Supplementary Materials for ECCV2024 Paper G^2fR: Frequency Regularization in Grid-based Feature Encoding Neural Radiance Fields

Shuxiang Xie[1,2], Shuyi Zhou[1], Ken Sakurada[2,3], Ryoichi Ishikawa[1], Masaki Onishi[2], and Takeshi Oishi[1]

[1] The University of Tokyo, Tokyo, Japan
[2] The National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
[3] Kyoto University, Kyoto, Japan

A Explanation and Extension of Equation 13

In the derivation of Equation 13, there is an important point that for two random index $t_1, t_2 \in \mathbb{N}$ and $t_1, t_2 \in [0, s-1]$, we have,

$$\Lambda((s-1)x - t_1) \cdot \Lambda((s-1)x - t_2) = 0, \quad \text{if } |t_1 - t_2| > 1. \tag{1}$$

Therefore, we can conclude the results as the summation of k-th power of triangular functions and the polynomials $P(x)$ related to neighboring functions.

In order to extend Equation 13 to more general cases, we first begin with discussing the situation where $L > 1$. From Equation 13 in the paper, we can have

$$\mathbf{z}^{(1)} = C + P(x) + \sum_{k=0}^{K} \sum_{t=0}^{s-1} h'_{(t;1),k} \cdot \Lambda^k((s-1)x - t). \tag{2}$$

Using the formulation of INR in Sec. 3.2, we have

$$\mathbf{z}^{(2)} = \rho^{(2)} \left(\mathbf{W}^{(2)} \mathbf{z}^{(1)} + \mathbf{b}^{(2)} \right). \tag{3}$$

Note that here $\mathbf{W}$ and $\mathbf{b}$ will not affect the order of $\Lambda(\cdot)$ in $\mathbf{z}^{(1)}$, therefore, we may simply model $\mathbf{z}^{(2)}$ as,

$$\mathbf{z}^{(2)} = \rho^{(2)}(\hat{\mathbf{z}}^{(1)}) = \rho^{(2)} \left(C + P(x) + \sum_{k=0}^{K} \sum_{t=0}^{s-1} h''_{(t;1),k} \cdot \Lambda^k((s-1)x - t) \right), \tag{4}$$

Where C can be arbitrary value to absorb the constant part. Keep using the polynomial approximation such that $\rho^{(2)}(z) = \sum_{k=0}^{K} \alpha_k z^k$ [7], then we can obtain

$$\mathbf{z}^{(2)} = \sum_{p=0}^{P} \alpha_p \left(C + P(x) + \sum_{k=0}^{K} \sum_{t=0}^{s-1} h''_{(t;1),k} \cdot \Lambda^k((s-1)x - t) \right)^p. \tag{5}$$

Recall multinomial theorem that for positive integer m and non-negative integer n,

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n} \gamma_{(k_1,k_2,\ldots k_m)} \cdot \prod_{i=1}^{m} x_i^{k_i}, \tag{6}$$

where $\gamma_{(k_1,k_2,\ldots k_m)} = \frac{n!}{k_1!k_2!\ldots k_m!}$. Consider the inner part in Eq. (5),

$$(\hat{\mathbf{z}}^{(1)})^p = \left(C + P(x) + \sum_{k=0}^{K} \sum_{t=0}^{s-1} h''_{(t;1),k} \cdot \Lambda^k((s-1)x - t) \right)^p \tag{7}$$

$$= \widehat{C} + \widehat{P}(x) + \sum_{p_1+p_2+\ldots p_k=p} \eta_{\{p_k\}} \cdot \prod_{i=0}^{p} \left[\Lambda^i((s-1)x - t) \right]^{p_i}. \tag{8}$$

Notice that here $\eta_{\{p_k\}}$ is calculated by multiplying some value to $\gamma_{(p_1,p_2,\ldots p_k)}$, and $\widehat{C}$ stands for some constant, $\widehat{P}(x)$ means the polynomials related to neighboring functions. We can see that $\mathbf{z}^{(2)}$ is still in a similar form of Eq. (2). Given the condition that the coefficients of the polynomial used in the approximation usually decay very rapidly as the order k increases, the rest analysis become the same as we discussed in the paper. Therefore, adding network depths will not affect our conclusion.

