Reconstructing Mercury's Southern Hemisphere

Amil Agrawal, Brandon Montoya, Joanne Ding
Updated implementation report · 31 October 2026

Abstract

We improve the original lunar-to-Mercury gravity reconstruction project while retaining its dense gravity network and separate topographic refinement. The new implementation uses aligned spherical-harmonic inputs, residual dense blocks, multiscale lunar pretraining, geographic validation, and inference that preserves the supplied low-degree gravity. A two-stage degree-25-to-100 cascade reduces mean held-out northern RMSE from 25.70 to 22.56 mGal. These measurements concern prediction of a reference gravity model. The southern output remains a model-dependent reconstruction, and no new interior structure or superiority to the published study is claimed.

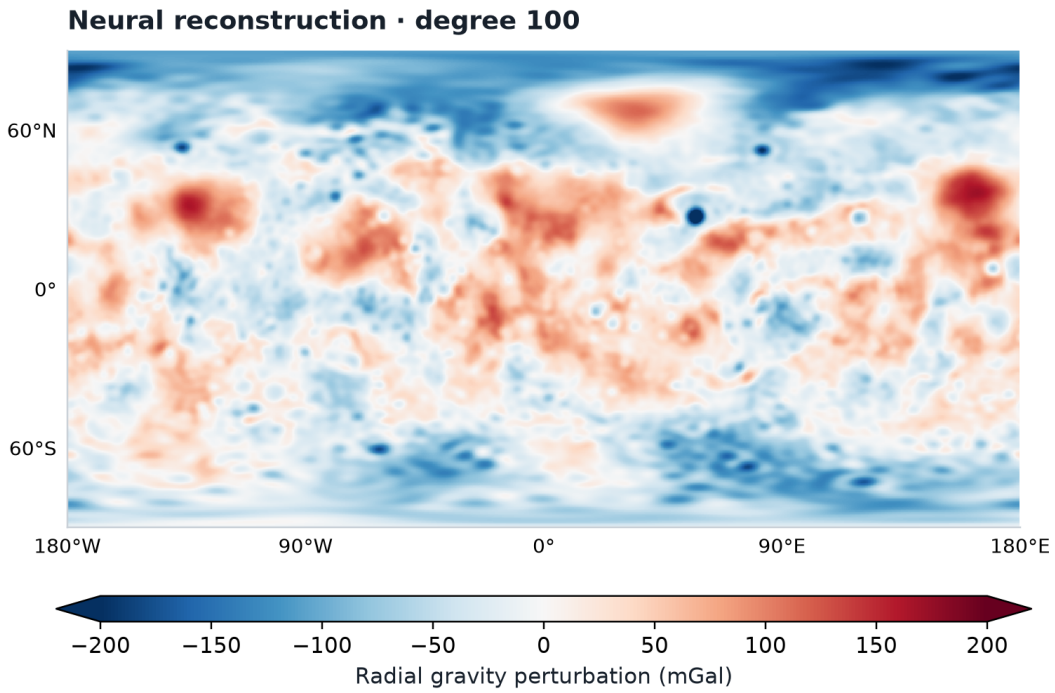


Figure 1. Full reconstructed radial gravity perturbation. This is the neural cascade averaged over four geographic training splits; no northern reference values are pasted into it. The displayed field is synthesized on a denser grid without adding degrees above 100. Color limits affect only the figure.

1. Scope and connection to the original project

Zhao et al. [1] motivate lunar pretraining, Mercury fine-tuning and a DEM refinement network. This revision develops that pipeline as a usable reconstruction system. It retains the original scientific question and poster identity. Improvements concern input geometry, feature representation, network stability, saved training state and full-map inference. It is an implementation extension, not a claim that the architecture or transfer-learning idea is new.

2. Data, coordinates and physical quantity

Mercury gravity is HgM008 through degree 100; the original repository coefficients match this model above degree 1. Lunar pretraining uses GRAIL GRGM900C through degree 200. Shape inputs are the public USGS stereo-derived Mercury model and LDEM128 in the lunar principal-axis frame [2–5]. URLs and SHA-256 hashes are saved in the manifest.

Fields share explicit north-to-south Driscoll–Healy grids with east-positive longitude. Both Mercury stages use 202 × 404 samples. Lunar grids match their respective target bandwidths. There is no polar-image stacking or independent gravity-map resizing. Monopole and degree-one terms are excluded. Values are inward radial gravity perturbations in mGal at the reference sphere, not vector gravity magnitude or free-air anomalies.

Degree-l coefficients are multiplied by $(l + 1)\ GM/R^2 \times 10^5$, using SI constants from each archive header. Inputs are low-degree gravity, its upper half-band, low-degree topography, and target-band topography. High-degree reference gravity is never an input. Normalization is fitted only to the training region and stored in each checkpoint.

