SuperLoRA: Parameter-Efficient Unified Adaptation for Large Vision Models

Xiangyu Chen, Jing Liu, Ye Wang, Perry Wang, Matt Brand, Guanghui Wang, Toshiaki Koike-Akino

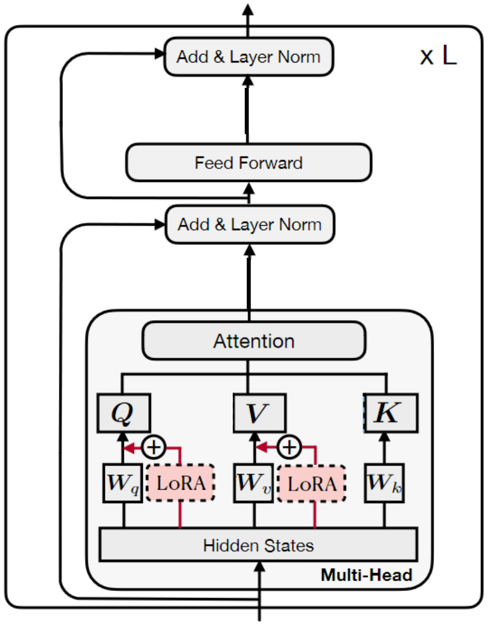
Efficient Deep Learning for Computer Vision CVPR Workshop 2024 (ECV)

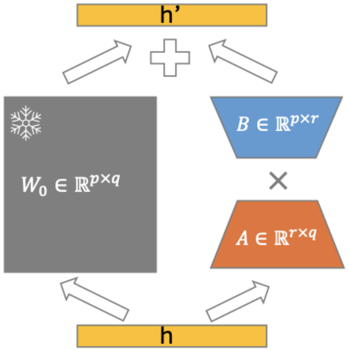
Jun 2024

MITSUBISHI ELECTRIC RESEARCH LABORATORIES (MERL)
Cambridge, Massachusetts, USA
http://www.merl.com

Introduction - LoRA

PEFT → Adapter-based fine-tuning → Low-Rank Adaptation (LoRA)





$$W' = W_0 + \Delta W = W_0 + BA$$

Hu, Edward J., et al. "LoRA: Low-rank adaptation of large language models." *ICLR 2022*.

