YOLOLens2.0: A Unified Super-Resolution and Detection Framework for High-Fidelity Crater Mapping in Lunar Permanently Shadowed Regions

Riccardo La Grassa[1], Cristina Re[1], Adriano Tullo[1], Natalia Amanda Vergara Sassarini[1], Gabriele Cremonese [1], Junichi Haruyama[2]

[1] INAF, Italian National Institute for Astrophysics, Italy - riccardo.lagrassa@inaf.it
[2] Institute of Space and Astronautical Science, JAXA, Japan

Keywords: Object detection, Super-Resolution, Craters, Moon, Kaguya, Shadowcam

Abstract

Accurate crater mapping in lunar permanently shadowed regions (PSRs) is hindered by extreme low-light and low-resolution imagery. We present YOLOLens2.0, a unified, end-to-end deep learning framework designed for high-fidelity crater detection and terrain reconstruction in these challenging environments. The architecture integrates a Dense-Residual-Connected Transformer (DRCT) for multimodal super-resolution (SR) with a YOLO-derived detection module and an affine calibrator to ensure geometric consistency at meter scale. Our framework exploits a bidirectional synergy where SR enhances feature discriminability for detection, while detection-driven supervision refines structural reconstruction. Validation on Kaguya data demonstrates a significant performance leap, achieving a Recall of 89.20% and an mAP@50 of 0.844 an improvement of over 33 percentage points in recall compared to the original YOLOLens. Out-of-distribution validation on ShadowCam imagery, performed without fine-tuning, confirms the model’s robustness and scalability. The framework successfully preserves quantitative elevation fidelity and supports detailed morphometric analyses, including the extraction of the crater size-frequency distributions (SFDs) that align with theoretical lunar production functions. YOLOLens2.0 provides a scalable, high-precision methodology for planetary mapping, offering critical insights for lunar surface evolution studies and future exploration missions.

1. Introduction

Planetary surface analysis in permanently shadowed regions (PSRs) is limited by extreme low illumination, where conventional optical imagery fails to resolve fine-scale features. Automatic crater detection and classification on planetary surfaces has progressively transitioned from handcrafted feature engineering to deep learning–based paradigms, with architectural adaptations tailored to the unique geometric and radiometric characteristics of impact craters. Recent contributions span single-stage and two-stage detectors, transformer architectures, and hybrid pipelines incorporating super-resolution (SR) and multimodal fusion.

Early deep learning approaches adapted established object detection frameworks to lunar imagery. Two-stage detectors such as Faster R-CNN have been extensively employed due to their strong localization accuracy. In (Zhong et al., 2025), a two-step pipeline was proposed in which a dedicated SR network enhances DEM inputs prior to crater detection. The optimized Faster R-CNN backbone and region proposal configurations were tuned to exploit the enhanced topographic contrast, yielding improved recall for sub-kilometer craters. The authors report gains in small-crater detection at the cost of additional computational overhead due to the SR preprocessing stage. Similarly, in (Liu et al., 2024), an improved Faster R-CNN framework was adopted, with anchor-scale adjustments and refined proposal strategies adapted to high-resolution Terrain Camera data. The study emphasizes the importance of illumination normalization and anchor tuning for small crater detection in localized landing regions. Single-stage detectors of the YOLO family have gained prominence due to their favorable speed–accuracy balance. The YOLO-SCNet framework, introduced in (Zuo et al., 2025), extends YOLOv11 by incorporating an additional high-resolution detection head for small objects, dynamic anchor optimization tailored to crater size distributions, and multi-scale feature fusion with re-weighted loss components. The authors report consistent improvements in precision, recall, F1-score, and AP relative to YOLOv11, particularly for sub-kilometer craters. Architecturally, the model retains the YOLOv11 backbone while expanding prediction capacity at shallow feature stages to preserve spatial detail. Further YOLO-based refinements are reported in (Gong et al., 2025), where adaptive frequency-domain aggregation modules and enhanced scale-fusion mechanisms are introduced to strengthen small-object representations. The model integrates a custom IoU-based loss for improved regression convergence. The authors report improved small-object mAP at the expense of additional architectural complexity and hyperparameter sensitivity. Beyond convolutional paradigms, transformer-based detectors have emerged. (Guo et al., 2024) introduces a DETR-style architecture augmented with Corresponding Regional Attention and dense intermediate supervision. Multiscale feature fusion is integrated prior to transformer encoding to improve small-object sensitivity. The authors report enhanced convergence stability and improved recall over baseline DETR, while noting increased computational demands. Efficiency-oriented designs are exemplified by (Yu and Xu, 2025), which employs a lightweight GhostNet backbone combined with bidirectional feature fusion and contrastive sampling strategies for partially annotated datasets. The hierarchical soft-NMS post-processing aims to better handle nested and clustered craters. The model achieves competitive accuracy with reduced parameter counts. Feature pyramid strategies have been systematically investigated in (Chaini et al., 2025), where multimodal inputs (optical imagery, DEMs, slope maps) are fused within a Mask R-CNN-style pyramid framework. The study highlights the importance of modality alignment and cross-scale integration for robust detection across varying crater sizes. Complementary perspectives are offered by review studies such as (Sinha and Paul, 2025),

(Chen et al., 2024), which synthesizes architectural trends including increasing reliance on multiscale feature pyramids, modality fusion, and transformer-based representations. A direct application of object detection is exemplified by (Pokorny et al., 2025) where a YOLO architecture multi-scale was deployed and applied to $22,256$ images characterizing the PSRs using KPLO ShadowCam imagery and identifying 1 billion impact craters with duplications due to multi-scale processing data. Previous work introduced YOLOLens (La Grassa et al., 2023), a deep learning model integrating super-resolution with YOLO-based crater detection to enhance mapping of planetary surfaces. YOLOLens demonstrated significant improvements in crater detection accuracy on multimodal datasets, and provided a basis for global crater catalogs on the Moon and Mercury (La Grassa et al., 2025a)(La Grassa et al., 2025b). Building upon these foundations, we present YOLOLens2.0 that integrates a Dense-Residual-Connected Transformer (DRCT) for SR reconstruction with a YOLO-derived crater detector, enabling robust morphometric mapping from multimodal inputs even under extreme low-light conditions. Ablation studies on model components demonstrate the contribution of each module: removing the DRCT SR branch reduces detection accuracy, while isolating the YOLO detection module without multimodal input leads to missed craters in deep shadow areas.

This study is designed to rigorously investigate the following research hypotheses:

(H1) *Whether the performance of an object detection task can be significantly enhanced through the integration of a super-resolution module.*

(H2) *Whether the super-resolution task itself can benefit, in measurable and statistically robust terms, from the incorporation of object detection–driven supervision or guidance.*

By systematically analyzing these bidirectional interactions, the present work aims to clarify the extent to which SR and object detection can act as mutually reinforcing processes within a unified learning framework.