3. Network and optimization

The 74,782-parameter network has a 32-channel stem, three residual dense blocks (three 12-channel growth layers each), local 1 × 1 fusion, global fusion and a gravity residual head. A three-convolution DEM refiner uses this residual and the input fields. SiLU and residual scaling of 0.2 support stable optimization. Zero-initialized output heads initially retain the supplied field. The historical TensorFlow model remains available; active training uses PyTorch.

Lunar pretraining interleaves 25 → 50, 50 → 100 and 100 → 200. Mercury adaptation fits the two lower pairs. AdamW uses 30-step warmup, cosine decay, norm-1 clipping, area-weighted MSE and a 0.05 finite-difference loss. Each batch has eight 54 × 54 patches with 24 × 24 supervised centers. Their 15-pixel halos cover the receptive field. Augmentations follow geographic assignment; labels outside the training mask receive zero weight.

Four 90° longitude sectors rotate through test, validation and two training sectors, with 8° edge guards and latitudes 15–75°N. Checkpoints minimize unprojected validation MSE; tests use band-projected predictions. A matched network trained from scratch tests the value of lunar initialization. These are reference-model development comparisons, not independent planetary measurements.

4. Inference and reproducibility

Longitude wraps periodically; pole crossing reflects latitude and shifts longitude by 180°. The absent south-pole sample is estimated from the last ring. Tiles write only valid centers. The residual is projected into its requested harmonic band and added to the supplied gravity. Stage two receives the stage-one prediction. No output blur, clipping or reference-pixel splice is applied.

Tests cover units, harmonic projection, DEM coordinates, normalization, patch alignment, pole handling, longitude equivariance, checkpoint loading and tiled/whole-grid agreement. Checkpoints include architecture and normalization; companion files preserve optimizer and RNG states. Configurations, training curves, fold predictions and evaluation metrics are retained.

5. Measured results

Reconstruction task	Supplied input	From scratch	Lunar transfer
25 → 50	20.24	17.18	17.29
50 → 100	15.23	14.29	14.07
25 → 100 cascade	25.70	22.63	22.56

Table 1. Mean of four area-weighted held-out northern RMSEs, in mGal. The single-stage tests use the supplied low-degree reference field. The cascade uses predicted intermediate gravity. Raw per-sector values are saved in metrics.json and reconstruction-*.json.

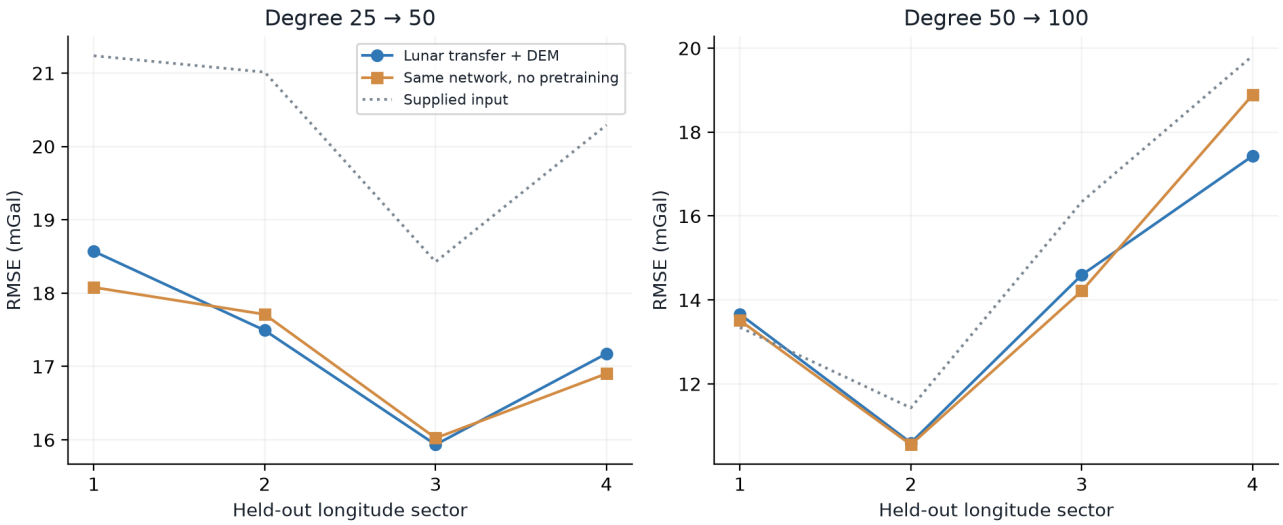


Figure 2. Per-sector results. Lunar initialization is not uniformly better than training the same network from scratch. Both variants and all folds are retained rather than selecting favorable sectors.

