SuperLoRA: Parameter-Efficient Unified Adaptation of Large Foundation Models

Xiangyu Chen[1,2], Jing Liu[2], Ye Wang[2], Pu (Perry) Wang[2], Matthew Brand[2], Guanghui Wang[3], Toshiaki Koike-Akino[2]

[1] University of Kansas, Lawrence, KS 66045, USA

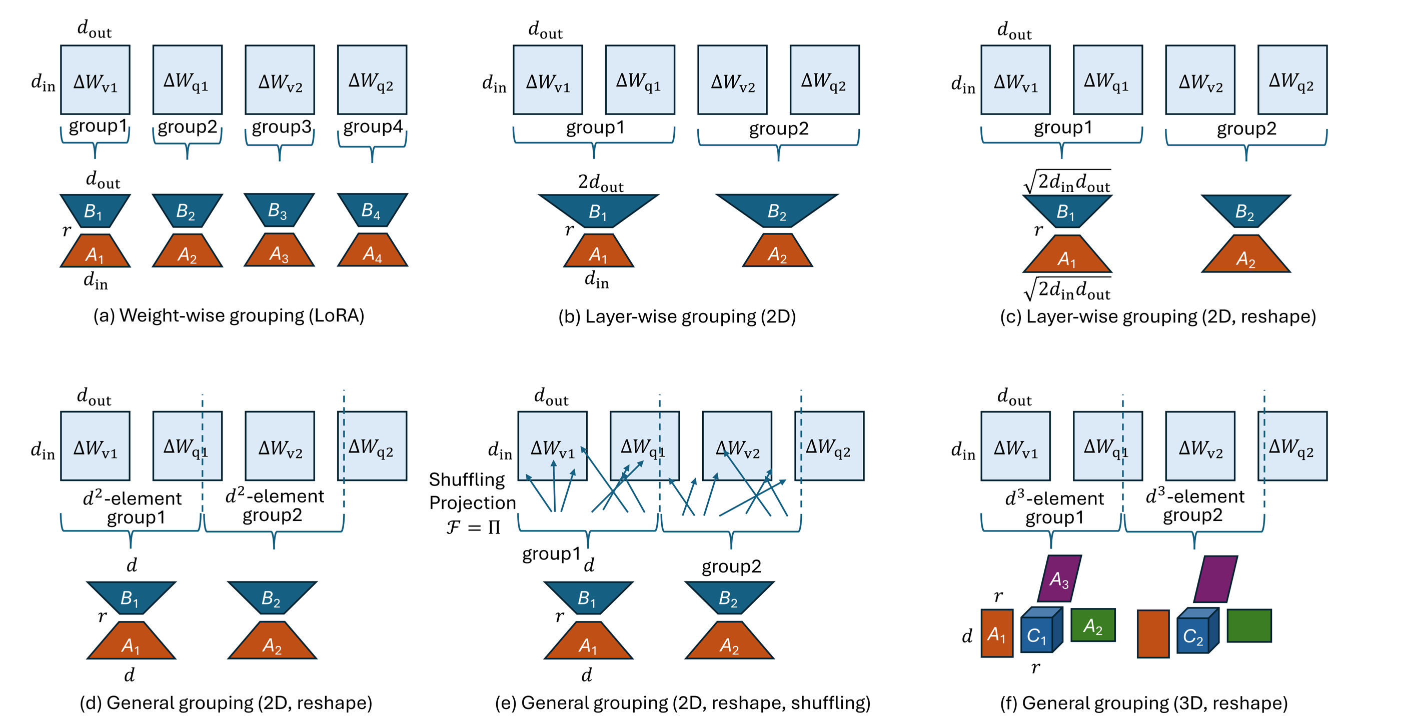
[2] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA

[3] Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada

xychen@ku.edu, {xiachen, jiliu, yewang, pwang, brand, koike}@merl.com, wangcs@torontomu.ca