2. Study Area and Ground Truth

The study focuses on the lunar polar regions, specifically the latitude bands between 70° and 90° in both the northern and southern hemispheres. These regions are characterized by extreme illumination conditions, including permanently shadowed regions (PSRs), low solar incidence angles, and strong topographic contrasts. Such conditions pose significant challenges for both SR and object detection, particularly for small and low-contrast impact craters. For the analysis of PSRs, we further use data from the ShadowCam instrument (Robinson et al., 2023), part of the Korea Pathfinder Lunar Orbiter (KPLO) mission (Jeon et al., 2024). ShadowCam is an optical instrument specifically designed to peer into the Moon’s PSRs, the camera is over $200\times$ more sensitive than the Lunar Reconnaissance Orbiter Camera (LROC) from which it was derived. By capturing high-resolution images (up to 1.7 m/pixel) using only secondary light reflected from nearby lunar terrain or earthshine, ShadowCam maps the morphology of these dark craters to search for evidence of water ice and other volatiles. Its data is critical for identifying potential resources and hazards for future human and robotic exploration, including NASA’s Artemis missions. In particular, we considered high-resolution imagery covering the following regions of interest: Nobile West Rim, Hinshelwood, Mouchez, LCROSS site, Nobile–Mouton, Ibn Bajja, Erlanger, Hermite A and Shackleton. These data are used to qualitatively and quantitatively assess the robustness of the proposed framework under extreme photometric constraints.
The Ground Truth (GT) topography is derived from the Terrain Camera (TC), a high-resolution stereo optical instrument onboard the Japanese Selenological and Engineering Explorer (SELENE) mission, better known as Kaguya (Haruyama et al., 2008). Launched by JAXA, the Kaguya TC provided global lunar coverage with a native spatial resolution of 10 meters per pixel, enabling the creation of highly accurate DTMs. To facilitate precise photometric analysis, these DTM tiles are resampled to match the specific spatial resolution of the corresponding visible imagery, ensuring pixel-level consistency between geometric surface properties and observed radiance. The resulting dataset is massive in scale, comprising two synchronized sets of approximately $37,500$ large-format tiles one for visible imagery and one for the associated DTM representing a total data volume of approximately 3.4 TB. This robust data pipeline provides the necessary foundation for systematic tiling and high-density grid generation across the lunar polar domains.

2.1 Grid Construction and Self-Label Generation

A complete regular grid was generated covering the 70°–90° latitude range in both hemispheres. The resulting grid contains more than $120,000$ tiles cropped to 416×416 pixels with a small percentage of overlap to mitigate boundary artifacts. As crater annotations are not available at this scale, labels were generated through a self-labeling strategy. Specifically, the previous multimodal YOLOLens model (La Grassa et al., 2025b) was used in inference mode to produce crater detections over the full grid. A high confidence threshold of 0.57 was enforced to retain only highly reliable detections and minimize noise propagation. Processing the full set of approximately 120,000 tiles required approximately three days of computation. To promote effective learning and avoid sparse supervision, tiles containing fewer than 15 predicted craters were discarded. After filtering and polishing, the final dataset consists of $118,135$ instances of the training set (approximately 152 GB) and $3,654$ instances of the validation set. This curated dataset provides dense crater distributions and diverse illumination conditions across both polar regions. The combination of visible imagery, topographic elevation (DTM), and dynamically generated hillshade enhances high-frequency surface structure while preserving quantitative elevation fidelity. The large-scale polar coverage, high-density crater annotations, and multimodal fusion strategy provide a comprehensive framework for evaluating both SR reconstruction and object detection under challenging illumination and geometric regimes.

3. Methodology

The YOLOLens2.0 architecture integrates SR reconstruction and crater detection into a unified end-to-end framework (Fig. 1). At its core, the system couples a Dense-Residual-Connected Transformer (DRCT) (Hsu et al., 2024) with an affine calibrator module (La Grassa et al., 2026), enabling simultaneous SR of both the panchromatic imagery and the corresponding coarse DTM. This multimodal enhancement is critical: the SR branch is explicitly designed to recover fine-scale topographic and photometric structures that are otherwise suppressed in low-illumination polar environments. The

DRCT module leverages dense residual connections and local self-attention to propagate high-frequency information while maintaining stability during training. Its output is further refined through an affine calibrator, which enforces local geometric consistency and corrects systematic biases in the predicted elevation field. The resulting high-fidelity representations serve as input to the YOLO-based detection head, improving crater localization, diameter estimation, and confidence scoring. The training strategy jointly optimizes the reconstruction and detection components through a composite loss function. The reconstruction branch is supervised via pixel-wise L1 losses, gradient-consistency constraints, refinement losses targeting high-frequency elevation details, and total-variation regularization to suppress artifacts. In parallel, the detection head is trained using standard YOLO loss terms (box regression, classification, and distribution focal loss). The final model has $77,600,292$ learnable parameters. During inference, ShadowCam images at 1.9 m/pixel are processed through the YOLOLens2.0 pipeline, which employs a dynamic multi-scale tiling approach to capture craters of varying sizes (from 416×416 to 3328×3328 by using factors [x1, x2, x4, x8]). The model reconstructs terrain structure and panchromatic appearance at a super-resolved 0.9 m/pixel. These enhanced multimodal products are then fed into the crater detector to extract crater centers, diameters, and morphometric parameters. To ensure geospatial accuracy across varying extraction scales, the high-resolution bounding box predictions are mathematically normalized and mapped back to the original 1.9 m/pixel coordinate space. Finally, the original affine geotransform is applied to generate a catalog of georeferenced crater locations and derived metrics, enabling precise geomorphological analyses within permanently shadowed regions.

3.1 Data Augmentation and Preprocessing

All preprocessing operations are performed dynamically during training, while the dataset itself preserves the original, unnormalized values. A hillshade channel is computed in real time from the Kaguya DTM to enhance local topographic gradients. Illumination parameters are randomly sampled for each tile to simulate varying solar geometries: Azimuth $\in [285^\circ, 345^\circ]$ (centered around $315^\circ \pm 30^\circ$) and altitude $\in [25^\circ, 65^\circ]$. The resulting hillshade is normalized to the $[0, 1]$ range and concatenated with the visible and DTM channels, forming a three-channel input tensor: (visible, DTM, hillshade). Normalization is performed during training, in particular, visible imagery is normalized using local normalization to preserve contrast while reducing illumination bias and DTM data are normalized using the global minimum and maximum absolute lunar elevation values and scaled to the $[0, 1]$ interval as for as hillshade values are linearly scaled to $[0, 1]$. All preprocessing steps were manually verified to ensure the correctness of the data fed to the model.

3.2 Formalization of Bidirectional Task Interactions

We formalize the mutual influence between the SR and object detection (DET) tasks in YOLOLens2.0. Let $X \in \mathcal{X}$ denote the low-resolution multimodal input, $Y_{SR} \in \mathcal{Y}^{SR}$ the high-resolution reconstruction target, and $Y_{DET} \in \mathcal{Y}^{DET}$ the detection labels. Let $Z = \phi(X; \theta_{shared})$ denote the shared representation used by both heads, with outputs

$$\hat{Y}_{SR} = f_{SR}(Z), \quad \hat{Y}_{DET} = f_{DET}(Z). \tag{1}$$

The joint multi-task loss is

$$\mathcal{L}_{joint}(\theta) = \mathcal{L}_{SR}(\theta) + \lambda \mathcal{L}_{DET}(\theta), \tag{2}$$

where $\lambda > 0$ weights the contribution of detection supervision.

