BANF: Band-limited Neural Fields for Levels of Detail Reconstruction

Supplementary Material

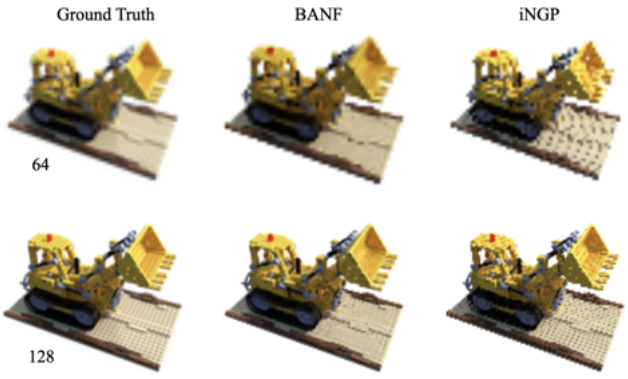
1. Radiance field decomposition

We extend the application of our frequency decomposition technique to the color field in neural radiance fields. Here, we adopt the spherical harmonics (SH) representation proposed in Plenoxels [3] and [10] to perform color frequency decomposition. Specifically, in addition to implementing frequency-bounded grids for the density field, we apply frequency constraints to the Spherical Harmonics (SH) coefficients. At each level of detail, both the density and SH coefficients are queried at a specific resolution and then trilinearly interpolated to determine the density and SH coefficients of the target point. Subsequently, the computed SH coefficients are transformed into RGB values. It's important to note that, due to the linearity of spherical harmonics, constraining the frequency of SH coefficients directly imposes a constraint on the predicted color frequency. Further, both the density and color heads are trained using the previously proposed cascaded scheme. We compare the results of this method to a vanilla iNGP queried at target resolutions at test time. In Figure 1, we report superior performance to the baseline, highlighting the robustness of our method to aliasing effects.

Implementation. We train iNGP and our variation of it for 50K iterations with a batch size of 4096 rays. We evaluate our method on the NeRF Synthetic Dataset [8] at resolutions $\{64^2, 128^2\}$, while the input images are at 800^2 resolution. These evaluations are performed using grid resolutions $\{32^3, 64^3\}$ respectively, which were empirically determined to yield the best results. The SH coefficients of second order were used similar to [3] and [10]. Further, we note that we do not employ the extra skip connection from hash grid to the color MLP, as the color is being filtered.

2. Image filtering (cont'd)

We show additional quantitative and qualitative results on the DIV2K [1] dataset in Figure 2 and Table 1. Our method is trained on original images downsampled to 256^2 resolution and compared to BACON [7] and PNF [5], trained in a similar fashion. The networks are trained on 5K samples of the images. Evaluation is done at resolution 512^2, comparing to original images downscaled to this resolution to compute PSNR. Qualitative results are shown at resolutions $\{64^2, 128^2, 256^2\}$. As our method is compatible with any neural field, we demonstrate that, being based on an efficient backbone, it can reconstruct high-quality multi-resolution images while maintaining the same number of parameters as other techniques.



Scene	iNGP-64	BANF-64	iNGP-128	BANF-128
Chair	26.27	**30.25**	28.89	**32.21**
Drums	21.85	**24.33**	23.98	**24.86**
Ficus	24.06	**27.87**	**27.59**	27.45
Hotdog	28.33	**30.48**	31.52	**32.77**
Lego	23.64	**26.59**	27.22	**28.12**
Material	23.35	**24.89**	26.30	**26.48**
Mic	24.32	**26.77**	27.73	**30.36**
Ship	22.66	**24.30**	24.17	**24.52**
Average	24.31	**26.93**	27.17	**28.34**

Figure 1. **Color decomposition** – We show an application of our method in frequency decomposition of the color field in NeRFs.

Method	BACON	PNF	BANF
PSNR ↑	29.266	29.470	**30.455**
# Parameters ↓	0.268M	0.276M	**0.244M**

Table 1. **Image fitting** – quantitative results on DIV2K [1].

Resolution	64	128	256
Bilinear	25.650	25.364	30.455
Bicubic	**26.559**	26.380	**30.697**
Lanczos	26.050	**26.415**	30.210

Table 2. PSNR reported on interpolation with higher order kernels.

