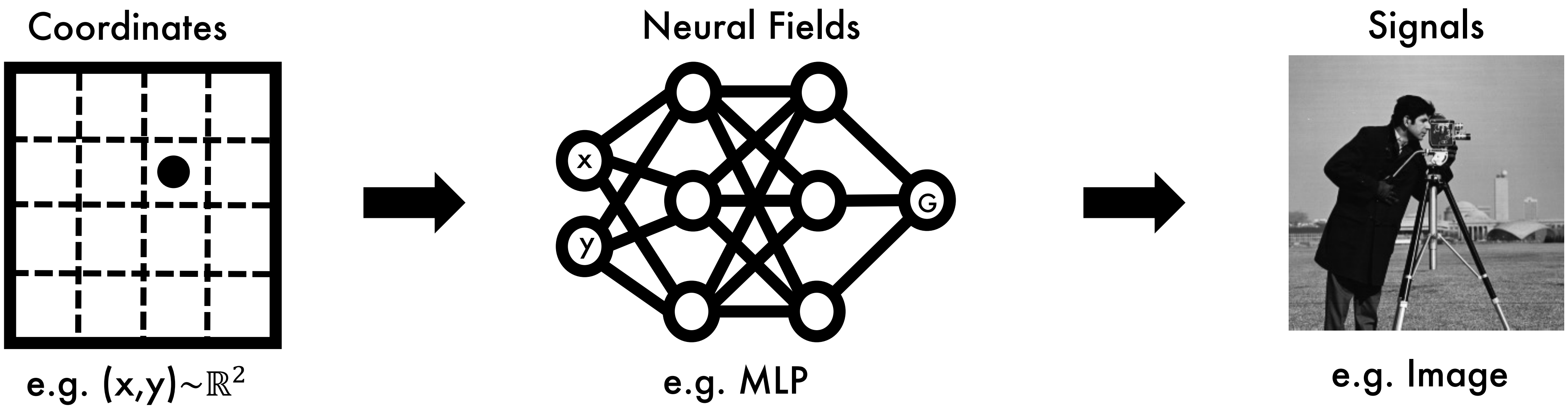
Fast Training of Sinusodial Neural Fields via Scaling Initialization

Taesun Yeom*   Sangyoon Lee*   Jaeho Lee

Efficient Learning Lab., Pohang University of Science and Technology (POSTECH)

ICLR 2025

Background: (Sinusoidal) Neural Fields



- Neural Fields (NFs) are neural nets designed to represent (various types of) signals.
- Sinusoidal Neural Fields (SNFs): NFs with sinusoidal activation functions.
  - ▶ The most widely used architectures for NFs.

$$f(X;\mathbf{W}) = W^{(l)}Z^{(l-1)} + b^{(l)}, \qquad Z^{(i)} = \sigma_i(W^{(i)}Z^{(i-1)} + b^{(i)}), \quad i \in [l-1]$$

$$\sigma_1(x) = \sin(\omega_0 \cdot x) \qquad \sigma_i(x) = \sin(\omega_h \cdot x)$$

(L-layer SNF)

Motivations

Problems.

- Training NFs is time-consuming, as it involves a per-datum fitting process.
  - Due to the inherent high-frequency components in the natural signals, known as spectral bias.

Motivations

Problems.

- Training NFs is time-consuming, as it involves a per-datum fitting process.
    - Due to the inherent high-frequency components in the natural signals, known as spectral bias.

Our Question.

Can we train SNFs with…

Faster speed + ***Good generalization*** + ***No additional cost*** ?

Motivations

Problems.

- Training NFs is time-consuming, as it involves a per-datum fitting process.
    - Due to the inherent high-frequency components in the natural signals, known as spectral bias.

Our Question.

Can we train SNFs with…

Faster speed + ***Good generalization*** + ***No additional cost*** *?*

Previous works.

- Utilize specialized activation functions. (e.g. Gabor wavelets, Gaussian function, Sinc functions, etc.)
- Develop novel encoding scheme for inputs. (e.g. Fourier features, Instant-NGP, Tri-plane, etc.)
- A limited number of works focus on SNF-based methods.

Our approach

- We revisit the principled initialization scheme [1] of SNFs.