3.3 Formulation of Bidirectional Task Synergy

Let θ^{DET}, θ^{SR}, and θ^{joint} denote parameters optimized for detection-only, SR-only, and joint SR+DET objectives, respectively. Let Ω_{in} denote pixels inside predicted RoIs, and Ω_{out} its complement. We define the expected losses (risks) as

$$\mathcal{R}_{SR}(\theta) = E_{X,Y_{SR}}\big[\ell_{SR}(f_{SR}(X;\theta), Y_{SR})\big], \tag{3}$$
$$\mathcal{R}_{DET}(\theta) = E_{X,Y_{DET}}\big[\ell_{DET}(f_{DET}(X;\theta), Y_{DET})\big], \tag{4}$$
$$\mathcal{R}^{in}_{SR}(\theta) = E\big[\ell_{SR} \mid x \in \Omega_{in}\big], \tag{5}$$
$$\mathcal{R}^{out}_{SR}(\theta) = E\big[\ell_{SR} \mid x \in \Omega_{out}\big]. \tag{6}$$

Under the assumption that shared representations capture overlapping geometric features relevant to both tasks, the following hold:

1. **Super-resolution enhances detection:**

$$\mathcal{R}_{DET}(\theta^{joint}) < \mathcal{R}_{DET}(\theta^{DET}). \tag{7}$$

2. **Detection enhances SR selectively within RoIs:**

$$\mathcal{R}^{in}_{SR}(\theta^{joint}) < \mathcal{R}^{in}_{SR}(\theta^{SR}), \tag{8}$$
$$\mathcal{R}^{out}_{SR}(\theta^{joint}) \approx \mathcal{R}^{out}_{SR}(\theta^{SR}). \tag{9}$$

3. **Mutual information perspective:**

$$I(Z_{joint}; Y_{DET}) > I(Z_{DET}; Y_{DET}), \tag{10}$$
$$I(Z_{joint}; Y_{SR} \mid R = 1) > I(Z_{SR}; Y_{SR} \mid R = 1), \tag{11}$$

where R is a spatial indicator for RoIs ($R = 1$ inside RoI, $R = 0$ outside).

3.4 Geometric Interpretation in Representation Space

Let $\mathcal{Z} \subset R^d$ denote the shared representation space. Each task defines a subspace of informative features:

$$\mathcal{Z}_{SR} = \{z \in \mathcal{Z} \mid z \text{ encodes high-frequency structure}\}, \tag{12}$$
$$\mathcal{Z}_{DET} = \{z \in \mathcal{Z} \mid z \text{ encodes object-discriminative features}\}. \tag{13}$$

Joint training reshapes the backbone θ_{shared} such that the learned representation subspace

$$\mathcal{Z}_{joint} = \mathcal{Z}_{SR} \cap \mathcal{Z}_{DET} \tag{14}$$

maximizes the overlap of features relevant to both tasks. This ensure us (1) increased mutual information between Z and

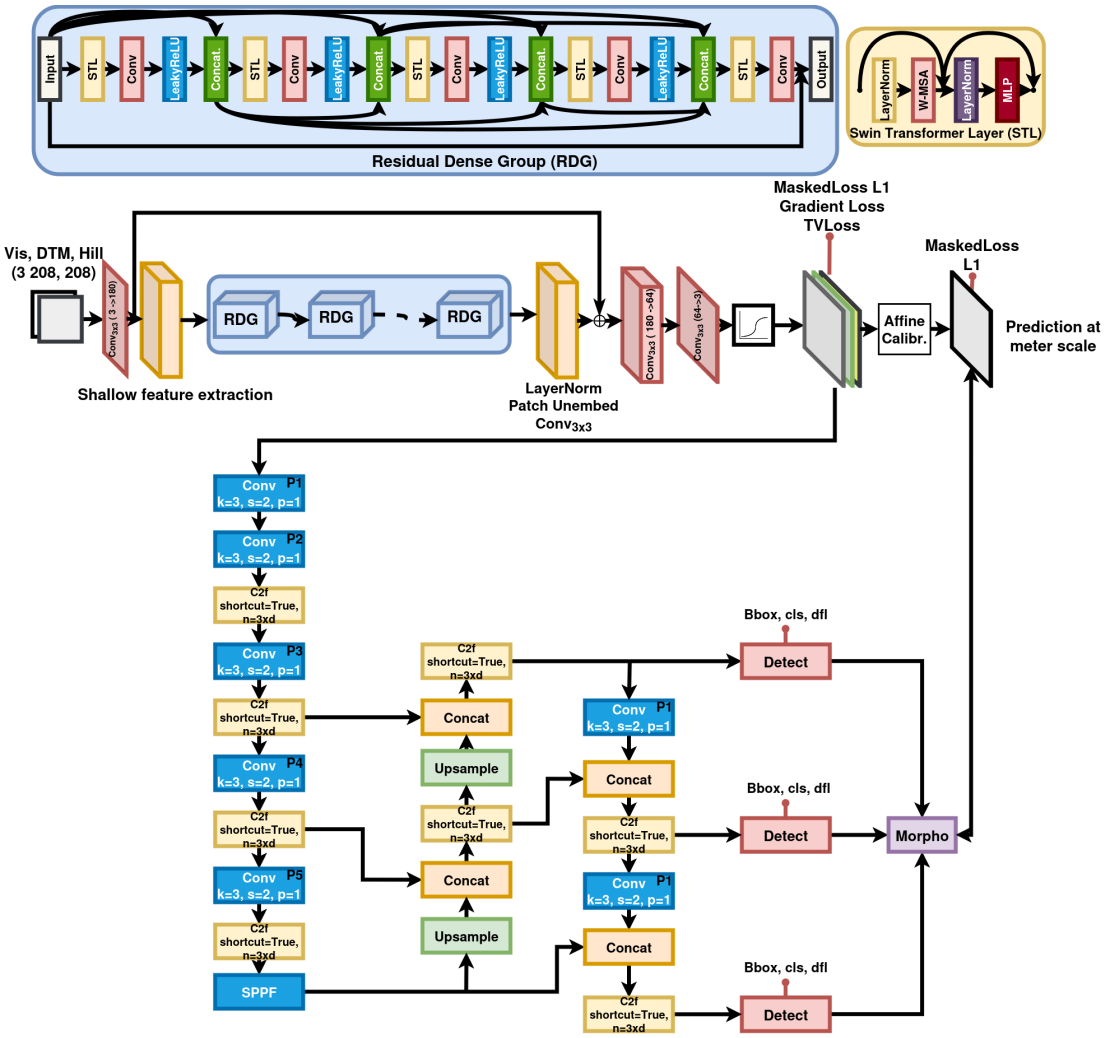

Figure 1. Architecture of YOLOLens2.0, integrating DRCT-based SR and YOLO-derived detection for multimodal lunar mapping.