The executed run contains 17 fits and 6,920 optimizer steps; measured time inside training and validation loops totals 2.5 minutes. These timings exclude data preparation, final fold evaluation, figure generation and process startup. Settings specify explicit step and wall-time caps. Exact environment and source hashes are preserved with the result.

6. Interpretation and remaining limits

The useful outcome is a complete neural reconstruction pipeline with aligned physical inputs, reliable tiled inference and numerical evidence that it predicts missing reference-model detail in held-out northern areas. It does not establish recovery of true southern gravity. HgM008 is a fitted mission solution with geographically varying information, and topography alone cannot resolve subsurface density. Fold-model spread is sensitivity to training regions, not a calibrated confidence interval.

The corrected fields, geographic splits and radial-gravity definition differ from the old poster and from Zhao et al.'s experiment. Their quoted errors therefore do not form a like-for-like performance comparison. The earlier development benchmark is retained under results/revision. No improvement over the published study is claimed.

7. Southern reconstruction detail

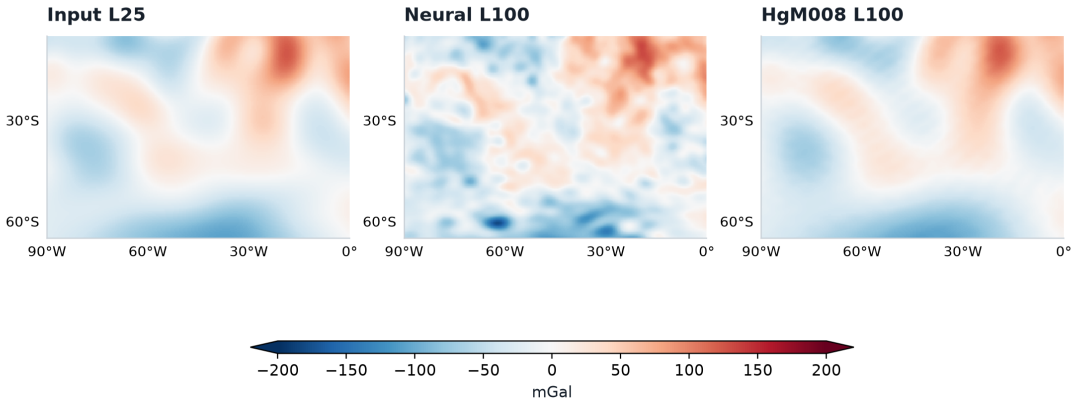


Figure 3. Fixed southern geographic crop shown at the same physical color scale. This is a qualitative view of the output, not a southern accuracy test. HgM008 is shown as a reference solution; apparent visual agreement or extra texture cannot establish recovered physical structure.

8. Reproduce and inspect

Install the repository with `pip install -e ".[dev,reports]"`, then run `python all_steps_run_pipeline.py`. The pipeline acquires checksum-verified public inputs, trains bounded models, reconstructs both variants and renders this report. To reuse completed models, pass `--skip-training`. Individual stages are available as `mercury-train`, `mercury-reconstruct` and `mercury-poster`.

The deployment arrays contain the supplied fields, reconstructed intermediate and final maps, individual fold predictions, member spread, latitude and longitude. Model checkpoints carry all inference normalization. The original implementation and poster remain in legacy/. The root numbered scripts route to the corrected stages. The root moon_model.py retains the TensorFlow architecture for old checkpoints, with serialization fixes.

References

[1] Zhao et al. (2020). A Global Gravity Reconstruction Method for Mercury Employing Deep Convolutional Neural Network. Remote Sensing 12, 2293.
https://doi.org/10.3390/rs12142293