Motivation

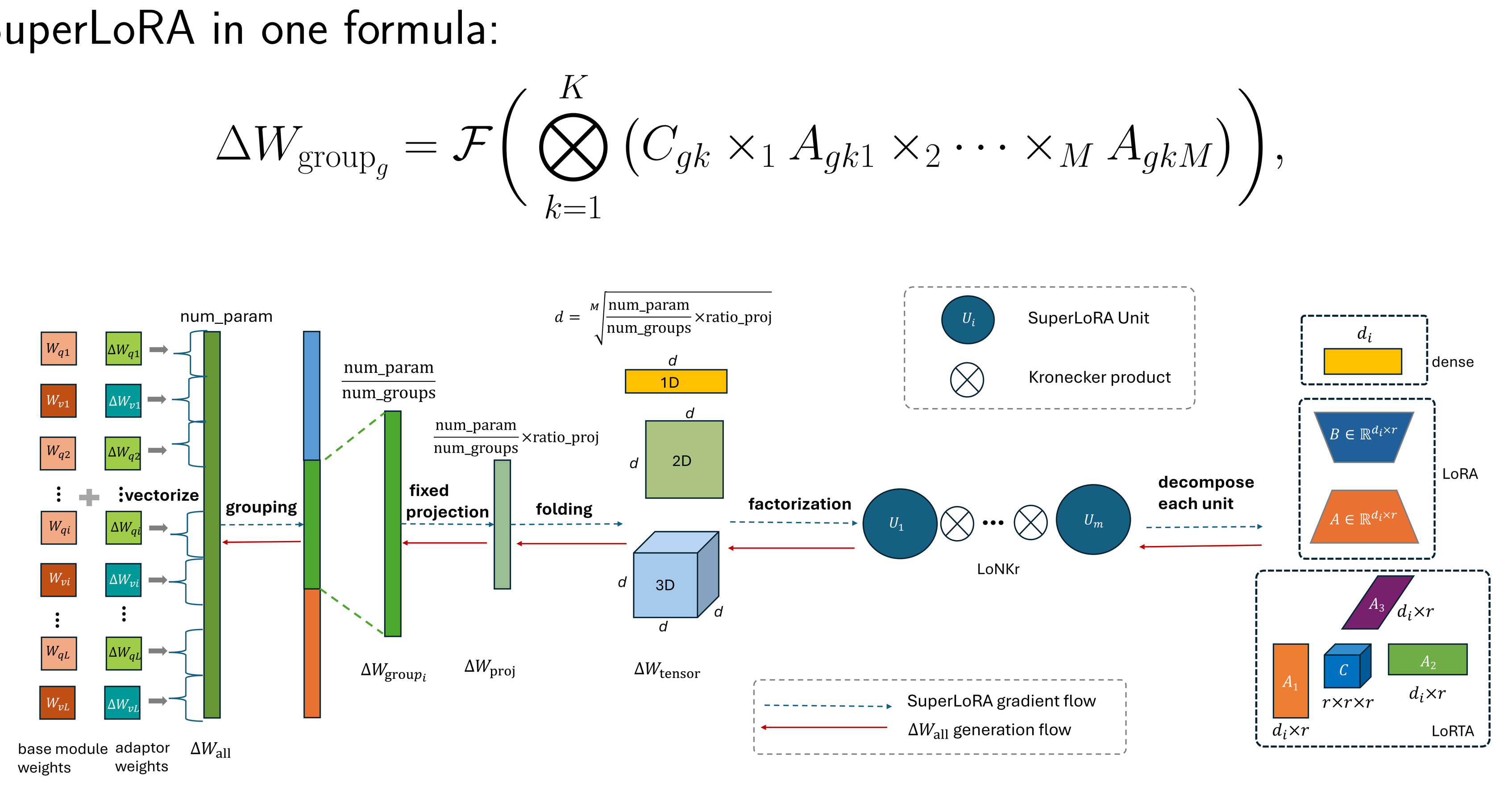
LoRA - look at each weight matrix W_0 separately; SuperLoRA - only care about the total number of parameters to update $W_{0_{total}}$