Y_{DET} (enhanced detection) and (2) spatially concentrated improvements of Z within Ω_{in}, lowering $\mathcal{R}_{SR}^{in}$. This geometric perspective explains both the empirical detection improvements and the observed inside/outside SR asymmetry.

3.5 Conditional Risk Decomposition Over Spatial Domains

The SR risk can be decomposed spatially as

$$\mathcal{R}_{SR}(\theta) = p_{in}\mathcal{R}_{SR}^{in}(\theta) + p_{out}\mathcal{R}_{SR}^{out}(\theta), \tag{15}$$

where $p_{in} = |\Omega_{in}|/|\Omega|$ and $p_{out} = 1 - p_{in}$.

Detection supervision effectively reweights the spatial gradients of the joint loss:

$$\|\nabla_\theta \mathcal{L}_{joint}(x)\| \propto 1 + \lambda \mathbf{1}_{x\in\Omega_{in}}, \tag{16}$$

where $\|\nabla_\theta \mathcal{L}_{joint}(x)\|$ denotes the gradient magnitude, and $\mathbf{1}_{x\in\Omega_{in}}$ is the indicator function for RoIs (1 inside, 0 outside). This formalizes the empirical observation that SR quality improves more inside object regions while remaining stable outside, consistent with the inside/outside asymmetry observed in our experiments.

3.6 Evaluation Metrics

To rigorously evaluate the performance of YOLOLens2.0, we utilize a comprehensive set of metrics designed to assess both the geometric accuracy of the reconstructed Digital Terrain Models (DTMs) and the photometric fidelity of the enhanced optical imagery. This validation process ensures that the SR branch preserves quantitative elevation integrity and structural coherence across the lunar surface.

Mean Absolute Error (MAE): Used to measure the average magnitude of errors in the predicted DTM elevation relative to the Kaguya ground truth.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{17}$$

Where y_i is the ground truth elevation and $\hat{y}_i$ is the predicted elevation for pixel i.

Mean Absolute Difference (MAD): It provides a robust measure of statistical dispersion for the elevation error, helping to identify the influence of extreme outliers in the terrain reconstruction.

$$\text{MAD} = \text{median}(|y_i - \hat{y}_i|) \tag{18}$$

Peak Signal-to-Noise Ratio (PSNR): Evaluates the reconstruction quality for visible imagery and hillshade channels by comparing the maximum possible power of the signal to the power of corrupting noise.

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{L^2}{\text{MSE}} \right) \tag{19}$$

Structural Similarity Index (SSIM): Assesses the preservation of structural information, including luminance, contrast, and local texture. This is vital for maintaining crater rim definition and overall topographic coherence.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{20}$$

Where μ and σ represent the mean and variance of the images x and y, and c_1, c_2 are stability constants.

Gradient Magnitude Similarity Deviation (GMSD): Using Sobel operators, we compute the horizontal (g_x) and vertical (g_y) gradients to determine the Gradient Magnitude (G) for both prediction and ground truth. The local similarity is defined as:

$$GMS(i) = \frac{2 \cdot G_{pred}(i) \cdot G_{gt}(i) + C}{G_{pred}(i)^2 + G_{gt}(i)^2 + C} \tag{21}$$

where $C = 0.0026$. The final GMSD score, acting as a perceptual fidelity indicator, is the standard deviation of the similarity values over the masked region of N pixels:

$$\text{GMSD} = \sqrt{\frac{1}{N} \sum_{i \in \text{mask}} (GMS(i) - \mu)^2} \tag{22}$$

where μ is the mean gradient similarity.

4. Experiments and Results

The training procedure, conducted over more than 30 epochs, exhibits stable and monotonic convergence in both the reconstruction and detection loss terms, thereby confirming the robustness and consistency of the proposed joint optimization strategy. On the Kaguya evaluation set, the final model achieves a recall of 76.9, mAP@50 of 0.682, and mAP@50–95 of 0.479. The gap between mAP@50 and mAP@50–95 reflects the expected increase in difficulty under stricter IoU thresholds, yet the retained performance confirms reliable localization precision. As illustrated in Fig. 2, the SR consistently reconstructs small and low-contrast craters following multimodal enhancement, demonstrating resilience under challenging polar illumination conditions. By activating the super-resolution (SR) module, the model achieves a recall of 89.20%, as reported in Table 1, thereby demonstrating its effectiveness relative to the previous version.

Model	Recall (%) N/S Poles
YOLOLens (La Grassa et al., 2025a)	55.65
YOLOLens Multimodal (La Grassa et al., 2025b)	70.20
YOLOLens2.0 Multimodal	**89.20**

Table 1. Recall comparison among the previous models and the new one configuration over the N/S polar regions.

Reconstruction performance, summarized in Table 2, further supports the effectiveness of the unified architecture. Across 3,654 validation samples, YOLOLens2.0 attains a MAD of 20.49 meters and an MAE of 21.80 meters in vertical elevation with respect to Kaguya ground-truth topography. The proximity between MAD and MAE indicates a limited influence of extreme outliers and suggests a stable error distribution across diverse terrain morphologies. These values confirm that the super-resolved DTMs preserve quantitative elevation fidelity rather than merely improving perceptual sharpness.

Model	MAD	MAE	PSNR vis.	PSNR Hill.
YOLOLens2.0	20.49	21.80	28.57	23.77

Table 2. Validation results across 3,654 samples. We report MAD and MAE jointly with PSNR for panchromatic and hillshade predictions. Metrics are computed on downsampled images to enable a fair comparison of SR against self-consistent GT.

Photometric consistency is likewise maintained. The enhanced panchromatic channel achieves a PSNR of 28.57, while the hillshade representation reaches 23.77. The PSNR difference between visible and hillshade domains is expected, as hillshade encodes nonlinear illumination effects that amplify minor geometric discrepancies. Importantly, all metrics are computed on downsampled images to ensure self-consistent comparison with reference data, avoiding artificial inflation of reconstruction quality. Qualitative visualizations corroborate the numerical results. Paired comparisons between original Kaguya inputs and YOLOLens2.0 outputs reveal improved delineation of crater rims, sharper edge transitions, and enhanced recovery of high-frequency terrain structures. Detection overlays confirm accurate localization even in regions characterized by low contrast and complex geomorphology. Additional geometric validation is provided through terrain profile analysis (Fig. 3). Across three randomly selected validation transects, the predicted DTM profiles closely follow the Kaguya ground-truth elevation curves. The agreement is preserved across varying slopes and rim boundaries, reinforcing confidence in the structural integrity of the reconstructed topography. Taken together, the quantitative metrics and qualitative evidence demonstrate that YOLOLens2.0 achieves a balanced and physically coherent optimization. The model sustains strong detection accuracy while preserving both geometric and photometric fidelity in the super-resolved outputs, even under the extreme illumination constraints characteristic of permanently shadowed lunar polar regions.

