MetricGrids: Arbitrary Nonlinear Approximation with Elementary Metric Grids based Implicit Neural Representation

Supplementary Material

A. Implementation Details

A.1. 2D Image Fitting

The decoder backbone network structure and initialization scheme follow SIREN, with all activation functions using a frequency value of $\omega_0 = 30$. For the experiment on the Kodak dataset, we set coarsest resolution $N_{min} = 16$, finest resolution $N_{max} = 256$, hash table size $T = 2^{14}$, number of levels $L = 13$, number of feature dimensions per entry $F = 2$ of Elementary Metric Grids. Our model has a total of 206,819 parameters, of which 187,040 are for the elementary metric grids and 19,779 for the high-order extrapolation decoder. For fully implicit methods we adjust the hidden layer width, and for hybrid methods we adjust the number of levels L to adapt to the number of parameters. All models are trained for 20,000 iterations in all experiments on the Kodak dataset with a batch size of target image pixels number.

In the experiment comparing with the recent method GaussianImage, two model sizes were evaluated. The medium-sized model (ours-m) utilized $N_{max} = 384$, $T = 2^{15}$, $L = 12$ and the larger model (ours-l) utilized $N_{max} = 384$, $T = 2^{16}$, and $L = 13$, other parameters remained consistent. These two models comprised 334,947 and 532,659 parameters, respectively. Evaluate metrics of the comparison methods in Tab. 2 are taken from GaussianImage.

In the gigapixel image fitting experiment, we set the finest resolution of grids to $1/2$ maximum resolution of the target image, hash table size $T = 2^{24}$, number of levels $L = 16$, and trained all models for 35,000 iterations with a batch size of 2^{18}.

A.2. 3D Shape Reconstruction

We adopt the same experimental hyperparameter settings and evaluation metric calculation methods in NeuRBF, where NeuRBF allocates a portion of the parameters to the adaptive radial basis functions by setting the number of feature dimensions per entry $F = 1$. we maintain $F = 2$, consistent with the baseline, and adjust the number of levels to $L = 15$ to balance the parameter increase introduced by the high-order extrapolation decoder.

A.3. Neural Radiance Field Reconstruction

We follow the training settings of NerfAcc, where the model parameters include coarsest resolution $N_{min} = 16$, finest resolution $N_{max} = 4096$, hash table size $T = 2^{19}$, number of levels $L = 16$, number of feature dimensions per entry $F = 2$, training with a batch size of 1024 for 30k iterations, and using occupancy grids for efficient sampling. For the NeRF reconstruction experiment, due to the complexity of the color and density estimation problems, we follow previous methods and use the rendering equation-based decoder proposed in NRFF (same as NeuRBF), the model size is slightly larger than the baseline model.

B. Limitations and Future Work

In this work, we focus on enhancing hybrid neural representations based on regular feature grids. Integrating our low-order and high-order terms with a level-of-detail structure represents a promising research direction. Regarding computational efficiency, although the current sequential implementation of MetricGrids processes grids iteratively, a parallel implementation is theoretically feasible. This is demonstrated by the multi-scale hash encoding, which can significantly reduce the method's time complexity.Additionally, the scalability of MetricGrids in reconstructing complex signals from images and shapes has been demonstrated, extending the approach to real-world and dynamic scenes in neural radiation fields remains an open challenge. we have explored the application of the proposed method for reconstructing large-scale scenes in real-world, such as the Mip-NeRF 360 dataset. While our method consistently outperforms the baseline, it still lags behind state-of-the-art methods. Combining our method with complex scenes represents another potential direction for future research.

C. Additional Results

Fig. S_1 and Fig. S_2 provide additional qualitative comparison results for other gigapixel images and the Kodak dataset fitting experiments, respectively. To better highlight the differences between the methods, we present L2 error maps, where the L2 error is calculated for each color channel and combined as a weighted average. Fig. S_3 provides additional 3D SDF reconstruction results, including examples of complex shapes. Our method consistently demonstrates high-accuracy fitting capabilities across a wide range of shapes. Fig. S_4 provides more close-up comparisons of NeRF reconstruction on the Blender dataset. Our method demonstrates a clear advantage in reconstructing fine structures. Tab. S_I and Tab. S_II present the quantitative results of our method for each scene in the 3D SDF and NeRF reconstruction tasks.

