KoRA: A Framework for Compositional Fine-Tuning

Avaya Aggarwal
Netaji Subhas University of Technology, India

December 18, 2025

Abstract

Low-Rank Adaptation (LoRA) is a standard for parameter-efficient fine-tuning (PEFT), but its layer-wise isolation limits compositional feature learning and out-of-distribution generalization. This proposal introduces KoRA (Kolmogorov-inspired Rank Adapters), a novel PEFT method that enables inter-adapter communication via lightweight, learnable connections. The core idea is to facilitate a directed flow of information, allowing the model to learn a unified, compositional adaptation strategy. Preliminary results on CIFAR-100 and Tiny ImageNet suggest that KoRA significantly outperforms LoRA in cross-domain transfer. The goal of this research is to extensively validate and refine the KoRA architecture to demonstrate its superiority for building robust, generalizable foundation models. The code for the project can be found at https://github.com/OnePunchMonk/KoRA

1 Introduction

The advent of large-scale foundation models, such as the Vision Transformer (ViT)[2], has marked a paradigm shift in computer vision. However, full fine-tuning has become computationally prohibitive. To address this, Parameter-Efficient Fine-Tuning (PEFT) methods[3] have gained prominence, with Low-Rank Adaptation (LoRA)[4] being one of the most popular techniques.

However, the very design that makes LoRA efficient also imposes a critical architectural limitation: **layer-wise isolation**. Each LoRA adapter operates independently, with no mechanism for coordination. We argue that this is a primary factor constraining the generalization capabilities of LoRA-tuned models, particularly when faced with out-of-distribution data.

This research proposes to move beyond isolated adaptation and introduce a model of structured compositionality. We present **KoRA (Kolmogorov-inspired Rank Adapters)**, a novel PEFT method that reframes adapters as nodes in a directed computational graph, facilitating a sequential flow of information from early to late layers. The goal of this work is to develop and validate this idea, showing that a compositional approach leads to more robust and generalizable models.

2 Methodology

2.1 Revisiting Low-Rank Adaptation (LoRA)

A pre-trained model can be described as a function $f(x; W_0)$. LoRA approximates the update matrix ΔW by factorizing it into two smaller matrices: $A \in \mathbb{R}^{r\times d_{in}}$ and $B \in \mathbb{R}^{d_{out}\times r}$. The forward pass is modified as:

$$h = W_0x + \Delta W x = W_0x + \alpha(BA)x \quad (1)$$

where α is a scaling scalar. The critical limitation is that the total adaptation is the sum of isolated perturbations, $\{\Delta W^{(1)}, \ldots, \Delta W^{(L)}\}$, preventing the model from learning dependencies between layer adaptations.

2.2 KoRA: A Compositional Adaptation Strategy

KoRA reframes adapters into nodes within a directed computational graph. This approach is motivated by the Kolmogorov-Arnold representation theorem[11, 7], which suggests any complex function can be decomposed into a composition of simpler functions.

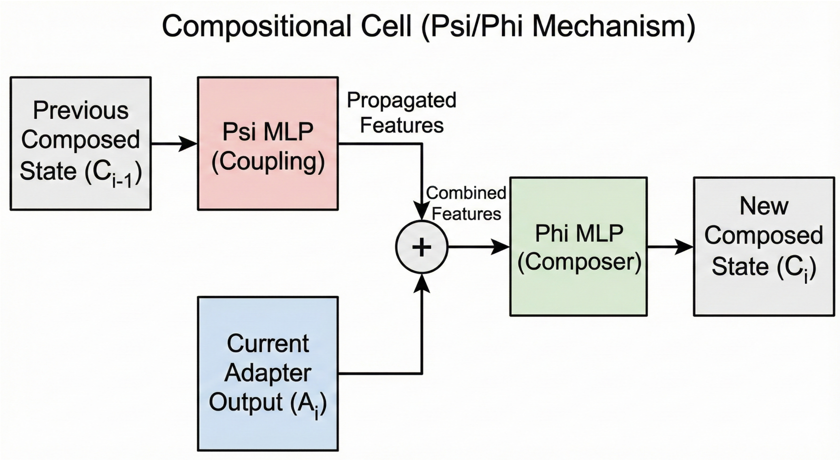

Figure 1: Compositional Cell

The KoRA architecture as shown in Fig.1, 2, 3 extends LoRA with a recursive flow:

1. Inner Adaptation and Projection. We compute the LoRA output, $o^{(i)} = (B^{(i)}A^{(i)})x^{(i)}$, and project it into a common composition space of dimension d_{comp}:

$$p^{(i)} = P^{(i)}_{\text{proj}}(o^{(i)}) \tag{2}$$

2. Sequential Compositional Flow. The projected vectors are sequentially integrated. The composed feature vector $c^{(i)}$ is updated recursively:

$$c^{(i)} = \begin{cases} \Phi^{(1)}(p^{(1)}) & \text{if } i = 1 \\ \Phi^{(i)}(p^{(i)} + \Psi^{(i-1)}(c^{(i-1)})) & \text{if } i > 1 \end{cases} \tag{3}$$

Here, Ψ is a lightweight *coupling network* propagating the state, and Φ is a *composer network* integrating new information.