3. Higher order kernels

In our main results we used *linear* interpolation in our filtering algorithm. However, it was noted that this interpolation exhibits leakage, attributed to the sinc^2 Fourier transform of the filter. We further investigate how higher-order interpolation kernels such as Lanczos and Bicubic can further improve the filtering quality. In Figure 3, we evaluate results on the DIV2K [1] dataset, with images downsampled to 256^2. We then evaluate both quantitatively and qualitatively at resolutions $\{64^2, 128^2, 256^2\}$. We show in Table 2 that higher-order interpolations can help give an extra boost the performance of our method, while producing (perceptually) better filtered images Fig. 3.

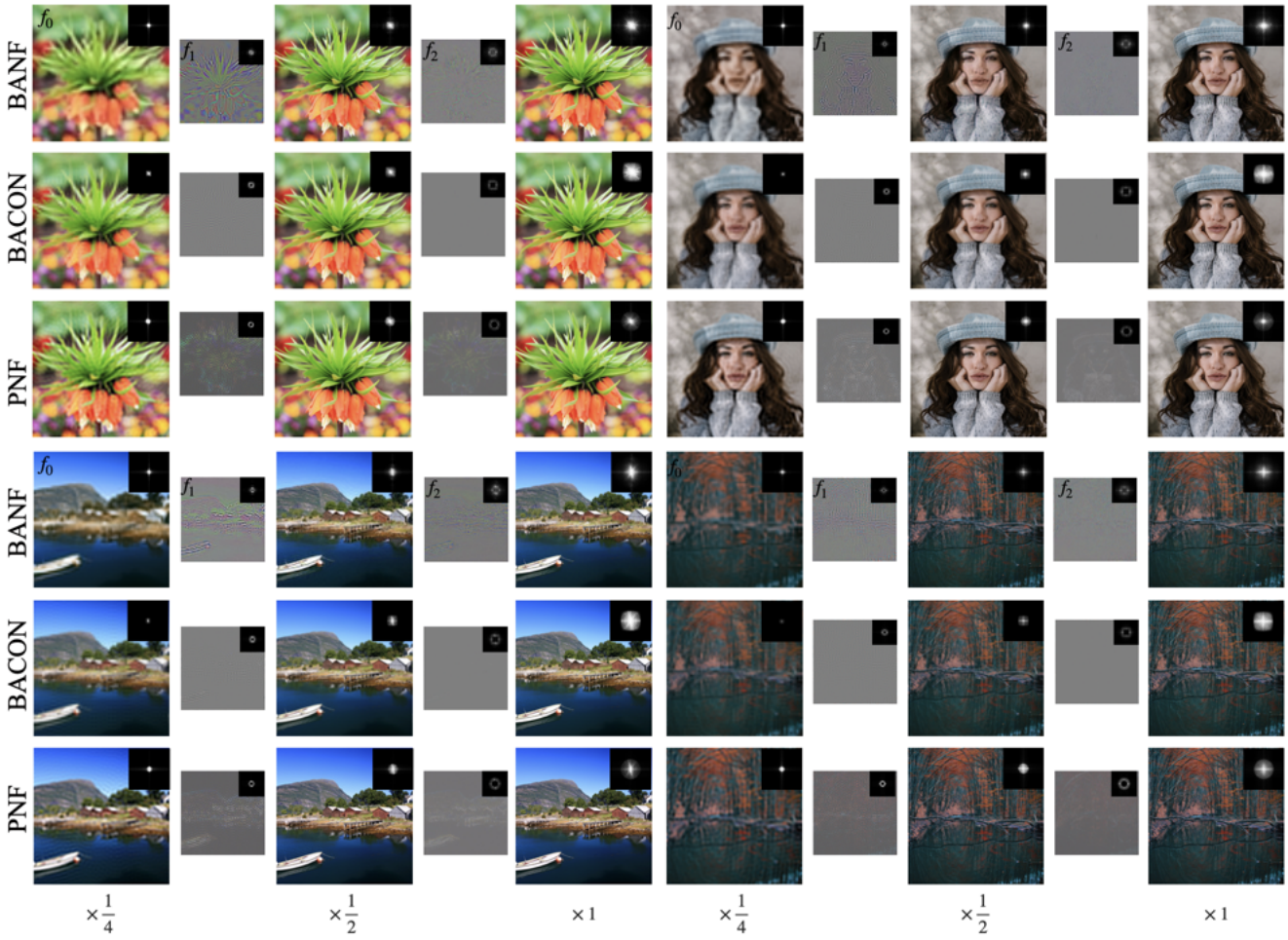


Figure 2. **Image Filtering –** Comparison of 2D filtering results on DIV2K [1] to BACON [7] and PNF [5]. Quantitative results are provided in Table 1.