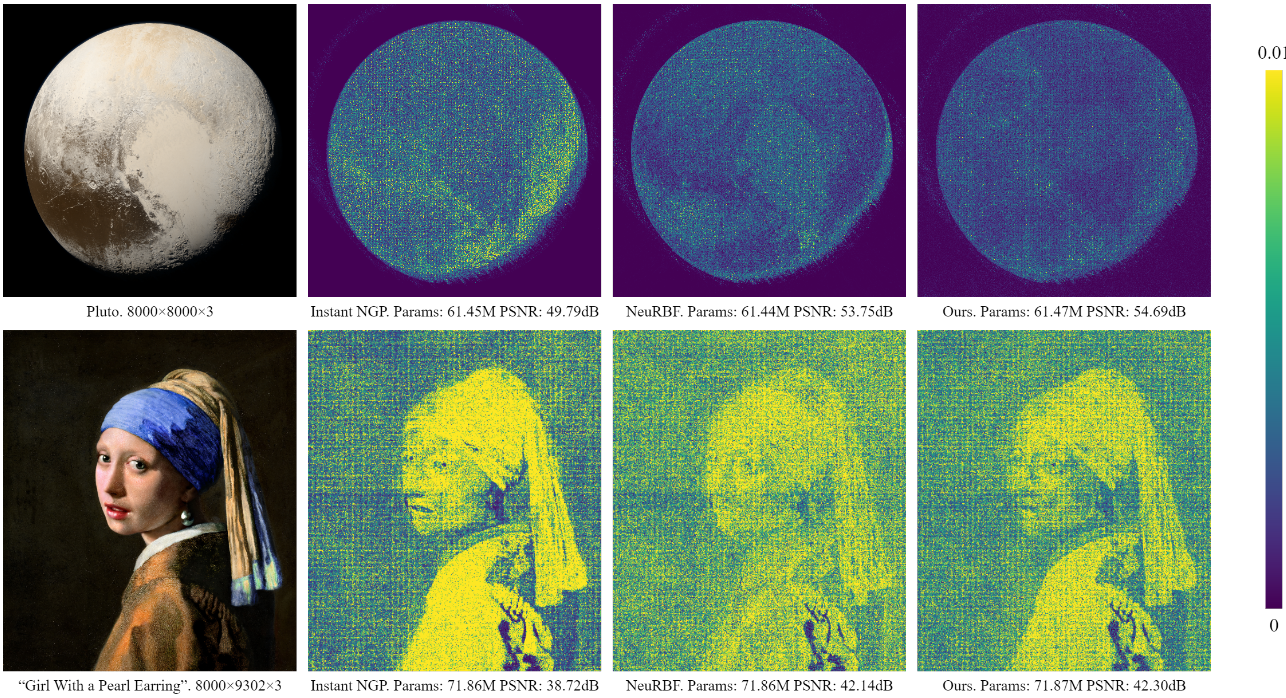


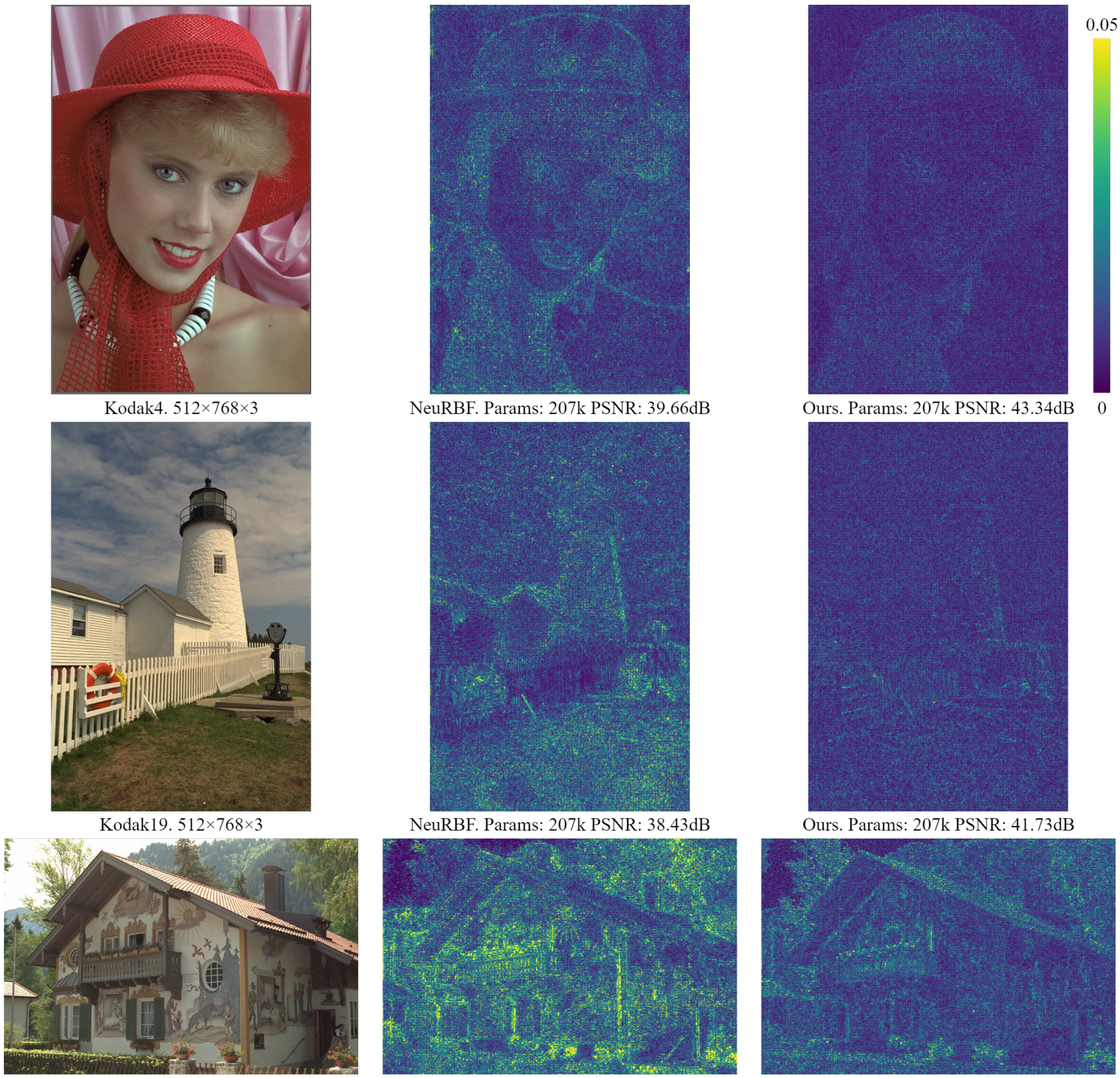
Figure S_1. Additional qualitative comparison on Gigapixel images. From left column to right displays our fitting results, the L2 error map of two compared methods, and the L2 error map of our method. Relevant details and evaluation metrics are provided below each image. We clamp the error range to 0 0.01 corresponding to pixel values ranging from 0 to 1. "Girl With a Pearl Earring" renovation ©Koorosh Orooj (CC BY-SA 4.0).

Metrics	Armadillo	Bunny	Dragon	Buddha	Lucy	XYZ Dragon	Statuette	Avg.
NAE ↑	2.6655	1.8194	2.0718	2.4848	3.185	4.088	5.343	3.0939
Chamfer ↓	0.0021	0.0028	0.0024	0.0022	0.0017	0.0018	0.0020	0.00214
IoU ↓	0.99997	0.99962	0.99997	0.99997	0.99998	0.99997	0.99994	0.999917

Table S_I. Quantitative evaluation of the proposed method on the 3D SDF reconstruction task. Per-scene breakdown of the quantitative metrics in Table 3 of the main text.

Metrics	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg.