4.1 Validation on ShadowCam Data

To assess the generalization capability of YOLOLens2.0, we performed a direct evaluation on ShadowCam imagery acquired over lunar permanently shadowed regions (PSRs). Importantly, the model was applied without any fine-tuning, domain adaptation, or parameter recalibration on ShadowCam data. The network weights remained exactly those obtained during training on the original dataset. This experimental protocol therefore constitutes a strict out-of-distribution validation. ShadowCam orthomosaics used in this study have a native ground sampling distance of approximately 2 m/px. Despite this limitation, YOLOLens2.0 produces super-resolved visible imagery and corresponding DTM predictions at an effective spatial resolution of approximately 1 m/px. This enhancement is achieved through the model’s integrated SR and detection pipeline, which jointly refines high-frequency spatial

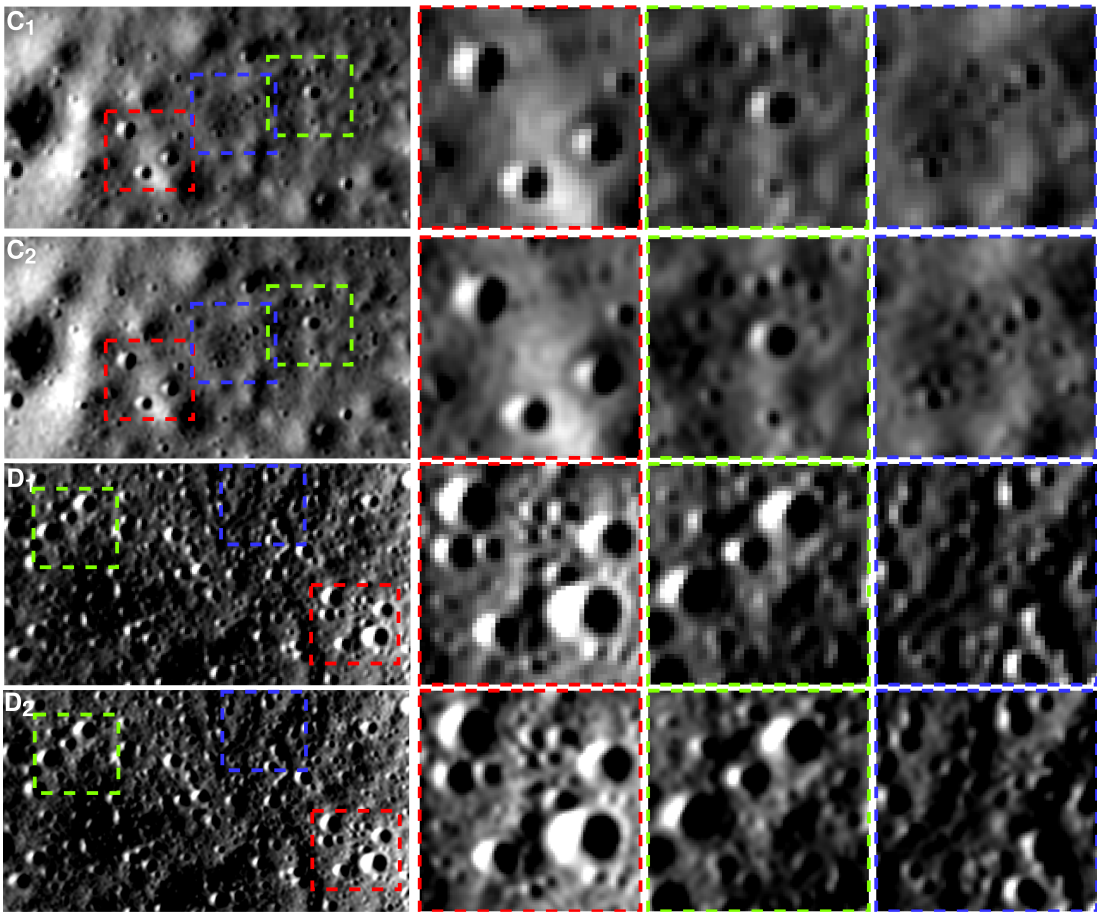

Figure 2. Direct comparisons without (C1, D1) and with SR effect (C2 and D2) of three different random location, shows significant improvements to catch subtle details of lunar terrain.

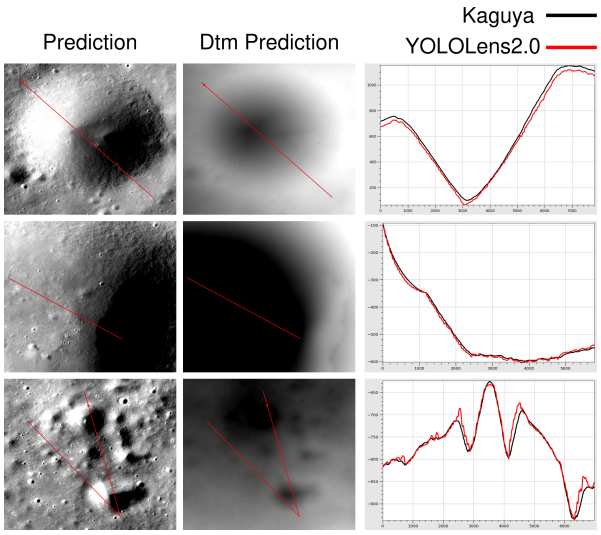

Figure 3. Terrain profiles of three random section (Column 1 visible images, Column 2 DTMs) shows the high similarity between DTM predictions and Kaguya source (Column 3).

content while preserving geomorphological consistency. Figure 10 presents the cumulative crater size–frequency distributions extracted within PSRs of six lunar craters (Figures 4, 5, 6, 7, 8, 9) using YOLOLens2.0 predictions on ShadowCam data. The red curves represent the cumulative crater distributions obtained from model detections, while the black curves correspond to the Neukum production function reference. The agreement between the empirical SFDs and the theoretical production function demonstrates that YOLOLens2.0 preserves the statistical scaling behavior expected for lunar crater populations, even under extreme illumination conditions and without domain-specific retraining.

5. Discussion

This section provides a structured analysis of the architectural and learning improvements introduced in YOLOLens2.0. The discussion proceeds along three complementary directions. First, we establish the performance gains of YOLOLens2.0 with respect to previous YOLOLens configurations. Second, we demonstrate that object detection performance is significantly improved by the inclusion of the SR module. Third, we analyze the reciprocal effect, namely whether the SR task benefits from joint optimization with object detection. Table 1 reports the recall comparison across 3,654 random tiles located in the North and South lunar polar regions. Recall is computed with respect to ground-truth annotations over identical latitude ranges to ensure strict comparability. The progression across model versions is clear and monotonic: the original YOLOLens achieves 55.65% recall, the multimodal extension increases performance to 70.20%, and YOLOLens2.0 Multimodal reaches 89.20%. This represents an absolute improvement of more than 33 percentage points over the original configuration and nearly 19 percentage points over the previous multimodal version. These results demonstrate that the architectural refinements and learn-