3. Final Aggregation. The final composed vector, $c^{(L)}$, is projected back to the model's hidden dimension and added to the [CLS] token representation:

$$h_{\text{final}} = h_{\text{CLS}} + \alpha \cdot \text{Proj}_{\text{final}}(c^{(L)}) \tag{4}$$

3 Experimental Setup

This section outlines the initial experiments conducted to provide a proof-of-concept for the KoRA architecture.

3.1 Implementation Details

The classification experiments were conducted on a Kaggle GPU (NVIDIA P100). The depth estimation experiments utilized a more powerful A100 GPU, accessed via the Lightning AI platform.

3.2 Task Specialization on CIFAR-100

Models were fine-tuned for 5 epochs on the CIFAR-100[6] training set. Table 1 compares the performance of PEFT methods.

3.3 Task Generalization on Tiny ImageNet

Models pre-trained on CIFAR-100 were then fine-tuned for one epoch on 1% of the Tiny ImageNet[1] training data. Table 2 shows the results.

3.4 Representation Analysis

Preliminary Centered Kernel Alignment (CKA) analysis[5] suggests that KoRA learns more structured, cross-layer dependencies than LoRA. Early visualizations show higher similarity scores between layers in the KoRA model, indicating a more stable feature hierarchy.

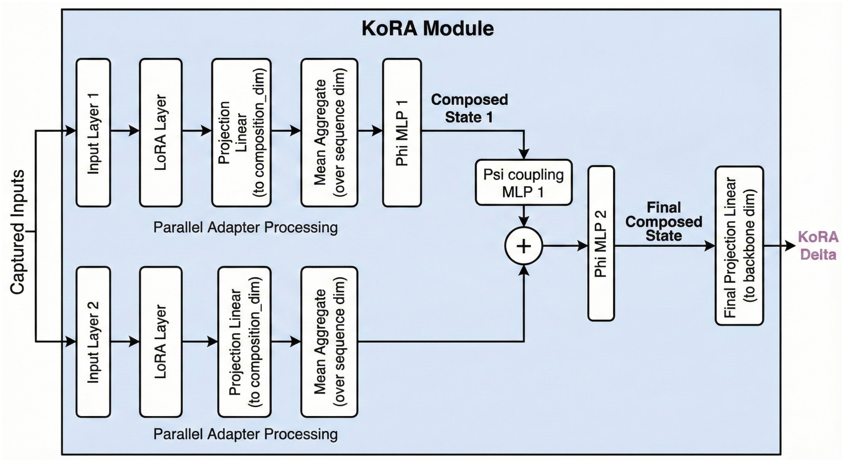

Figure 2: KoRA Module Dataflow

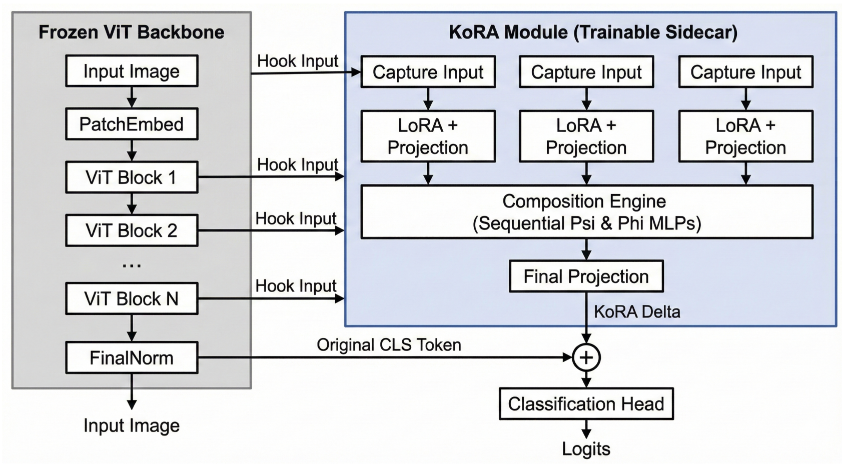

Figure 3: System-Wide Integration

Table 1: Comparison of ViT Fine-Tuning Methods on CIFAR-100 (Task Specialization).

Tuning Method	**Params Tuned (%)**	**CKA Sim.**	**Accuracy (%)**	**F1 Score**
PEFT Methods				
LoRA ($r = 8$)	1.45	0.73	**92.48**	**0.924**
Adapter Fusion[9]	1.45	0.71	92.22	0.922
KoRA ($d_{\text{comp}} = 4$)	1.80	**0.76**	83.96	0.840
KoRA ($d_{\text{comp}} = 8$)	2.18	**0.76**	84.19	0.842

Table 2: Task Generalization: Classification Performance on Tiny ImageNet.