PSNR ↑	37.04	26.80	34.95	38.92	37.11	32.65	39.65	33.11	35.028
SSIM ↑	0.989	0.951	0.987	0.988	0.985	0.971	0.996	0.943	0.976
LPIPS$_{\text{VGG}}$ ↓	0.0132	0.0526	0.0163	0.0197	0.0169	0.0394	0.0056	0.0985	0.0328

Table S_II. Quantitative evaluation of the proposed method on the NeRF reconstruction task. Per-scene breakdown of the quantitative metrics in Table 4 of the main text.



Kodak4. 512×768×3 | NeuRBF. Params: 207k PSNR: 39.66dB | Ours. Params: 207k PSNR: 43.34dB

Kodak19. 512×768×3 | NeuRBF. Params: 207k PSNR: 38.43dB | Ours. Params: 207k PSNR: 41.73dB

Kodak8. 768×512×3 | NeuRBF. Params: 207k PSNR: 36.11dB | Ours. Params: 207k PSNR: 38.83dB

Figure S_2. Additional qualitative comparison on Kodak dataset. From left column to right displays our fitting results, the L2 error map of baseline method, and the L2 error map our method. Relevant details and evaluation metrics are provided below each image. We clamp the error range to 0 0.05 corresponding to pixel values ranging from 0 to 1.