Then we consider the case of multi-resolution. Let $L = 1, M = 2$, and $s_1 = s, s_2 = 2s$. The encoded function $\Psi(\mathbf{x})$ becomes

$$\Psi(x) = \left[\sum_{t=0}^{s-1} h_{(t;1)} \cdot \Lambda((s-1)x - t), \sum_{t=0}^{2s-1} h_{(t;2)} \cdot \Lambda((2s-1)x - t) \right]^\top. \tag{9}$$

When $\mathrm{G}^2f\mathrm{R}$ or any other binary masking techniques as in [2,3,5,6] are applied, we just need to multiply a weighting coefficient for each element in $\Psi(\mathbf{x})$. Then we have

$$\mathbf{z}^{(1)} = \rho^{(1)} \left(\mathbf{W}^{(1)} \cdot \begin{bmatrix} \sum_{t=0}^{s-1} h_{(t;1)} \cdot \Lambda((s-1)x - t) \\ \sum_{t=0}^{2s-1} h_{(t;2)} \cdot \Lambda((2s-1)x - t) \end{bmatrix} + \mathbf{b}^{(1)} \right). \tag{10}$$

Let $\mathbf{W}^{(1)} = [W_1, W_2]^\top$, we may further write $\mathbf{z}^{(1)}$ as

$$\mathbf{z}^{(1)} = \rho^{(1)} \left(\sum_{i=1}^{2} W_i \cdot \sum_{t=0}^{i \cdot s-1} h_{(t;i)} \cdot \Lambda((i \cdot s - 1)x - t) + \mathbf{b}^{(1)} \right), \tag{11}$$

where the inner part is a linear combination of triangular pulse functions with resolution of s and $2s$. We can still use multinomial theorem to expand $\rho^{(1)}(\cdot)$, which will finally lead to

$$\begin{aligned} \mathbf{z}^{(1)} \approx & \sum_{k=0}^{K} \sum_{t=0}^{s-1} u_{(t;k)} \cdot \Lambda^k((s-1)x - t) + \sum_{k=0}^{K} \sum_{t=0}^{2s-1} v_{(t;k)} \cdot \Lambda^k((2s-1)x - t) \\ & + \sum_{k_1+k_2 \le K} w_{(t;k_1,k_2)} \Lambda^{k_1}((s-1)x - t) \cdot \Lambda^{k_2}((2s-1)x - t). \end{aligned} \tag{12}$$

We can see that the main part of $\mathbf{z}^{(1)}$ is still a combination of product of triangular functions. Note that here we omit constant and neighboring polynomials for simplicity.

The Fourier transform of $\mathbf{z}^{(1)}$ becomes complicated. Here we only show the results of $\Lambda(x)\cdot(2x+1)$, which is the basic scenario of $\sum_{k_1+k_2\leq K} w_{(t;k_1,k_2)}\Lambda^{k_1}((s-1)x-t)\cdot\Lambda^{k_2}((2s-1)x-t)$ part in $\mathbf{z}^{(1)}$. After integral we can obtain,

$$\begin{aligned}\mathcal{F}[\Lambda(x)\cdot\Lambda(2x+1)](\omega) &= \int_{-1}^{-\frac{1}{2}}(x+1)^2e^{-j\omega x}\,\mathrm{d}x + \int_{-\frac{1}{2}}^{0}(-x^2-x)e^{-j\omega x}\,\mathrm{d}x \\ &= \frac{1}{\omega^2}(1-e^{\frac{1}{2}j\omega}) + \frac{2}{\omega^3}\cdot\frac{(e^{\frac{1}{2}j\omega}+1)^2}{j}.\end{aligned} \tag{13}$$

Similar analysis corresponding to the bandwidth can be conducted as shown in the paper.

To conclude, extending Equation 13 to additional layers or dimensions increases the complexity of the analysis, but the underlying principles and concepts remain the same.

A.1 Calculation of Fourier Transform

Recall that the triangular function is defined as $\Lambda(x)$,

$$\Lambda(x) = \max(0, 1-|x|). \tag{14}$$

We can calculate the Fourier transform by the following steps:

$$\begin{aligned}\mathcal{F}[\Lambda(x)](\omega) &= \int_{-\infty}^{\infty}\Lambda(x)e^{-j\omega x}\,\mathrm{d}x \\ &= \int_{-1}^{0}(x+1)e^{-j\omega x}\,\mathrm{d}x + \int_{0}^{1}(-x+1)e^{-j\omega x}\,\mathrm{d}x.\end{aligned} \tag{15}$$

Let $k=-j\omega$, we can obtain:

$$\begin{aligned}\mathcal{F}[\Lambda(x)](\omega) &= \left.\frac{e^{kx}(kx+k-1)}{k^2}\right|_{-1}^{0} + \left.\frac{e^{kx}(-kx+k+1)}{k^2}\right|_{0}^{1} \\ &= \frac{1}{k^2}(e^{-k}+e^{k}-2) \\ &= \frac{1}{(-j\omega)^2}(e^{j\omega}+e^{-j\omega}-2).\end{aligned} \tag{16}$$

Since $e^{j\omega}+e^{-j\omega}=2\cos\omega$, we can see that $\mathcal{F}[\Lambda(x)](\omega)=\frac{2-2\cos\omega}{\omega^2}$.