Tuning Method	**Accuracy (%)**	**F1 Score**
LoRA ($r = 8$)	71.04	0.8307
Adapter Fusion	46.67	0.6364
KoRA ($d_{\text{comp}} = 4$)	97.37	0.9867
KoRA ($d_{\text{comp}} = 8$)	**98.24**	**0.9911**

3.5 Depth Estimation on NYU Depth V2

The models were also fine-tuned for one epoch on 1% of the NYU Depth V2[12] training data.

These initial results show LoRA performing better on this specific dense prediction task, highlighting an important area for future investigation of domain generalisation[8].

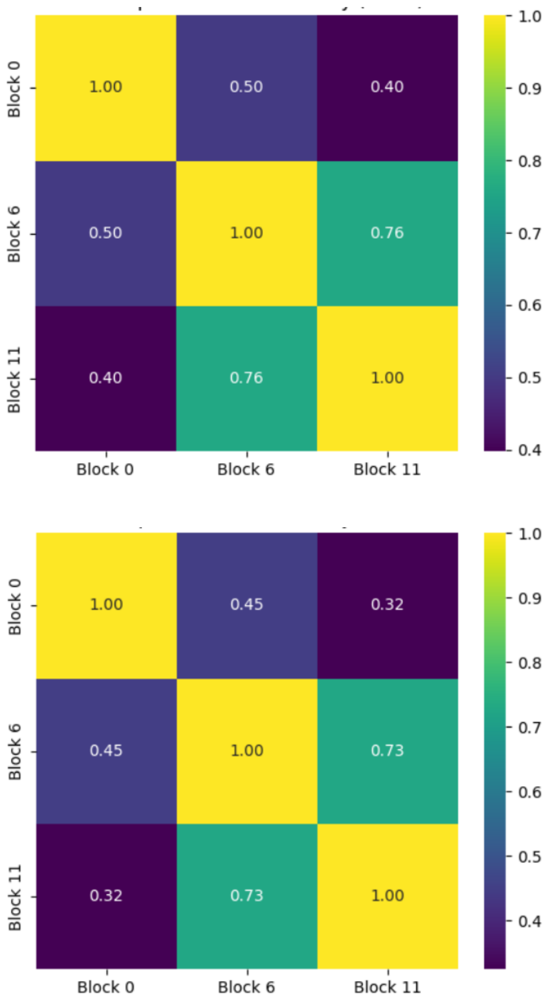

Figure 4: CKA analysis: KoRA (Top) vs LoRA (Bottom)

Table 3: Transfer learning on NYU Depth V2. **Note: For these metrics, lower is better.**

Method	**Metric**	**Score**
LoRA	RMSE	**0.2800**
	AbsRel	**0.5629**
KoRA	RMSE	0.3327
	AbsRel	0.6271

4 Discussion and Future Work

While preliminary results demonstrate that KoRA offers a significant advantage in cross-domain transfer (Table 2), they also reveal a "specialization-generalization" trade-off. LoRA remains highly effective for narrow, in-domain tasks, whereas KoRA's strength lies in learning a unified adaptation strategy that transcends specific datasets. Our future work is structured around three key research pillars:

Architectural Refinement via Attention Mechanisms. The current iteration of KoRA utilizes static MLPs [10] for the composer (Φ) and coupling (Ψ) networks. A primary goal is to transition from this rigid sequential flow to a *dynamic compositional attention* mechanism. By allowing adapters to selectively attend to the states of multiple preceding layers, we hypothesize that the model can learn more complex, non-linear feature hierarchies. We will also investigate the use of Orthogonal Projections in the composition space to minimize redundant information across adapters.

Scaling and Dense Prediction Robustness. As evidenced by Table 3, KoRA currently faces challenges in dense prediction tasks like depth estimation. We aim to investigate whether this is a result of the composition space (d_{comp}) acting as a bottleneck for high-resolution spatial information. Future experiments will scale the composition dimensionality and evaluate KoRA's performance in object detection and semantic segmentation.

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

[3] Zeyu Han, Chao Gao, Jialiang Liu, Jeff Zhang, and Sai Qian Zhang. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*, 2024.

[4] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[5] Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 3519–3529. PMLR, 2019.

[6] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[7] Ziming Liu, Yixuan Wang, Vaibhav Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2024.

[8] Yue Ni, Shixiong Zhang, and Piotr Koniusz. Pace: Marrying generalization in parameter-efficient fine-tuning with consistency regularization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[9] Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. Adapterfusion: Non-destructive task composition for transfer learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pages 487–503, 2021.

[10] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[11] Johannes Schmidt-Hieber. The kolmogorov–arnold representation theorem revisited. *Neural Networks*, 137:119–126, 2021.

[12] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision (ECCV)*, pages 746–760. Springer, 2012.