Low-Rank Adaptation of Neural Fields - Supplemental Materials

ANH TRUONG, Massachusetts Institute of Technology, USA
AHMED H. MAHMOUD, Massachusetts Institute of Technology, USA
MINA KONAKOVIĆ LUKOVIĆ, Massachusetts Institute of Technology, USA
JUSTIN SOLOMON, Massachusetts Institute of Technology, USA

A Low-rank Structure of Fine-tuned Weights

As motivation for the use of LoRA in our setting, we conducted the following experiment which reveals that the final weight updates obtained from fully fine-tuning an image neural field often have low rank across different layers.

Suppose $f_\theta : \mathbb{R}^2 \to \mathbb{R}^3$ is a neural field with m hidden layers trained to regress the RGB colors of a 2D image $\mathcal{D}$. Denote the hidden layer weights of f_θ by $W^i_{\text{base}} \in \mathbb{R}^{n\times n}$ where $i = 1, \dots, m$. We perform a minor edit to $\mathcal{D}$ to obtain a new image $\mathcal{D}'$. We then fine-tune all the weights of f_θ to regress $\mathcal{D}'$, yielding a new neural field $f_{\theta'}$ with weights $W^i_{\text{fine-tuned}} \in \mathbb{R}^{n\times n}$. By solving the rank-constrained optimization problem

$$\min_{\Delta^i \in \mathbb{R}^{n\times n}} \mathop{\mathbb{E}}_{x} \left[\left\| ((W^i_{\text{base}} + \Delta^i) - W^i_{\text{fine-tuned}})x \right\|_2^2 \right] \quad (1)$$
$$s.t. \quad \text{rank}(\Delta^i) \le k$$

for each layer i, where $k \in \mathbb{N}$ controls the maximum rank of Δ^i, and $x \in \mathbb{R}^n$ is sampled from the empirical output distribution of the previous layer $(i-1)$ of $f_{\theta'}$, we find that **a low-rank factorization of Δ^i approximates the full update $W^i_{\text{base}} - W^i_{\text{fine-tuned}}$ with minimal error** while accounting for the layer's input distribution.

For any choice of maximum rank k, this problem has a closed-form solution (normalized optimal values shown in Figure 2 of the paper for three different examples). This is because the distribution of the input queries x_0—representing normalized image coordinates—is assumed to be Unif $\left([0,1]^2\right)$. By pushing input samples through the network, we are able to obtain an empirical distribution of each layer's output activations. In the following, we derive the closed-form optimal value of (1) by transforming it into a problem where the rank-k truncated SVD of a matrix is optimal.

Define $D^i \coloneqq W^i_{\text{base}} - W^i_{\text{fine-tuned}}$. The objective function of (1) can be rewritten as

$$\mathop{\mathbb{E}}_{x} \left[\left\| (D^i + \Delta^i)x \right\|_2^2 \right]$$
$$= \mathop{\mathbb{E}}_{x} \left[x^T(\Delta^i)^T\Delta^i x + 2x^T(D^i)^T(\Delta^i)x + x^T(D^i)^TD^i x \right]$$

The final term is constant with respect to the decision variable Δ^i, so we can safely ignore it for the purposes of solving problem 1:

$$= \text{const} + \mathop{\mathbb{E}}_{x} \left[x^T(\Delta^i)^T\Delta^i x + 2x^T(D^i)^T(\Delta^i)x \right]$$
$$= \text{const} + \mathop{\mathbb{E}}_{x} \left[x^T(\Delta^i)^T\Delta^i x \right] + 2\mathop{\mathbb{E}}_{x} \left[x^T(D^i)^T(\Delta^i)x \right]$$
$$= \text{const} + \mathop{\mathbb{E}}_{x} \left[\text{Tr}(xx^T(\Delta^i)^T\Delta^i) \right] + 2\mathop{\mathbb{E}}_{x} \left[\text{Tr}(xx^T(D^i)^T(\Delta^i)) \right]$$
$$= \text{const} + \text{Tr}\left(\mathop{\mathbb{E}}_{x} \left[xx^T \right] (\Delta^i)^T\Delta^i \right) + 2\,\text{Tr}\left(\mathop{\mathbb{E}}_{x} \left[xx^T \right] (D^i)^T\Delta^i \right)$$

by the cylic property of the trace operator and linearity of expected value. Let $S \coloneqq \mathbb{E}_x[xx^T] \in \mathbb{R}^{n\times n}$. S is positive definite, so we can apply the Cholesky factorization to express it as $S = LL^T$ where L is lower-triangular. So the above becomes

$$= \text{const} + \text{Tr}((\Delta^i L)^T\Delta^i L) + 2\,\text{Tr}((\Delta^i L)^T DL)$$

Let $M \coloneqq \Delta^i L$ and $C \coloneqq -D^i L$.

$$= \text{const} + \text{Tr}(M^TM) - 2\,\text{Tr}(M^TC)$$
$$= \text{const} + \|M - C\|_F^2 \quad (2)$$

where the final equality follows from noticing that

$$\|M - C\|_F^2 = \text{Tr}(M^TM) - 2\,\text{Tr}(M^TC) + \text{Tr}(C^TC)$$

and that C is constant w.r.t. the decision variable Δ^i.

