heavily restricted by governing agencies, limiting
 141 their practical use. Specifically, data from Chinese lunar missions
 142 is tightly controlled, and usage typically requires affiliation with
 143 a Chinese academic institution or research organization. Despite
 144 inquiries, permissions to use these datasets could not be secured.
 145 However, small samples of the Chang’e dataset were utilized in the
 146 paper as a validation set, which proved useful.
 147

148 To obtain the Apollo mission imagery, we developed a script to
 149 download all publicly available images from NASA’s online archives.
 150 The script utilized Python’s requests library to automate the re-
 151 trieval of image files. Using a recursive script, images were down-
 152 loaded sequentially from mission subdirectories. Metadata files
 153 accompanying the images were also retrieved where available. The
 154 resulting dataset comprised approximately 10,000 high-resolution
 155 images, which had to be sorted to determine project relevance. See
 156 ?? the incorporation of this sorting model into DESPINA’s system.
 157

158 **A.2.1 Segmentation Process.** We manually segmented 150 images
 159 using CVAT to fine-tune a DINOv2 instance segmentation model.
 160 This model segmented objects in scenes rather than every pixel,
 161 requiring post-processing with inpainting techniques to fill unclas-
 162 sified areas at class boundaries. While this automated approach
 163 saved time and resources, it introduced quality limitations that
 164 could be addressed in future work.

165 Most Apollo mission images include fiducial markers, which
 166 are small crosshair-like artifacts introduced for photogrammetric
 167 purposes. These markers interfere with existing computer vision
 168 techniques by introducing unnatural geometry that disrupts seg-
 169 mentation models. To address this, we developed a preprocessing
 170 pipeline for detecting and removing fiducial markers. The process
 171 involves identifying markers based on their geometric properties
 172 and applying inpainting techniques to remove them while preserv-
 173 ing surrounding image features.

175 **A.2.2 Data Augmentation.** To enhance dataset diversity and miti-
 176 gate overfitting, we applied controlled augmentation techniques:
 177

- 178 • Brightness adjustments ($\pm 20\%$) to simulate varied lunar
 179 lighting conditions
- 180 • Contrast alterations ($\pm 15\%$) to adapt to different surface
 181 reflectance levels
- 182 • Horizontal flips (50% probability) to double training data
 183 for non-directional features

184 Small random rotations were initially implemented but later re-
 185 moved due to overfitting concerns.

186 **A.2.3 BLIP Prompt Generation Example.** We used BLIP [?] to gen-
 187 erate structured text prompts from binary annotations for Apollo
 188 imagery. Below is an example showing the image and correspond-
 189 ing prompt:



200 **Figure 2: Apollo image used in BLIP prompt generation.**

201 **Generated Prompt:** *A realistic image of the Moon’s surface, from the
 202 perspective of a rover. A lunar landscape with hills in the background.
 203 A lunar crater is in view. A rover is present in the image. Dark shadows
 204 cast by the terrain. An astronaut is visible in the image.*

207 A.3 Hardware and Training Configuration

208 The SDXL LoRA and the DPT Fine-Tuning was trained using:

- 209 • Hardware: Six NVIDIA A100 GPUs (40GB memory) with
 210 CUDA 12.1.1
- 211 • Software: Python 3.11, PyTorch 2.1.2 (foss-2023a toolchain),
 212 TorchVision 0.16.0
- 213 • Dataset: 3,500 preprocessed images (2,900 training, 600 vali-
 214 dation)
- 215 • Key parameters: Batch size 6, three discriminators, dropout
 216 enabled

217 A.4 Code Availability

219 The DEM-to-Depth pipeline, processed datasets, and supporting
 220 tools for dataset preparation will be made available at:
 221

222 <https://github.com/UH-DAIS/TIF-to-image>
 223 https://github.com/UH-DAIS/DESPINA_dataset

224
 225
 226
 227
 228
 229
 230
 231

233						
234						
235						
236						
237						
238						
239						
240						
241						
242						
243						

Figure 3: Several Examples of Images Easily Identified as *Not Horizon*. See ??.

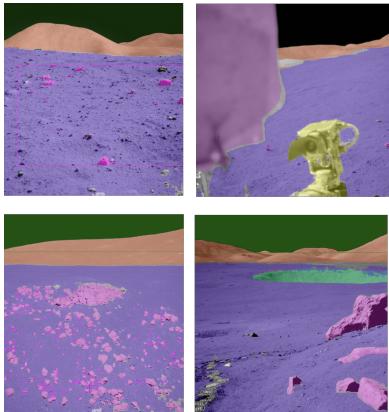


Figure 4: Examples of Manual Segmentation Through CVAT.

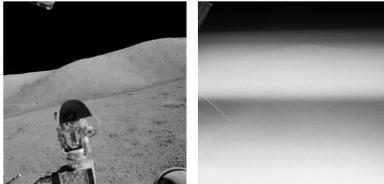


Figure 5: Examples of usable and unusable images from the Apollo dataset.

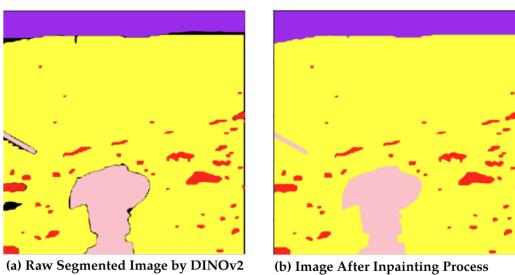


Figure 6: Before and After the Inpainting Process on Raw DINOv2 Generated Segmentation Maps.

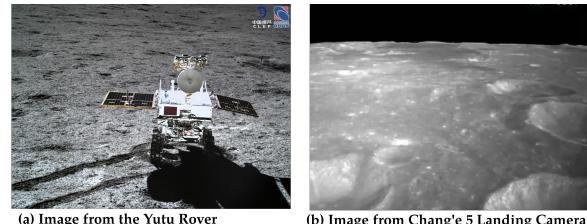


Figure 7: Images of the Lunar Horizon from Non-Apollo Datasets. Sourced from the Yutu Rover (a) and Chang'e 5 (b).

The image consists of four panels arranged in a 2x2 grid. Each panel shows a different perspective of a simulated Mars-like environment. The terrain is primarily a light purple color, representing soil, with numerous small, irregularly shaped pinkish-red objects scattered across it, representing rocks. In the background, there is a brown, rocky horizon line against a dark, possibly black, sky. The top-left panel shows a wide, flat expanse of this terrain. The top-right panel features a yellow-colored robotic arm or rover in the lower right corner, positioned on the purple ground. The bottom-left panel is a close-up shot of the rocky surface, highlighting the texture and distribution of the pinkish-red rocks. The bottom-right panel shows a larger, more prominent rocky outcrop or boulder on the right side of the frame, with some smaller rocks scattered in the foreground.

Figure 4: Examples of Manual Segmentation Through

The image is a composite of two photographs. The left side shows a wide-angle view of the Moon's surface, featuring a prominent crater and a rugged, hilly terrain under a dark sky. The right side is a close-up shot of a scientific instrument, specifically a sensor, mounted on a robotic arm. The sensor has a circular lens and a protective cover. A thin wire or cable extends from the sensor towards the bottom edge of the frame.

Figure 5: Examples of usable and unusable images from Apollo dataset.

Figure 6: Before and After the Inpainting Process on Raw DINOv2 Generated Segmentation Maps.