The calculation of Fourier transform of $\Lambda^2(x)$ is also similar. Let $k = -j\omega$, we calculate the Fourier transform as,

$$
\begin{aligned}
\mathcal{F}[\Lambda^2(x)](\omega) &= \int_{-\infty}^{\infty} \Lambda^2(x) e^{-j\omega x}\,\mathrm{d}x \\
&= \int_{-1}^{0} (x+1)^2 e^{-j\omega x}\,\mathrm{d}x + \int_{0}^{1} (-x+1)^2 e^{-j\omega x}\,\mathrm{d}x \\
&= \frac{1}{k^3}\left[e^{kx}(k^2(x+1)^2 - 2k(x+1)+2)\right]\Big|_{-1}^{0} + \\
&\quad \frac{1}{k^3}\left[e^{kx}(k^2(x-1)^2 - 2k(x-1)+2)\right]\Big|_{0}^{1} \\
&= \frac{1}{k^3}(-2e^{-k} + 2e^{k} - 4k) \\
&= \frac{1}{\omega^3}(\frac{-2e^{j\omega} + 2e^{-j\omega}}{j} + 4\omega).
\end{aligned} \tag{17}
$$

Since $\sin\omega = \frac{e^{j\omega} - e^{-j\omega}}{2j}$, we can see that $\mathcal{F}[\Lambda^2(x)](\omega) = \frac{4\omega - 4\sin\omega}{\omega^3}$.

A.2 Bandwidth of Single Pulse

In this section, we are going to prove $\exists \eta_1, \eta_2 \in \mathbb{R}$, and $\eta_1, \eta_2 > 0$ that make $\mathcal{F}[\Lambda(x)](\omega) \leq \eta_1 \cdot \frac{1}{\omega^2}$ and $\mathcal{F}[\Lambda^2(x)](\omega) < \eta_2 \cdot \frac{1}{\omega^2}$.

For $\mathcal{F}[\Lambda(x)](\omega)$, it is very obvious that $\cos\omega \in [-1, 1]$, then $2 - 2\cos\omega \in [0, 4]$. Therefore, we just to make $\eta_1 \geq 4$ then we shall ensure $\mathcal{F}[\Lambda(x)](\omega) \leq \eta_1 \cdot \frac{1}{\omega^2}$.

For $\mathcal{F}[\Lambda^2(x)](\omega)$, we begin by considering the minimum and maximum value of $g(\omega) = \frac{\sin\omega}{\omega}$. We have $\lim_{\omega\to 0} = 1$ and $\lim_{\omega\to\infty} = 0$. Then we consider,

$$g'(\omega) = \frac{\omega\cos\omega - \sin\omega}{\omega^2}. \tag{18}$$

Note that $\frac{\mathrm{d}}{\mathrm{d}\omega}(\omega\cos\omega - \sin\omega) = -\omega\sin\omega < 0$, when $\omega \in (0, 1]$. Considering $\omega\cos\omega - \sin\omega = 0$ when $\omega = 0$, we can have $g'(\omega) < 0$ for $\omega \in (0, 1]$. Then we may conclude that the minimum value of $g(\omega)$ when $\omega \in (0, 1]$ is $g(1) = \sin(1)$ when $\omega = 1$. Situations are the same for $\omega \in [-1, 0)$. As for the case where $\omega > 1$, it is obvious that $g(\omega) \in (-1, 1)$ since $\sin(\omega)$ is bounded by -1 and 1. Considering $g(1) = \sin(1) > -1$, then we can have $g(\omega) > -1$.

Since $\mathcal{F}[\Lambda^2(x)](\omega) = \frac{4\omega - 4\sin(\omega)}{\omega^3}$, it is easy to see that

$$\mathcal{F}[\Lambda^2(x)](\omega) = \frac{1}{\omega^2}\cdot(4 - 4g(\omega)) < \frac{1}{\omega^2}\cdot 8. \tag{19}$$

Then η_2 just need to take the value of 8 to ensure $\mathcal{F}[\Lambda^2(x)](\omega) < \eta_2 \cdot \frac{1}{\omega^2}$.

B Experimental Results

To remind, we test the following methods for the comparative experiments: BARF [4]: frequency regularization in PE NeRF; L2G-NeRF [1]: local-to-global camera pose representation; RCPR[4] [2]: smooth interpolation instead of linear; CamP [5]: preconditioners for camera pose and intrinsic parameters.