Now, since S is positive definite, we have that L is invertible. This implies $\text{rank}(M) = \text{rank}(\Delta^i)$. Hence, by ignoring constant terms w.r.t. Δ^i in equation (2), we see that the minimizer of (1) is the same as that of

$$\min_{M\in\mathbb{R}^{n\times n}} \|M - C\|_F^2 \quad (3)$$
$$s.t. \quad \text{rank}(M) \le k$$

By the Eckart-Young-Mirsky Theorem, the minimizer of (3) is attained by the rank-k truncated singular value decomposition of C, that is,

$$M_{\text{opt}} = U\ \text{diag}(\sigma_1, \dots, \sigma_k, 0, \dots, 0)\ V^T$$

where $C = U\Sigma V^T, \Sigma \coloneqq \text{diag}(\sigma_1, \dots, \sigma_n)$. Therefore, $\Delta^i_{\text{opt}} = M_{\text{opt}}L^{-1}$. We report the optimal objective value in Figure 2 of the paper, normalized across different choices of k independently per network layer.

Authors' Contact Information: Anh Truong, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA, anh_t@mit.edu; Ahmed H. Mahmoud, Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA, ahdhn@mit.edu; Mina Konaković Luković, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA, minakl@mit.edu; Justin Solomon, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA, jsolomon@mit.edu.

SA Conference Papers '25, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2137-3/2025/12
https://doi.org/10.1145/3757377.3763882

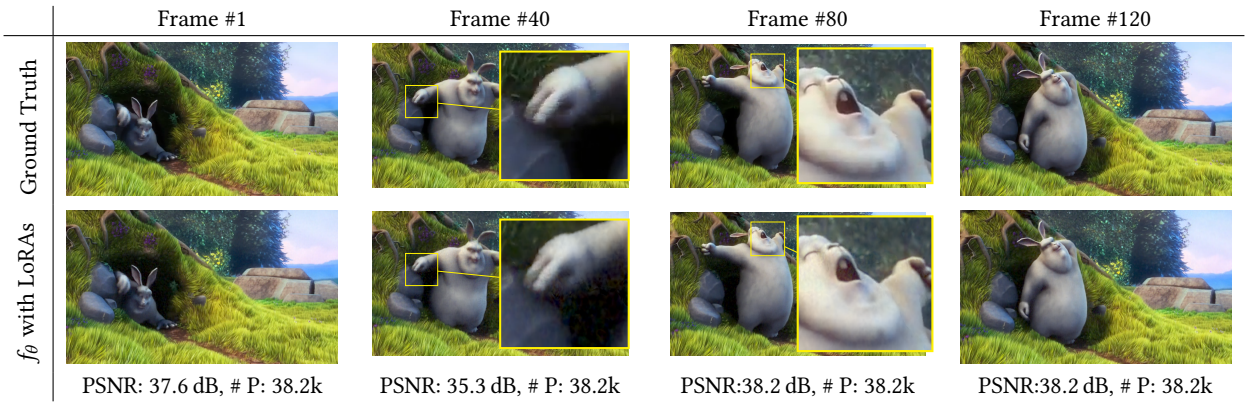

Fig. 1. We also test our sequential LoRA algorithm for encoding an animated sequence. We encode 120 frames from the Big Buck Bunny animation.

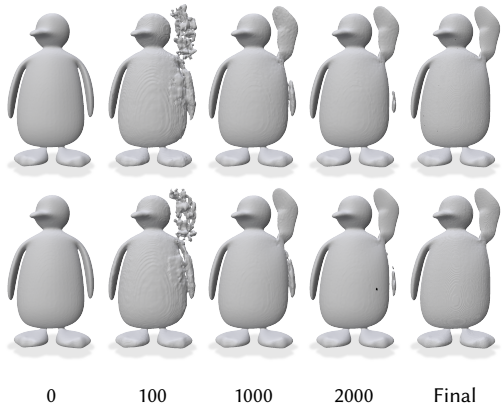

Fig. 2. Progression of SDF surface reconstructions using LoRA (top) and full fine-tuning (bottom). Each column shows the extracted surface after a given number of training steps (0, 100, 1000, 2000, Final). Despite the compressed representation, LoRA rapidly tracks the improvements as of full fine-tuning and recovers clean geometry at convergence.