Q. : Does the standard initialization scheme lead to suboptimal results?

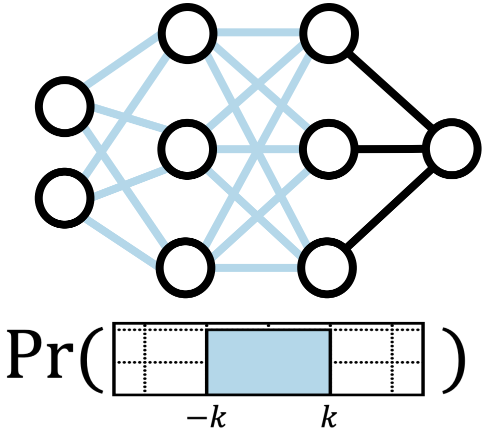
A. Yes…

Q. Then how can we change it?

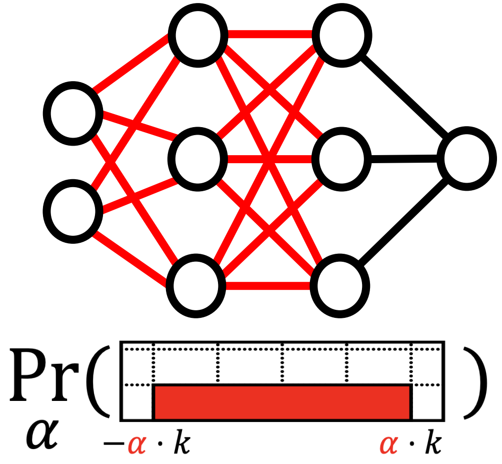
A. Simply scaling the initialization distribution leads to significant performance gain!

[1] Sitzmann et al., *Implicit Neural Representations with Periodic Activation Functions, NeurIPS 2020*

Weight Scaling Initialization



Standard Init. [1]



Weight Scaling (WS) Init. (ours)

First layer.

$$w_{j,k}^{(1)} \overset{\text{i.i.d.}}{\sim} \text{Unif}\left(-\frac{1}{d_0}, \frac{1}{d_0}\right) \qquad w_{j,k}^{(1)} \overset{\text{i.i.d.}}{\sim} \text{Unif}\left(-\frac{\alpha}{d_0}, \frac{\alpha}{d_0}\right)$$

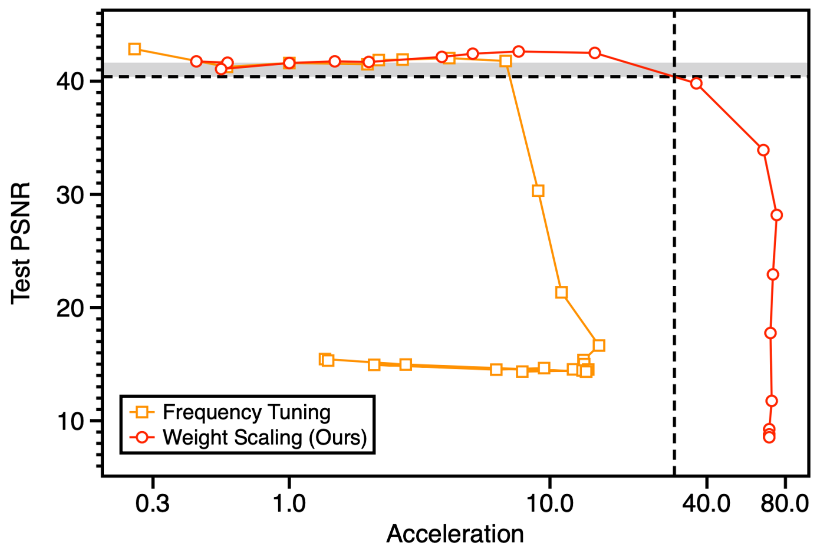
Hidden layer.

$$w_{j,k}^{(i)} \overset{\text{i.i.d.}}{\sim} \text{Unif}\left(-\frac{\sqrt{6}}{\omega_h\sqrt{d_{i-1}}}, \frac{\sqrt{6}}{\omega_h\sqrt{d_{i-1}}}\right) \qquad w_{j,k}^{(i)} \overset{\text{i.i.d.}}{\sim} \text{Unif}\left(-\frac{\alpha\sqrt{6}}{\omega_h\sqrt{d_{i-1}}}, \frac{\alpha\sqrt{6}}{\omega_h\sqrt{d_{i-1}}}\right)$$

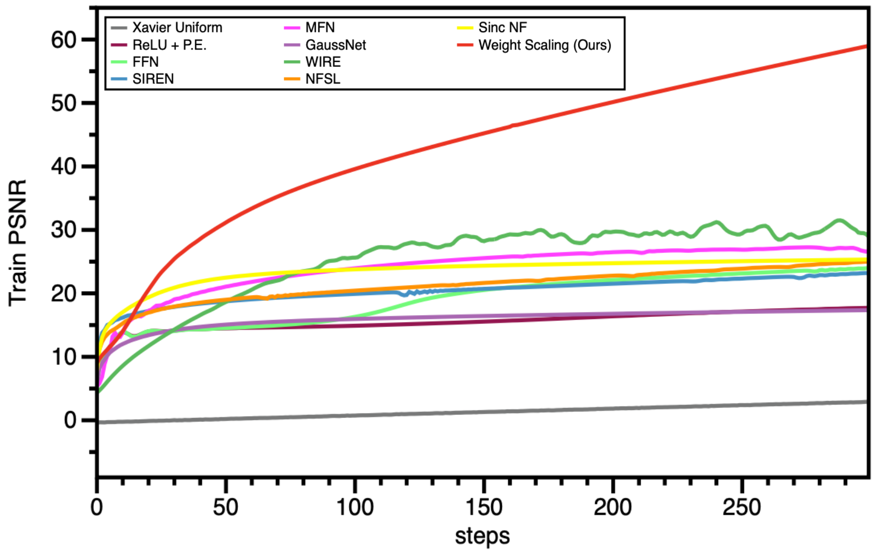
- In **Proposition 1**, we prove that 'WS initialization' does not harm signal propagation property.