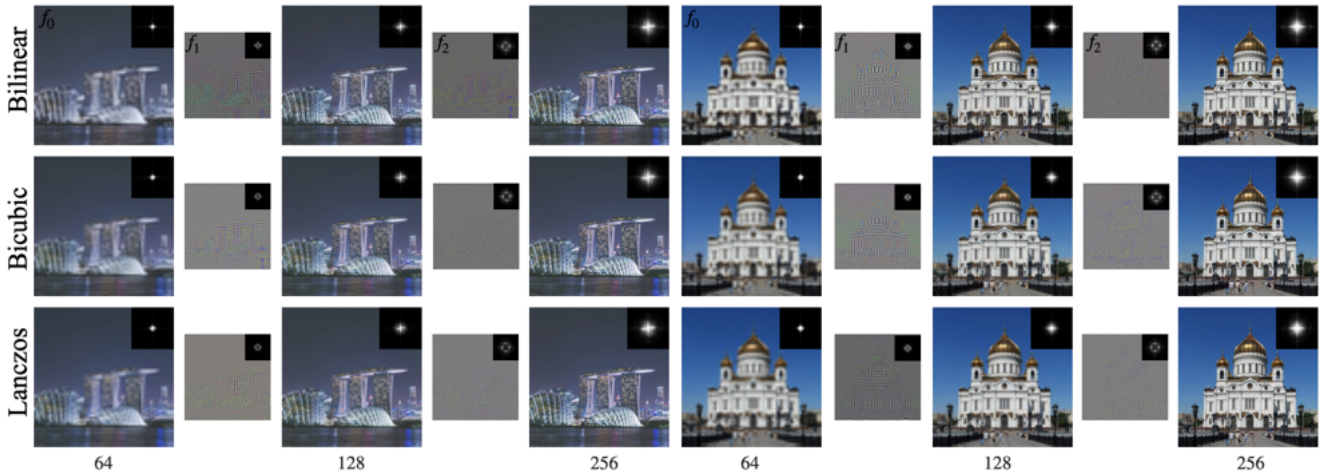


Figure 3. **Higher Order Kernels –** We evaluate the effect of higher order interpolation kernels on DIV2K [1].

4. Ablations

We validate our design choices in terms of Chamfer Distance (CD ↓) evaluated on the NeRF Synthetic dataset [2] averaged over all objects, and at $1/8\times$ scale.

Color MLP decomposition. When we train the color MLP output in a cascaded manner , performance drops CD=8.19→9.16. We attribute this to the color MLP relying on the SDF MLP output. As color and geometry exhibit different frequencies in terms of their signal, (*e.g.*, a highly complex shape with solid coloring), this reliance would be detrimental. For this ablation study we evaluate at $1/4\times$ scale, as the difference is less marked at $1/8\times$ scale.

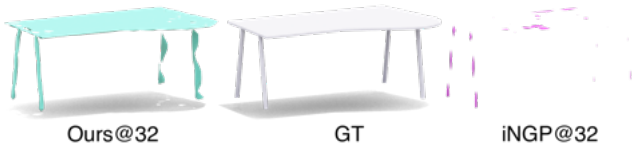
Color MLP Input. Removing the hash features input to the color MLP results in a performance drop from CD=11.4→13.6. Without the hash feature input, the color MLP relies only on the SDF features. At low scales, the SDF features will contain mostly low frequency content. We hypothesize that relying only on band-limited features reduces color reconstruction accuracy, resulting in inaccurate optimization of the the neural fields.

Resolution warmup. When we remove the resolution warmup from the coarsest level field, the performance drops CD=11.4→11.9. This is because starting optimization with lower resolution encourages smooth structures early on, leading to an easier optimization landscape as also shown in [4, 9].