Method

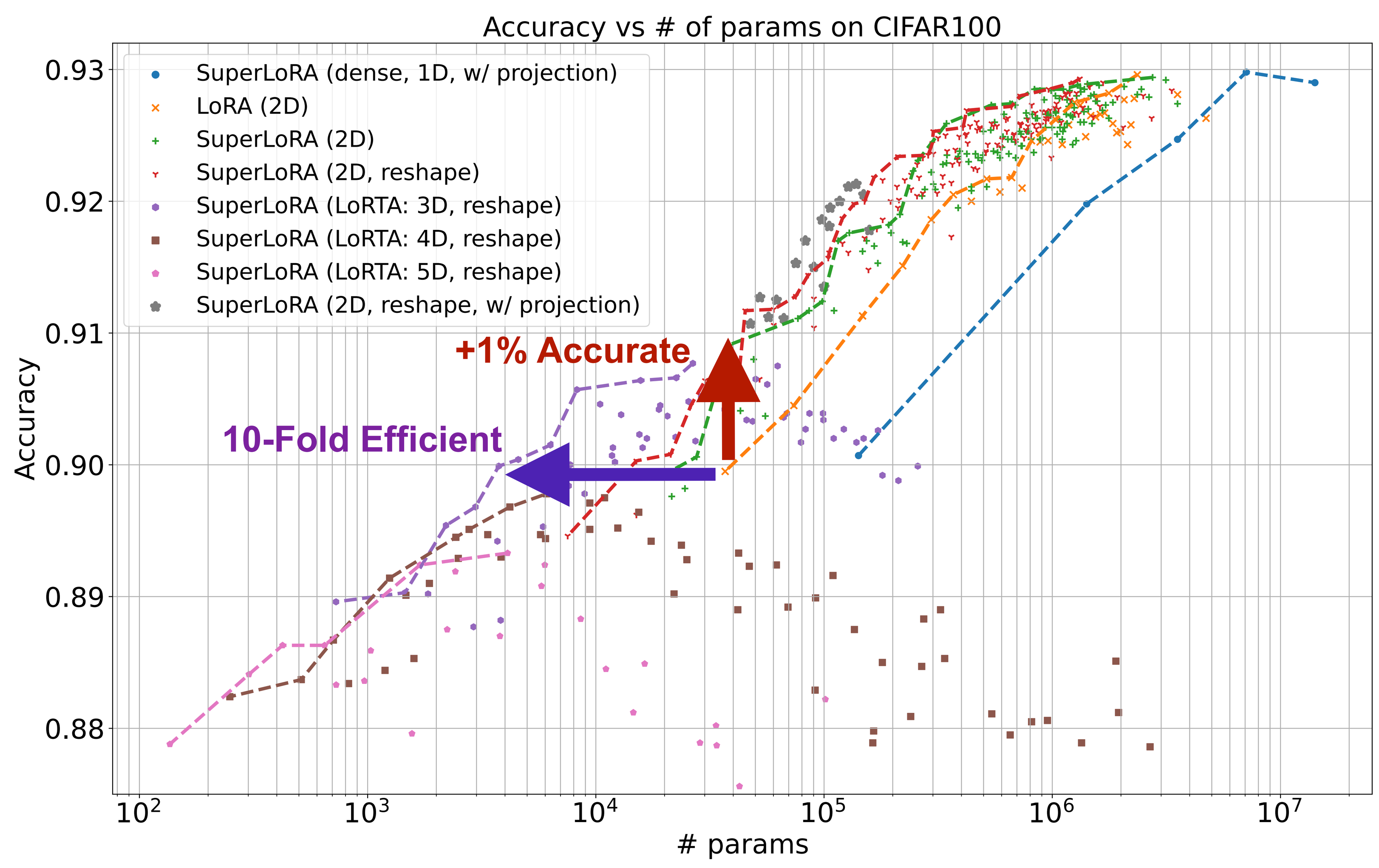
SuperLoRA in one formula:

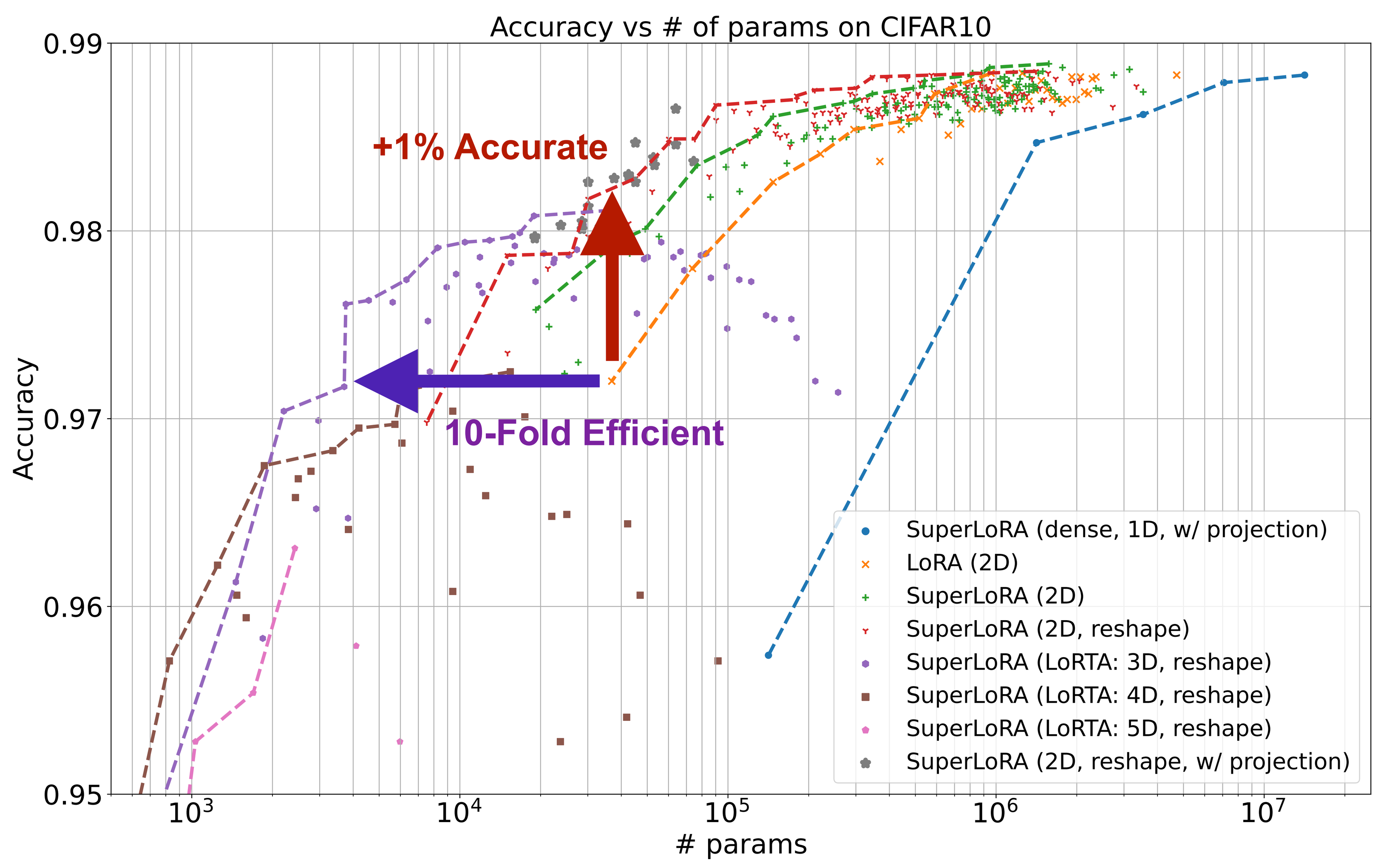
$$\Delta W_{\text{group}_g} = \mathcal{F}\left(\bigotimes_{k=1}^{K}\left(C_{gk}\times_1 A_{gk1}\times_2\cdots\times_M A_{gkM}\right)\right),$$