Scene	Camera pose registration: Rotation error (°) ↓						Camera pose registration: Translation error ↓ ($\times 10^2$)						View synthesis quality: PSNR ↑					
	BARF	L2G-NeRF	RCPR	CamP	NGP $+G^2 f$R	Mtrf $+G^2 f$R	BARF	L2G-NeRF	RCPR	CamP	NGP $+G^2 f$R	Mtrf $+G^2 f$R	BARF	L2G-NeRF	RCPR	CamP	NGP $+G^2 f$R	Mtrf $+G^2 f$R
Chair	0.099	0.117	0.202	0.232	0.185	0.105	0.360	0.449	1.279	1.344	0.689	0.636	31.08	31.01	30.77	38.05	35.01	32.17
Drums	0.045	0.064	0.418	0.449	0.028	0.031	0.276	0.340	4.840	2.575	0.135	0.096	23.90	23.79	19.03	23.17	25.36	23.56
Ficus	0.075	0.180	0.230	1.187	0.043	0.069	0.444	0.849	1.396	3.764	0.204	0.236	26.29	26.21	26.29	18.86	26.98	26.65
Hotdog	0.246	0.253	0.830	3.105	0.128	0.682	1.204	1.273	0.309	1.437	0.745	1.554	34.58	34.59	33.02	32.83	36.89	34.79
Lego	0.076	0.10	0.064	0.164	0.043	0.039	0.300	0.433	0.280	0.884	0.143	0.148	28.32	27.97	32.21	32.33	33.55	29.71
Materials	0.837	0.051	1.156	0.598	0.378	0.028	2.729	0.268	2.789	3.764	2.042	0.094	27.86	27.71	25.09	14.78	27.70	28.65
Mic	0.080	0.093	0.933	0.504	0.041	0.040	0.402	0.372	1.004	1.312	0.124	0.129	31.18	31.03	29.89	33.82	34.32	32.80
Ship	0.074	0.179	0.896	0.431	0.099	0.108	0.341	0.746	1.178	1.344	0.605	0.830	27.54	27.44	30.64	29.84	30.19	27.03

Table 1: Comparative results of the camera pose optimization experiments using different methods. The data inside row *Lego* is the data on the paper.

Scene	View synthesis quality: PSNR ↑		SSIM ↑		LPIPS ↓		Scene	View synthesis quality: PSNR ↑		SSIM ↑		LPIPS ↓	
	Mtrf w/	Mtrf w/o	Mtrf w/	Mtrf w/o	Mtrf w/	Mtrf w/o		Mtrf w/	Mtrf w/o	Mtrf w/	Mtrf w/o	Mtrf w/	Mtrf w/o
Chair	28.63	27.91	0.949	0.946	0.039	0.039	Bicycle	15.17	11.61	0.278	0.229	0.868	0.880
Drums	17.18	13.69	0.781	0.694	0.216	0.430	Bonsai	12.99	9.95	0.438	0.393	0.687	0.680
Ficus	20.18	21.93	0.865	0.892	0.099	0.072	Counter	15.21	15.73	0.437	0.449	0.657	0.601
Hotdog	27.77	27.15	0.939	0.926	0.065	0.083	Garden	20.33	19.87	0.344	0.368	0.695	0.680
Lego	25.14	25.62	0.899	0.916	0.068	0.063	Kitchen	13.22	12.13	0.385	0.336	0.695	0.680
Materials	22.78	21.73	0.887	0.867	0.040	0.068	Room	19.22	19.04	0.595	0.621	0.493	0.503
Mic	20.83	30.01	0.922	0.971	0.090	0.024	–						
Ship	17.62	19.57	0.688	0.732	0.259	0.222	–						

Table 2: Quantitative results of few-shot reconstruction on both the NeRF-synthetic and MipNeRF-360 datasets utilizing Mtrf.

References

1. Chen, Y., Chen, X., Wang, X., Zhang, Q., Guo, Y., Shan, Y., Wang, F.: Local-to-global registration for bundle-adjusting neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8264–8273 (2023)
2. Heo, H., Kim, T., Lee, J., Lee, J., Kim, S., Kim, H.J., Kim, J.H.: Robust camera pose refinement for multi-resolution hash encoding. arXiv preprint arXiv:2302.01571 (2023)

[4] The codes of [2] are not publicly available when we write this paper. In this study, we use our own implementation according to the description in the paper.

3. Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H.: Neuralangelo: High-fidelity neural surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8456–8465 (2023)
4. Lin, C.H., Ma, W.C., Torralba, A., Lucey, S.: Barf: Bundle-adjusting neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5741–5751 (2021)
5. Park, K., Henzler, P., Mildenhall, B., Barron, J.T., Martin-Brualla, R.: Camp: Camera preconditioning for neural radiance fields. ACM Transactions on Graphics (TOG) **42**(6), 1–11 (2023)
6. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3295–3306 (2023)
7. Yüce, G., Ortiz-Jiménez, G., Besbinar, B., Frossard, P.: A structured dictionary perspective on implicit neural representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19228–19238 (2022)