5. Additional 3D reconstruction results

ModelNet10 data. We extend our evaluation of 3D shape fitting to include five models from the ModelNet10-tables dataset, assessed using the Chamfer-L2 metric (↓ $\cdot 10^{-2}$). Specifically, we examined 3D objects with the following identifiers: Table403, Table435, Table463, Table468, and Table479. Notably, our method demonstrates the capability to reconstruct thin structures even in low resolution, a feature not exhibited by the iNGP baseline:

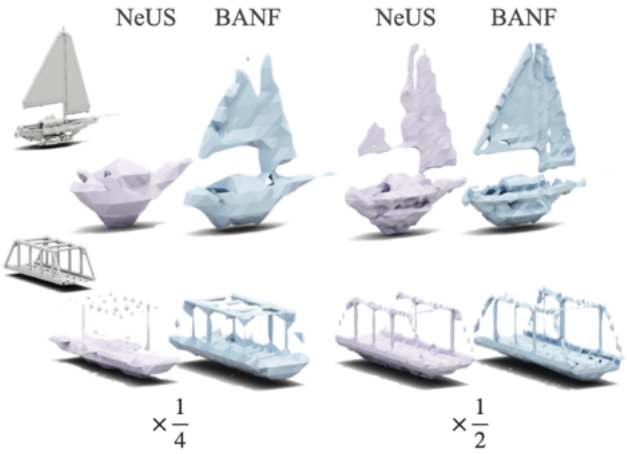
BANF			iNGP		
32	64	128	32	64	128
2.38	**2.08**	**0.31**	18.62	2.31	0.44



Additional Baselines for 3D Shape Fitting. In our comparative analysis, we include BACON [7] for 3D shape reconstruction. However, we observed that BACON exhibits slower training times and yields lower-quality results, particularly at higher decimation rates, in comparison to BANF. This observation is supported by the Chamfer-L2 metric (↓ $\cdot 10^{-2}$) evaluated on the four objects from the Stanford dataset used in BACON. Notably, for this series of experiments, we utilized the BACON codebase for data processing and point sampling pipelines.

BANF			iNGP			BACON		
32	64	128	32	64	128	32	64	128
1.35	**0.82**	**0.54**	1.98	0.91	0.61	2.65	1.00	0.58

Qualitative results for MobileBrick dataset. we also provide a qualitative comparison using the MobileBrick dataset [6]. Notably, our method demonstrates superior performance, particularly in handling low-resolution data and thin structures, when compared to NeUS.



6. Evaluation of geometry and color reconstruction on NeRF Synthetic Dataset

In the NeRF Synthetic Dataset [8], not all aspects of the geometries are observable from the training/validation cameras. This includes:

1) Internal structures that evade capture by any means, such as the stem of the plant inside the pot in the "Ficus" scene.

2) Some parts of the geometry that remain invisible from all cameras (both during training and testing). For example, the bottom of the chair in the "Chair" scene.

This inherent limitation results in a skewed assessment of metrics designed to measure the quality of reconstruction. To address this issue when computing the Chamfer Distance, we modify the process for densely sampled points on the surface of a mesh. Specifically, we filter the point cloud so that only points visible from at least one camera are retained. This filtering procedure is applied to both the ground truth and predicted meshes, ensuring a more accurate evaluation.

In inverse rendering, it is expected that the image reconstruction quality degrades as more emphasis is put on

(filtered) meshes, since regardless of mesh quality, neural fields can *cheat* to make renderings look good. Still, this degradation is minimal (<1 dB): average test PSNR for NeuS is 29.31 and 28.45 for BANF.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. *CVPRW*, 2017. 1, 2

[2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. *ICCV*, 2021. 3

[3] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. *CVPR*, 2022. 1

[4] Amir Hertz, Or Perel, Raja Giryes, Olga Sorkine-Hornung, and Daniel Cohen-Or. Sape: Spatially-adaptive progressive encoding for neural optimization. *NeurIPS*, 2021. 3

[5] Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. Panoptic Neural Fields: A Semantic Object-Aware Neural Scene Representation. *CVPR*, 2022. 1, 2

[6] Kejie Li, Jia-Wang Bian, Robert Castle, Philip H.S. Torr, and Victor Adrian Prisacariu. Mobilebrick: Building lego for 3d reconstruction on mobile devices. *CVPR*, 2023. 3

[7] David B. Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. *CVPR*, 2022. 1, 2, 3

[8] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*, 2020. 1, 3

[9] Jiawei Yang, Marco Pavone, and Yue Wang. Freenerf: Improving few-shot neural rendering with free frequency regularization. *CVPR*, 2023. 3

[10] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. *ICCV*, 2021. 1