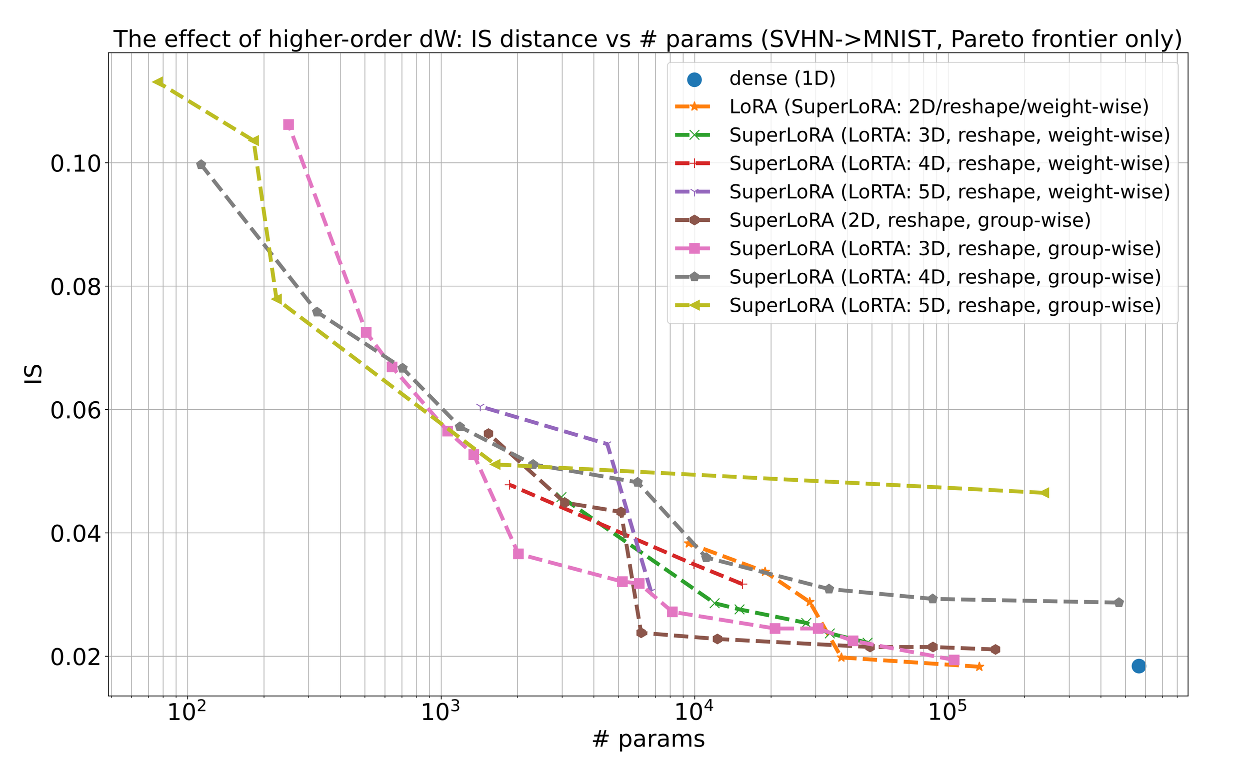
Paper QR

Experiments - Image Classification (ImageNet–>CIFAR)





Experiments - Image Generation



Vis. (# params 32)

Experiments - E2E Challenge

Table: GPT-2 medium with different adaptation methods on E2E NLG Challenge. For all metrics, higher is better. * indicates numbers published in prior works, as compiled by [2].

Method	# Trainable Parameters	E2E NLG Challenge				
		BLEU	NIST	MET	ROUGE-L	CIDEr
FT*	354.92M	68.2	8.62	46.2	71.0	2.47
Adapter^{L}*	0.37M	66.3	8.41	45.0	69.8	2.40
Adapter^{L}*	11.09M	68.9	8.71	46.1	71.3	2.47
Adapter^{H}*	11.09M	$67.3_{\pm.6}$	$8.50_{\pm.07}$	$46.0_{\pm.2}$	$70.7_{\pm.2}$	$2.44_{\pm.01}$
FT$^{\text{Top2}}$*	25.19M	68.1	8.59	46.0	70.8	2.41
FT$^{W_q W_v}$	48.00M	$69.4_{\pm.1}$	$8.74_{\pm.02}$	$46.0_{\pm.0}$	$71.0_{\pm.1}$	$2.48_{\pm.01}$
LoRA	0.40M	$69.28_{\pm.01}$	$8.73_{\pm.08}$	$46.51_{\pm.00}$	$71.4_{\pm.00}$	$2.49_{\pm.02}$
SuperLoRA	**0.12M**	$\mathbf{69.82}_{\pm.00}$	$\mathbf{8.76}_{\pm.02}$	$\mathbf{46.54}_{\pm.00}$	$\mathbf{71.5}_{\pm.00}$	$\mathbf{2.50}_{\pm.01}$

SuperLoRA and its derivation

Table: Hyperparameter settings in SuperLoRA and resultant LoRA variant.

hyper-parameters settings	method
$\mathcal{F}=I$, weight-wise, $K=1$, $C_{g1}=I$, $M=1$	dense FT
$\mathcal{F}=I$, weight-wise, $K=1$, $C_{g1}=I$, $M=2$	LoRA [2]
$\mathcal{F}=I$, weight-wise, $K=2$, $C_{gk}=I$, $M=2$	LoKr [3]
$\mathcal{F}=I$, group-wise, $G=1$, $M>2$	LoTR [1]
$\mathcal{F}=I$, group-wise, $K>2$, $C_{gk}=I$, $M=2$	LoNKr
$\mathcal{F}=I$, group-wise, $K=1$, $M>2$	LoRTA

Table: Hyperparameters and notation.

notation	description
r	rank of factorization
$\mathcal{F}$	mapping function
ρ	compression ratio
G	number of groups
M	order of tensor modes
K	number of splits

References

[1] Bershatsky, D., Cherniuk, D., Daulbaev, T., Oseledets, I.: LoTR: Low tensor rank weight adaptation. arXiv preprint arXiv:2402.01376 (2024)

[2] Hu, E.J., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2021)

[3] Yeh, S.Y., Hsieh, Y.G., Gao, Z., Yang, B.B.W., Oh, G., Gong, Y.: Navigating text-to-image customization: From lyCORIS fine-tuning to model evaluation. In: The Twelfth International Conference on Learning Representations (2024)