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1239-2026

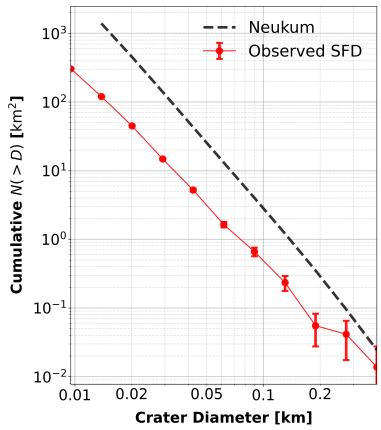

Figure 4. (a) LCROSS

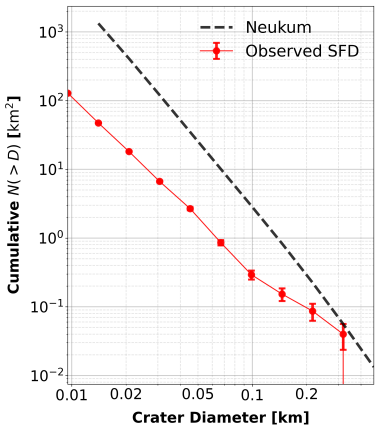

Figure 5. (b) Shackleton

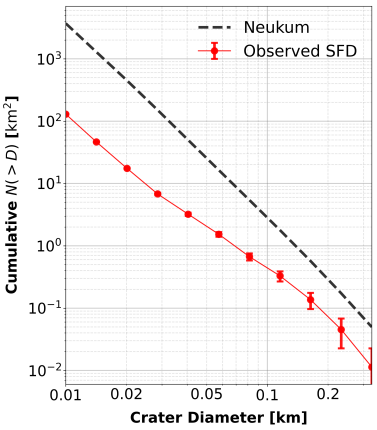

Figure 6. (c) Ibn Bajja

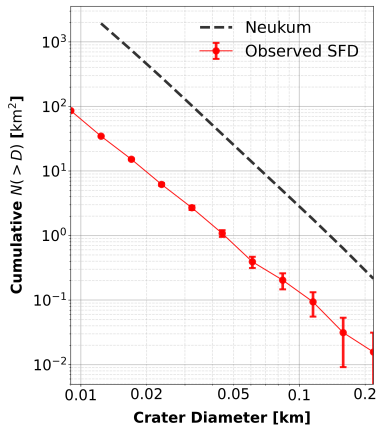

Figure 7. (d) Erlanger

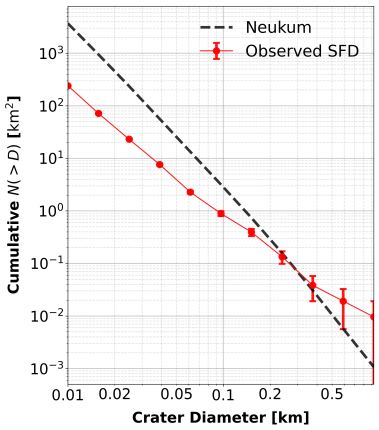

Figure 8. (e) Hinshelwood

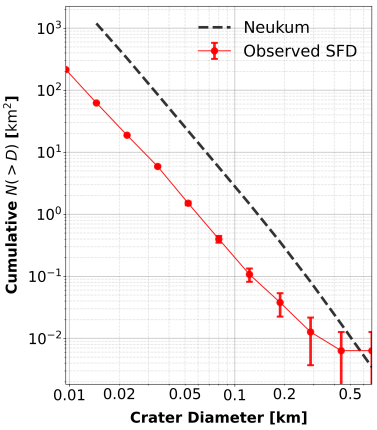

Figure 9. (f) Mouchez L

Figure 10. Cumulative crater size–frequency distributions (SFDs) derived within permanently shadowed regions (PSRs) of six lunar craters using ShadowCam data and model predictions. The red curve represents the cumulative SFD obtained from the detected craters within each PSR, while the black curve corresponds to the Neukum production function power-law reference model. Panels: (a) LCROSS; (b) Shackleton ; (c) Ibn Bajja; (d) Erlanger; (e) Hinshelwood; (f) Mouchez L.

ing strategies introduced in YOLOLens2.0 materially enhance the model’s ability to retrieve true positives in challenging high-latitude lunar environments. Importantly, the evaluation is conducted on geographically extreme regions characterized by low illumination angles and complex morphology, thereby reinforcing the robustness of the observed gains. From an ablation perspective, the comparison in Table 1 isolates the contribution of progressive architectural enhancements. The transition from YOLOLens to YOLOLens Multimodal quantifies the benefit of multimodal feature integration. The subsequent improvement observed in YOLOLens2.0 reflects the cumulative effect of the refined multimodal design and the integration of the SR component within a unified framework. Because the geographic coverage, and evaluation protocol remain constant, the performance increments can be attributed to architectural and training modifications rather than distributional variation. This staged comparison provides the necessary experimental foundation for the subsequent analyses. Having established the superiority of YOLOLens2.0, we next examine the internal mechanisms responsible for these gains, specifically assessing the impact of SR on detection performance described in Subsection 5.1 and the reciprocal influence of detection supervision on SR quality in Subsection. 5.2.

5.1 Does object detection task improved by super-resolution effect?

We evaluate whether the integration of the SR module leads to measurable improvements in object detection performance. To isolate the effect of SR, we compare YOLOLens2.0 trained and evaluated with the SR branch against the same architecture without SR, while keeping all remaining components and training conditions unchanged. The evaluation is conducted over $3,654$ randomly sampled tiles between $\pm 70°$ and $\pm 90°$ latitude. Detection performance is quantified using Recall, mAP@50, and mAP@50–95, thereby assessing both detection sensitivity and localization precision across increasingly strict IoU thresholds. The results reported in Table. 3 demonstrate substantial and consistent gains when SR is enabled (see Figures. 12, 13, 14, 15). Recall increases from 76.90% to 89.20%, corresponding to an absolute improvement of 12.3 percentage points. Similarly, mAP@50 rises from 0.682 to 0.844, and mAP@50–95 improves from 0.479 to 0.605. The improvement at mAP@50–95 is particularly noteworthy, as this metric penalizes localization inaccuracies more strictly. Therefore, the observed gain reflects not only enhanced detection sensitivity but also improved spatial precision and bounding-box alignment. These findings provide strong empirical evidence that SR materially enhances object detection performance. By recon-

structing high-frequency details and refining structural boundaries, the SR module produces inputs with sharper gradients and improved local contrast. This enrichment of fine-scale spatial information directly benefits the feature extraction process within the detection backbone, resulting in more discriminative representations and more reliable object localization. Importantly, the improvements are consistent across all reported detection metrics, indicating that the effect is not confined to a single evaluation criterion. Rather, the inclusion of SR enhances both object retrieval capability (Recall) and geometric accuracy (mAP@50–95). Taken together, these results establish a clear causal direction: the SR task strengthens the object detection task within the unified YOLOLens2.0 framework.

YOLOLens2.0	Rec. (%)	mAP@50	mAP@50–95
without SR	76.90	0.682	0.479
with SR	**89.20**	**0.844**	**0.605**

Table 3. Performance comparison of YOLOLens2.0 with and without SR across 3,654 random tiles.