[1] Sitzmann et al., *Implicit Neural Representations with Periodic Activation Functions*, NeurIPS 2020

Benefits of WS initialization



(A)



(B)

(A) Pareto plot. Acceleration vs. Generalization (Test PSNR)

- WS-SNF is superior to frequency scaling. (i.e. scaling first layer frequency gain)
- **Note.** Frequency scaling does not amplify inner frequencies.

(B) Learning curve. Compared to SOTA NF architectures

- WS-SNF outperforms SOTA architectures.

Understanding WS Initialization

Additional details and exact derivations can be found in Appendix.

1. Initial Functional Perspective

- Consider a reparameterized simple three-layer SNF, i.e. $f(x;\mathbf{W}) = 2W^{(3)} \sum_{\ell \in \mathbb{Z}_{odd}} J_\ell(W^{(2)}) \sin(\ell W^{(1)} x)$
- The relative harmonic is bounded w.r.t. Bessel function of the first kind.

$$\frac{(W^{(2)})^2}{(2\ell+2)(2\ell+4)} < \frac{J_{\ell+2}(W^{(2)})}{J_\ell(W^{(2)})} < \frac{(W^{(2)})^2}{(2\ell+1)(2\ell+3)}$$

- Scaling hidden layer's weights (i.e. $W^{(2)} \mapsto \alpha \cdot W^{(2)}$), which is the unique characteristic of WS init., leads to $J_{\ell+2}/J_\ell \propto \alpha^2$

Hidden layer scaling induces an increase in relative harmonics.

---

2. Gradient Perspective

- In a simple two-layer SNF, $f(x) = w^{(2)}\sigma(w^{(1)}x)$, the gradients are

$$\nabla_{\mathbf{w}} f = \left(w^{(2)} x \sigma'(w^{(1)}x), \sigma(w^{(1)}x)\right) \text{ where } \begin{cases} \left|w^{(2)} x \sigma'(w^{(1)}x)\right| \propto \left|w^{(2)}\right| = \alpha\, w^{(2)} \\ \left|\sigma(w^{(1)}x)\right| \in [-1,1] \end{cases}$$

- Thus, the gradients of the earlier layers are relatively larger than those of the later layers, due to the properties of the sine.

Weight scaling induces unbalanced gradients across layers.

Understanding WS Initialization

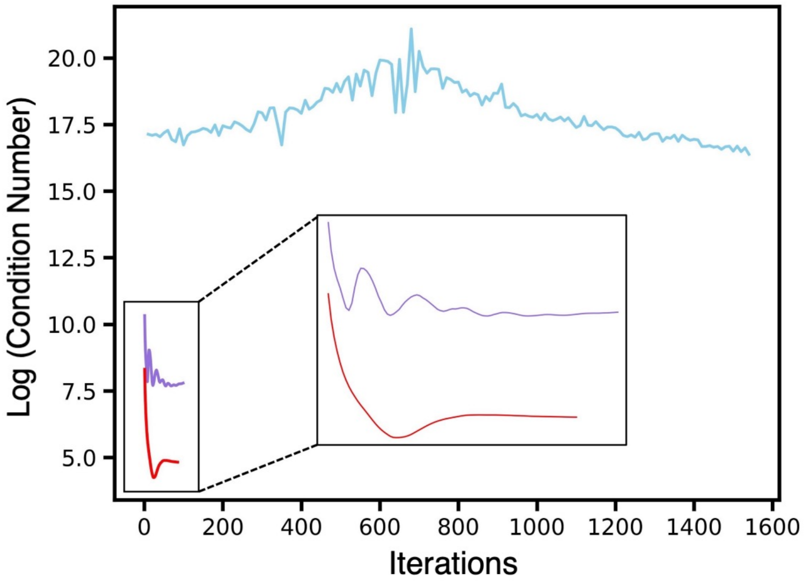
3. Neural Tangent Kernel Analysis

With some assumptions, neural net training is akin to kernel learning, called Neural Tangent Kernel (NTK).

The dynamics of typical neural network training are intractable, due to highly nonlinear evolution of params.