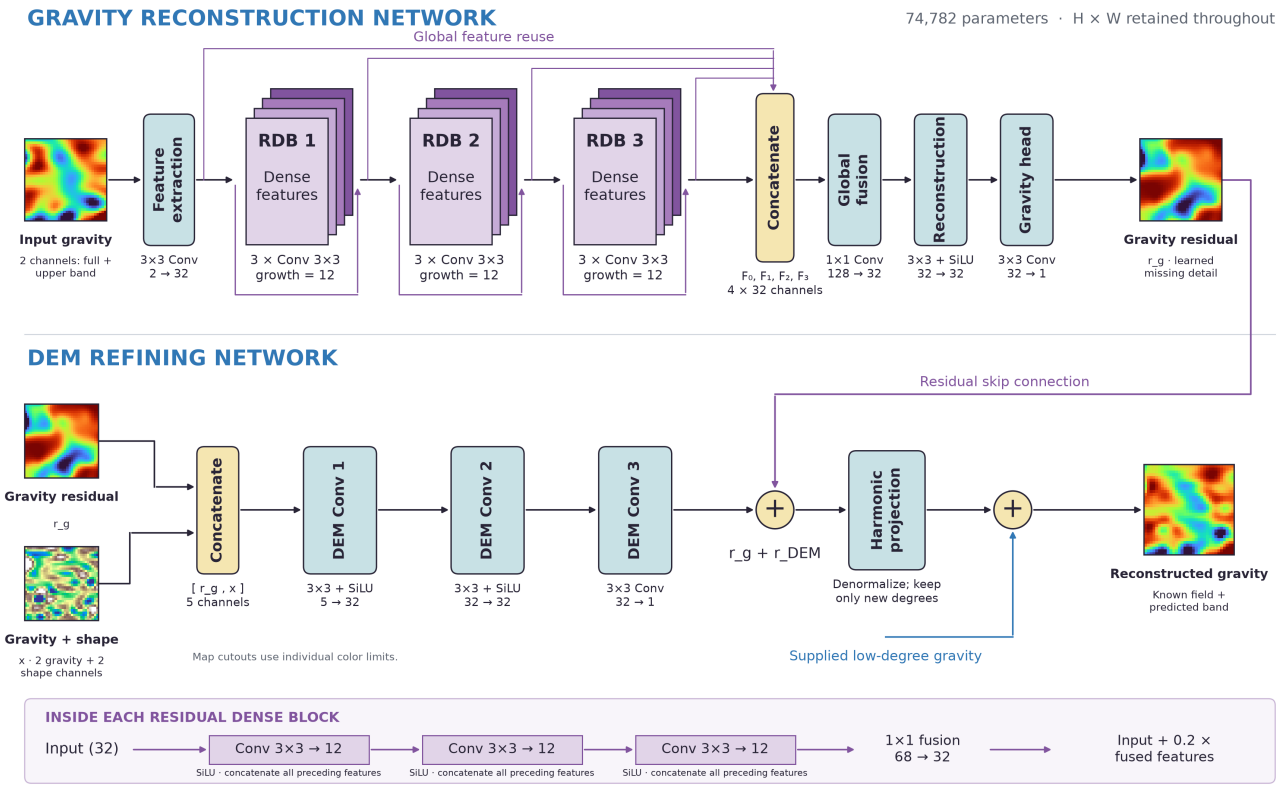
[2] Genova et al. (2019). Geodetic evidence that Mercury has a solid inner core. HgM008 gravity model.
https://doi.org/10.1029/2018GL081135

[3] Lemoine et al. (2014). GRGM900C: A degree 900 lunar gravity model from GRAIL primary and extended mission data.
https://doi.org/10.1002/2014GL060027

[4] Maia (2024). Spherical harmonic models of the shape of Mercury. USGS stereo-photogrammetric shape.
https://doi.org/10.5281/zenodo.10809345

[5] Wieczorek (2024). Spherical harmonic models of the shape of the Moon. LDEM128, principal-axis coordinates.
https://doi.org/10.5281/zenodo.11533784

Architecture and comparison with the source method



Layer-level view of the implemented network, redrawn in the original poster's visual style. Stacks show dense feature reuse, purple paths show residual and fusion connections, and yellow nodes show concatenation or addition. Map cutouts come from the saved degree-50-to-100 model and public inputs; the diagram is also provided as an editable SVG.

What extends the implementation

Lunar transfer, multiscale training and DEM refinement already belong to Zhao et al. [1]. Our added harmonic projection preserves the supplied coefficient band exactly. On the existing four-sector transfer fits, its relative RMSE reductions are 0.86% at 25 → 50 and 1.53% at 50 → 100 relative to the same network without projection. This is an internal ablation, not a head-to-head result against their model.

Their Mercury scale-transfer experiment fine-tunes on 25 → 50 and then evaluates 50 → 100. Our main reported experiment also fine-tunes on northern 50 → 100 pairs, so its task differs. In a separate post-hoc check, applying our 25 → 50 checkpoint directly to 50 → 100 gives 17.44 mGal versus 15.23 mGal for the unchanged input. Scale transfer therefore remains an unresolved limitation; better accuracy than the paper has not been demonstrated.

A valid superiority claim requires a faithful implementation of the published network, matched inputs, the same geographic holdouts and scale-transfer protocol, and comparable optimization budgets. The 425,762-parameter historical repository checkpoint is not the paper's full network. Current measurements support software corrections and exact coefficient preservation, not a stronger southern gravity measurement.