5.2 Does the SR task improve as a result of object-detection learning?

We analyze whether the inclusion of an object-detection objective induces measurable improvements in SR quality within semantically relevant regions. To this end, we partition each validation image into two complementary pixel sets: *Inside RoI*, defined as the union of all bounding boxes predicted by the detection head, and *Outside RoI*, corresponding to the remaining pixels. All metrics are computed over valid pixels only. SSIM is evaluated on visible imagery, MAE is computed between predicted and ground-truth DTMs, and GMSD is used as an additional perceptual fidelity indicator (lower values indicate better performance for MAE and GMSD, whereas higher values indicate better performance for SSIM). As reported in Table 4, reconstruction quality is consistently superior inside predicted RoIs. In particular, SSIM values are substantially higher within RoIs, while both MAE and GMSD are lower compared to outside regions. The MAE reduction, although numerically small due to normalization in the $[0, 1]$ range, is systematic and accompanied by a clear perceptual improvement reflected by SSIM and GMSD. Importantly, the magnitude of the SSIM gap indicates that structural consistency is significantly enhanced in object-relevant regions. These findings suggest that, despite the absence of any explicit spatially weighted SR supervision, the joint optimization with the detection objective implicitly biases the SR branch toward allocating greater representational capacity to regions that are discriminative for object localization. In other words, the detection task appears to act as an attention-inducing regularizer, guiding the feature hierarchy to prioritize semantically meaningful structures. The SR network does not merely improve global fidelity; rather, it exhibits spatially selective enhancement aligned with the detection head’s predictions. This behavior supports the hypothesis of a bidirectional interaction between SR and object detection within a unified model. The detection objective encourages localized structural refinement, while the improved high-frequency reconstruction inside RoIs plausibly enhances the quality of intermediate features available to the detection branch. The observed inside/outside performance asymmetry therefore provides empirical evidence that SR and detection can operate as mutually reinforcing processes when trained jointly.

In Figure. 11 we show the evaluation methodology and performance analysis of SR enhancement in object-relevant regions. Validation images are divided into Inside RoI (the union of predicted bounding boxes) and Outside RoI (the background pixels) to assess the impact of the joint detection objective on SR quality. A quantitative comparison of MAE, SSIM, and GMSD across the random regions is compared between inside/outside RoI. The significant performance gap most notably in SSIM and GMSD demonstrates that the model exhibits spatially selective enhancement, prioritizing structural consistency and perceptual fidelity within semantically meaningful regions aligned with the detection head’s predictions. Error bars represent standard deviation as reported in Table 4. These results indicate that the detection objective acts as a selective regularizer, prioritizing semantically meaningful features during the SR process.

Metric	Inside RoI	Outside RoI
MAE	**0.0035** $\pm$ 0.0028	0.004 $\pm$ 0.002
SSIM	**0.9805** $\pm$ 0.0064	0.828 $\pm$ 0.018
GMSD	**0.1484** $\pm$ 0.0082	0.171 $\pm$ 0.008

Table 4. Performance comparison of reconstruction metrics inside versus outside the Region of Interest (RoI). Values represent Mean ± Standard Deviation, with bold text indicating superior performance.

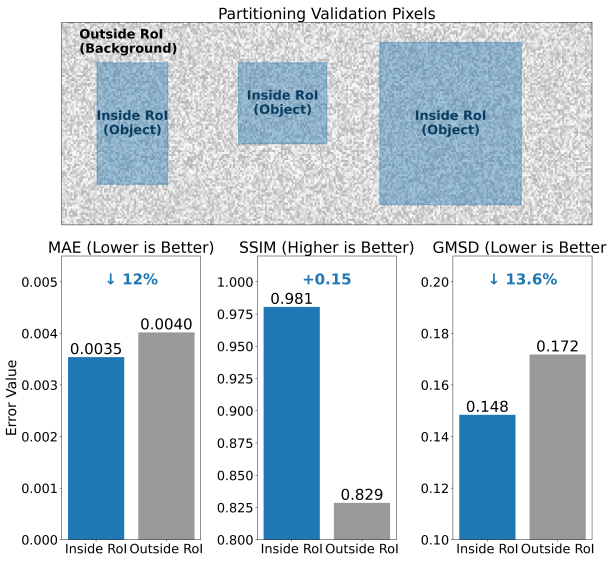

Figure 11. Performance analysis of SR enhancement based on object-detection RoIs. (Top) Partitioning of validation images into Inside RoI (union of predicted bounding boxes) and Outside RoI (background). (Bottom) Comparative metrics (MAE, SSIM, GMSD) showing significant improvements in structural consistency and perceptual fidelity within object-relevant regions.

6. Conclusions

The development of YOLOLens2.0 represents a significant advancement in the unified processing of multimodal lunar data for permanently shadowed regions. By integrating SR and object detection into a single end-to-end framework, this study successfully demonstrates the bidirectional synergy between these two tasks, as evidenced by the validation of both research hypotheses. The validation of H_1 (sec. 5.1) confirms that the integration of a super-resolution module significantly enhances object detection performance. The inclusion of the SR branch

led to an absolute recall increase from 76.90% to 89.20% and a substantial gain in mAP@50-95 (from 0.479 to 0.605), proving that SR not only improves detection sensitivity but also refines localization precision and bounding-box alignment. By reconstructing high-frequency topographic details and sharpening gradients, the SR module provides more discriminative features for the detection backbone. Simultaneously, the results support H_2 (sec. 5.2), demonstrating that the SR task benefits from detection-driven supervision. We observed a marked asymmetry in reconstruction quality, with significantly higher SSIM (0.9805) and lower MAE (0.0035) within predicted Regions of Interest compared to outside areas. This indicates that the detection task acts as an attention-inducing regularizer, guiding the network to prioritize representational capacity for semantically meaningful structures like crater rims. Finally, the application of YOLOLens2.0 to ShadowCam imagery demonstrates the framework's robust generalization, providing high-fidelity georeferenced catalogs at 0.9 m/pixel that adhere to theoretical crater production functions. This unified approach offers a scalable and precise methodology for geomorphological analyses in the Moon's most challenging illumination environments.

References

Chaini, C., Jha, V. K., Rajnish, K., 2025. Multi-scale feature pyramid-based crater detection on lunar surface. *Earth Science Informatics*, 18(3), 305.

Chen, D., Hu, F., Zhang, L., Wu, Y., Du, J., Peethambaran, J., 2024. Impact crater recognition methods: A review. *Science China Earth Sciences*, 67(6), 1719–1742.

Gong, X., Yu, J., Zhang, H., Dong, X., 2025. Aed-yolo11: A small object detection model based on yolo11. *Digital Signal Processing*, 166, 105411.

Guo, Y., Wu, H., Yang, S., Cai, Z., 2024. Crater-DETR: A novel transformer network for crater detection based on dense supervision and multiscale fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–12.

Haruyama, J., Matsunaga, T., Ohtake, M., Morota, T., Honda, C., Yokota, Y., Torii, M., Ogawa, Y., Group, L. W., 2008. Global lunar-surface mapping experiment using the Lunar Imager/Spectrometer on SELENE. *Earth, planets and space*, 60(4), 243–255.

Hsu, C.-C., Lee, C.-M., Chou, Y.-S., 2024. Drct: Saving image super-resolution away from information bottleneck. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6133–6142.