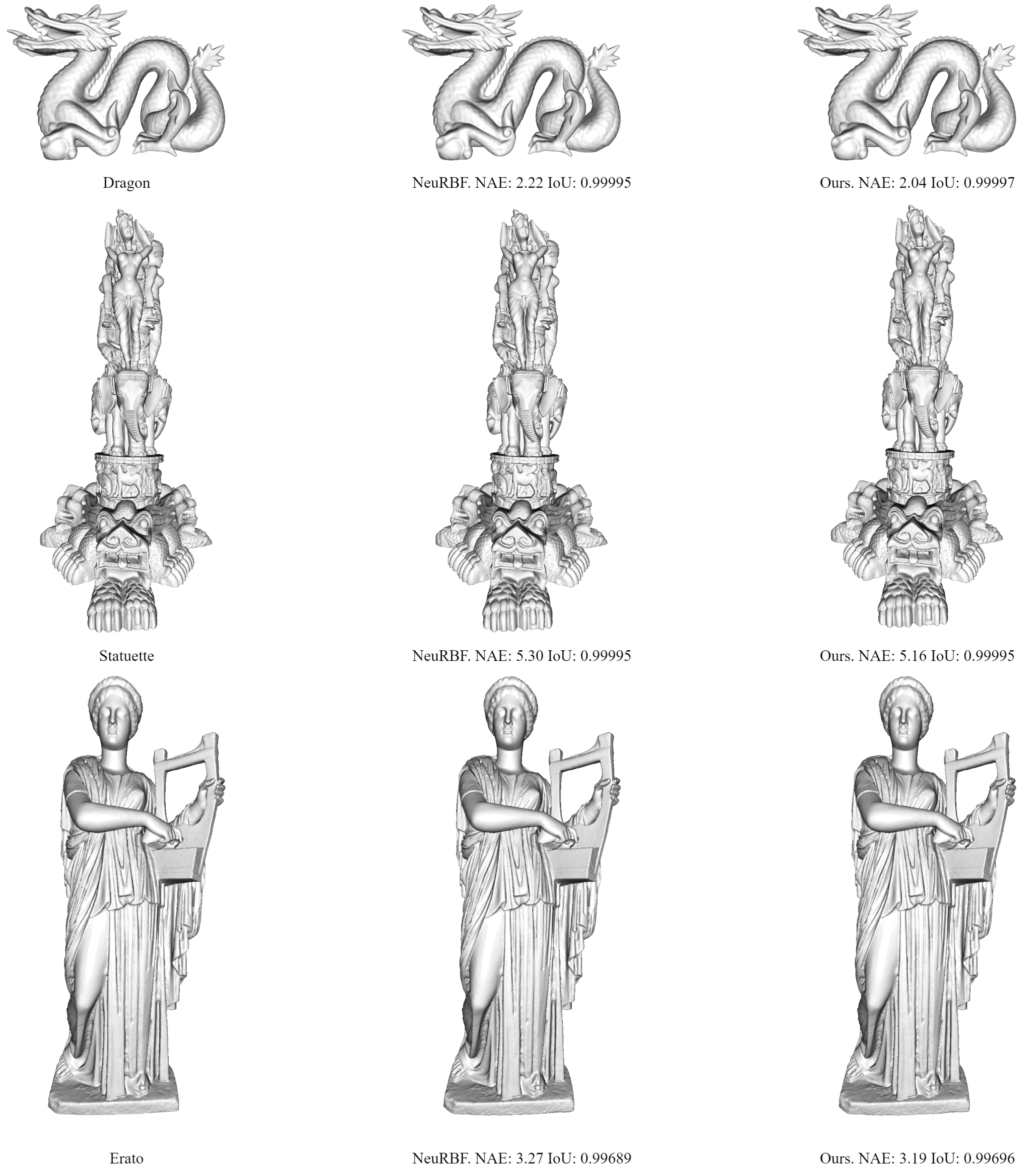


Figure S_3. Additional qualitative comparison on 3D SDF Reconstruction. Beyond the two example scenes from the Stanford 3D Scanning Repository, we also evaluate an additional sculpture scene using the same experimental settings. (Erato Model: © 2016 Geoffrey Marchal, CC BY 4.0)

Instant NGP NeuRBF Ours Ground Truth

Figure S_4. Additional close-up comparison on Neural Radiance Field Reconstruction.