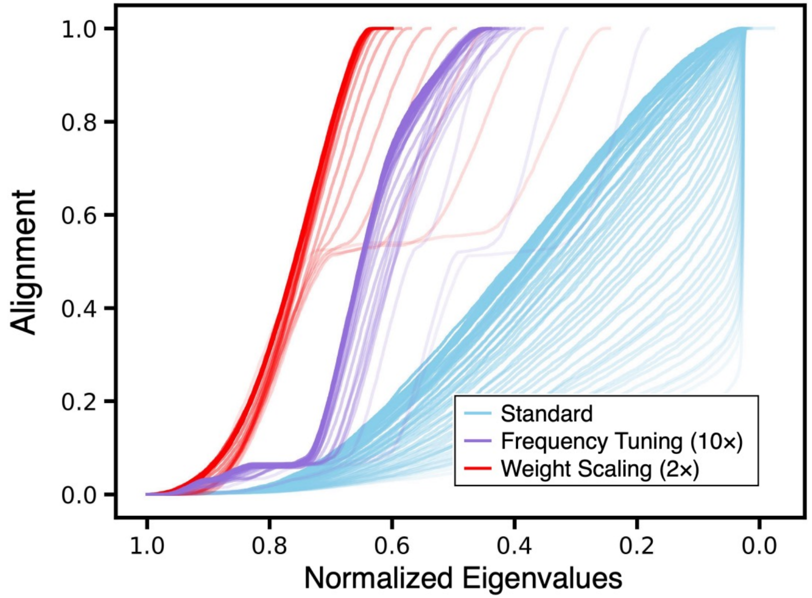
▶ Measures from empirical NTK (eNTK) serve as proxies for learning dynamics.

Condition number (CN) of eNTK



WS-SNF enjoys well-conditioned optimization trajectory.

Kernel – Task Alignment



WS-SNF ‘well-behaved’ throughout training.

Main Experiments

	Activation	Image (PSNR)		Occ. Field (IoU)	Spherical (PSNR)	Audio (PSNR)
		KODAK	DIV2K	Lucy	ERA5	Bach
Xavier Uniform (Glorot & Bengio, 2010)	Sinusoidal	0.46±0.10	0.39±0.10	0.0000±0.0000	4.11±0.66	7.77±0.20
ReLU + P.E. (Mildenhall et al., 2020)	ReLU	18.60±0.08	16.72±0.08	0.9896±0.0003	33.30±0.54	24.98±0.19
FFN (Tancik et al., 2020)	ReLU	20.52±0.60	19.81±0.48	0.9843±0.0020	38.69±0.27	16.66±0.28
SIREN init. (Sitzmann et al., 2020b)	Sinusoidal	24.58±0.05	22.86±0.06	0.9925±0.0001	38.72±0.07	37.37±3.11
GaussNet (Ramasinghe & Lucey, 2022)	Gaussian	21.94±2.48	19.22±0.14	0.9914±0.0005	38.56±0.51	27.47±2.10
MFN (Fathony et al., 2021)	Wavelet	28.54±0.12	26.42±0.10	0.9847±0.0003	36.89±0.80	16.16±0.05
WIRE (Saragadam et al., 2023)	Wavelet	28.94±0.21	28.20±0.13	0.9912±0.0005	31.27±0.53	16.83±1.85
NFSL (Saratchandran et al., 2024b)	Sinusoidal	24.93±0.07	23.39±0.09	0.9925±0.0001	38.92±0.07	37.17±2.88
Sinc NF (Saratchandran et al., 2024c)	Sinc	27.73±0.27	26.42±0.24	0.9936±0.0002	36.15±0.51	23.03±0.40
Weight scaling (ours)	Sinusoidal	**42.83±0.35**	**42.03±0.41**	**0.9941±0.0002**	**45.28±0.03**	**45.04±3.23**

- Qualitative results and results of additional experiments on NeRFs, PDEs, and weight space learning can be found in Appendix.

WS-SNF outperforms baselines (i.e. recent NF methods) across all tasks.

Summary

- We propose a simple yet effective initialization scheme for sinusoidal neural fields using a scaling factor, coined weight scaling initialization.

- Weight scaling initialization affects both the initial functionals and the gradients, both of which are beneficial.

- We provide both empirical and theoretical analyses for understanding weight scaling initializations.

- WS-SNF significantly outperforms recent NF methods across various tasks.

Thank you for your attention!

Feel free to contact me if there are any questions.

▶ **Contact**. `tsyeom@postech.ac.kr`

At conference, swing by our poster for more details!

▶ **Session**. Poster session 1 (Thu, April 24)

Also, consider visiting our group's website for more interesting work!

▶ **Efficient Learning Lab**. (a.k.a. EffL) @ POSTECH (Hompage: effl.postech.ac.kr)

Paper QR

Poster QR