Jeon, M.-J., Cho, Y.-H., Kim, E., Kim, D.-G., Song, Y.-J., Hong, S., Bae, J., Bang, J., Yim, J. R., Kim, D.-K. et al., 2024. Korea pathfinder lunar orbiter (kplo) operation: From design to initial results. *Journal of Astronomy and Space Sciences*, 41(1), 43–60.

La Grassa, R., Cremonese, G., Gallo, I., Re, C., Martellato, E., 2023. YOLOLens: A deep learning model based on super-resolution to enhance the crater detection of the planetary surfaces. *Remote Sensing*, 15(5), 1171.

La Grassa, R., Martellato, E., Cremonese, G., Re, C., Tullo, A., Bertoli, S., 2025a. LU5M812TGT: An AI-Powered global database of impact craters 0.4 km on the Moon. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220, 75–84.

La Grassa, R., Re, C., Martellato, E., Tullo, A., Bertoli, S., Cremonese, G., Vergara Sassarini, N. A., Faletti, M., Galluzzi, V., Giacomini, L., 2025b. From the Moon to Mercury: Release of Global Crater Catalogs Using Multimodal Deep Learning for Crater Detection and Morphometric Analysis. *Remote Sensing*, 17(19). https://www.mdpi.com/2072-4292/17/19/3287.

La Grassa, R., Re, C., Tullo, A., Gallo, I., Cremonese, G., 2026. Transformer-driven monocular high-resolution DTM generation on mars via multimodal integration of CaSSIS imagery and MOLA altimetry. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 19, 100118.

Liu, Y., Lai, J., Xie, M., Zhao, J., Zou, C., Liu, C., Qian, Y., Deng, J., 2024. Identification of lunar craters in the Chang'e-5 landing region based on Kaguya TC morning map. *Remote Sensing*, 16(2), 344.

Pokorny, P., Mazarico, E., Robinson, M. S., Mahanti, P., Williams, J., Fassett, C., 2025. Machine learning driven detection of 1 billion+ lunar impact craters in permanently shadowed regions using shadowcam data. *56th Lunar and Planetary Science Conference (LPSC)*.

Robinson, M. S., Brylow, S. M., Caplinger, M. A., Carter, L. M., Clark, M. J., Denevi, B. W., Estes, N. M., Humm, D. C., Mahanti, P., Peckham, D. A. et al., 2023. ShadowCam instrument and investigation overview. *Journal of Astronomy and Space Sciences*, 40(4), 149–171.

Sinha, M., Paul, S., 2025. Automated detection of craters on the lunar surface using deep learning: A review with insights from Chandrayaan-2 TMC-2 data. *Results in Earth Sciences*, 3, 100094.

Yu, R., Xu, Z., 2025. Crater-MASN: A Multi-Scale Adaptive Semantic Network for Efficient Crater Detection. *Remote Sensing*, 17(18), 3139.

Zhong, Z., Lai, J., Zhong, Y., Xu, Y., Cui, F., 2025. Enhancing lunar DEM data using super-resolution techniques and optimizing the faster R-CNN network for sub-kilometer crater detection. *Icarus*, 430, 116483.

Zuo, W., Gao, X., Wu, D., Liu, J., Zeng, X., Li, C., 2025. YOLO-SCNet: A framework for enhanced detection of small lunar craters. *Remote Sensing*, 17(11), 1959.

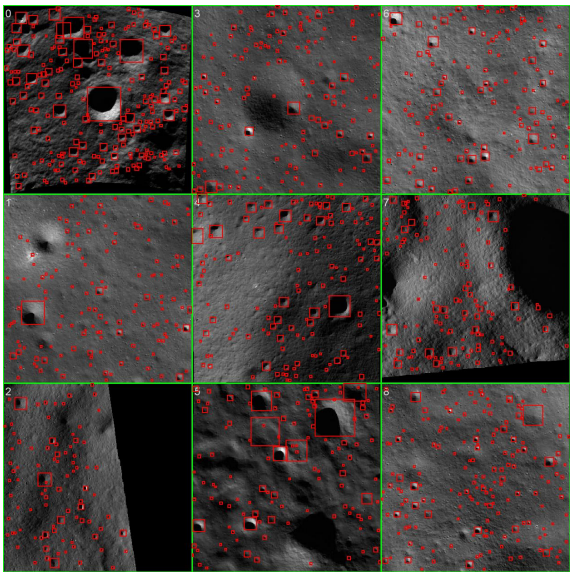

Figure 12. (a) Ground Truth (GT)

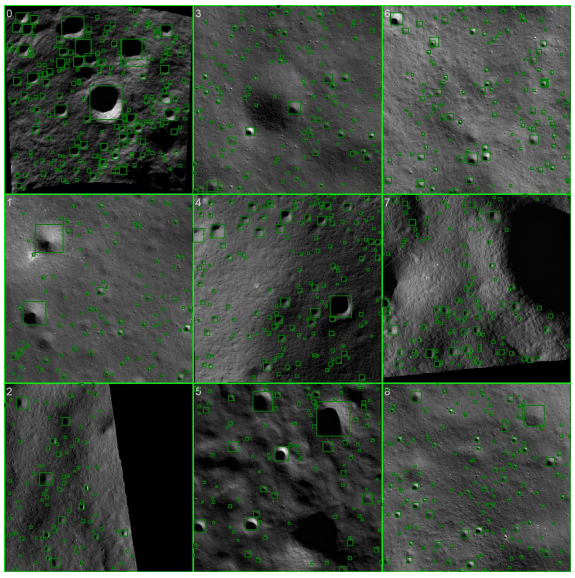

Figure 13. (b) Prediction without SR

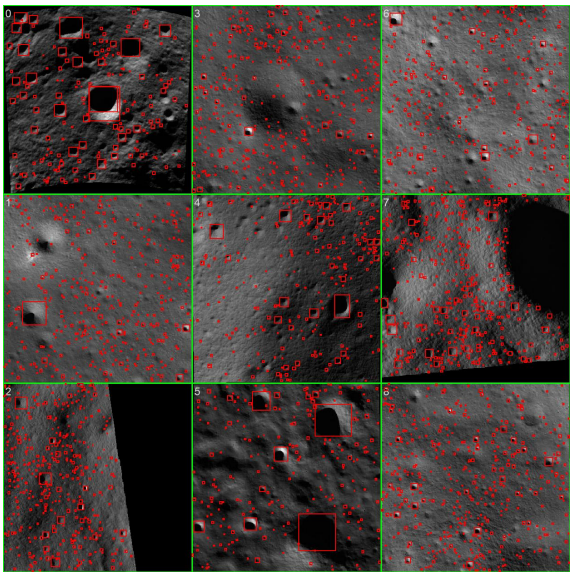

Figure 14. (c) Prediction with SR

Figure 15. Qualitative comparison between the reference elevation map and model predictions. (a) Ground truth (GT). (b) Prediction obtained without the SR module. (c) Prediction obtained with SR enhancement. The SR-integrated configuration demonstrates improved structural coherence and enhanced reconstruction of fine-scale topographic features.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1239-2026