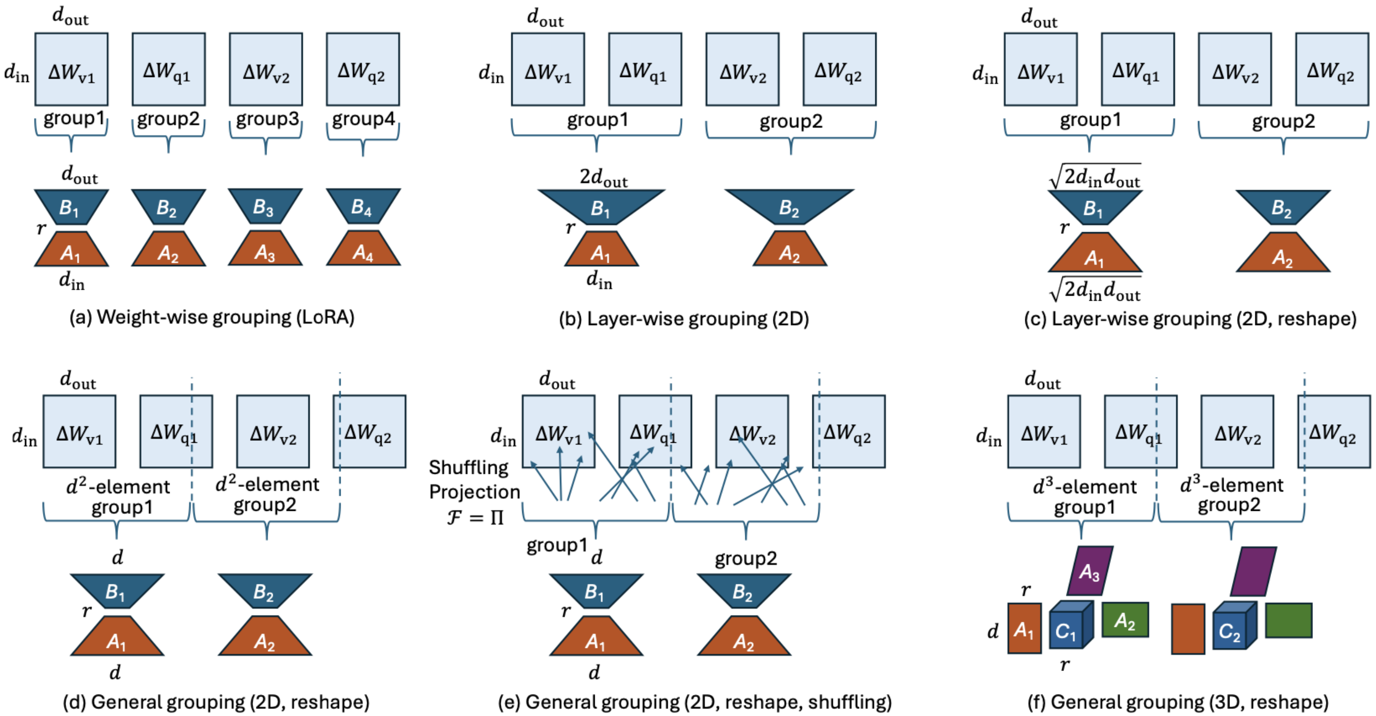
Introduction - SuperLoRA

PEFT → Adapter-based fine-tuning → Low-Rank Adaptation (LoRA) → SuperLoRA

However,
LoRA and its variants regard adapter dW as a 2D matrix
with the same shape ($d_{in} \times d_{out}$) as the weight W_0

What SuperLoRA care is:
The adapters can generate $d_{in} \times d_{out}$ values for each W_0
Not where and how those values come from

SuperLoRA: Grouping for Joint Weight Adaptation



LoRA simply looks at each weight matrix separately
SuperLoRA improves by only caring about the total number of parameters

SuperLoRA: Tensor Rank Adaptation

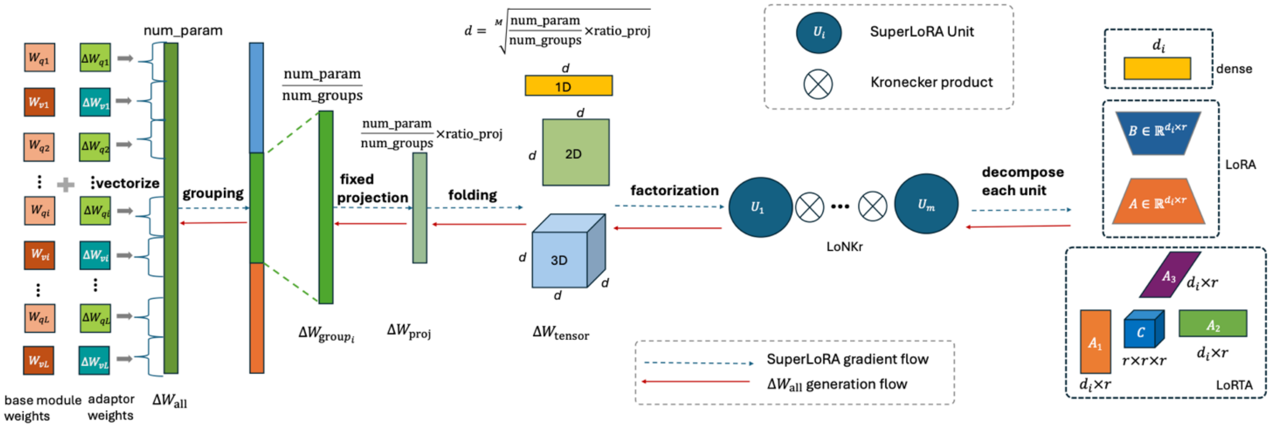


Fig. 1: Schematic of SuperLoRA to fine-tune multi-layer attention modules at once with vectorizing, grouping, projection, folding, and factorization.

SuperLoRA in one formula

$$\Delta W_{\text{group}_g} = \mathcal{F}(\Delta W_{\text{lora}_g}) = \mathcal{F}\left(\bigotimes_{k=1}^{K}\left(C_{gk} \prod_{m=1}^{M} \times_m A_{gkm}\right)\right)$$

SuperLoRA: Unified LoRA Variants

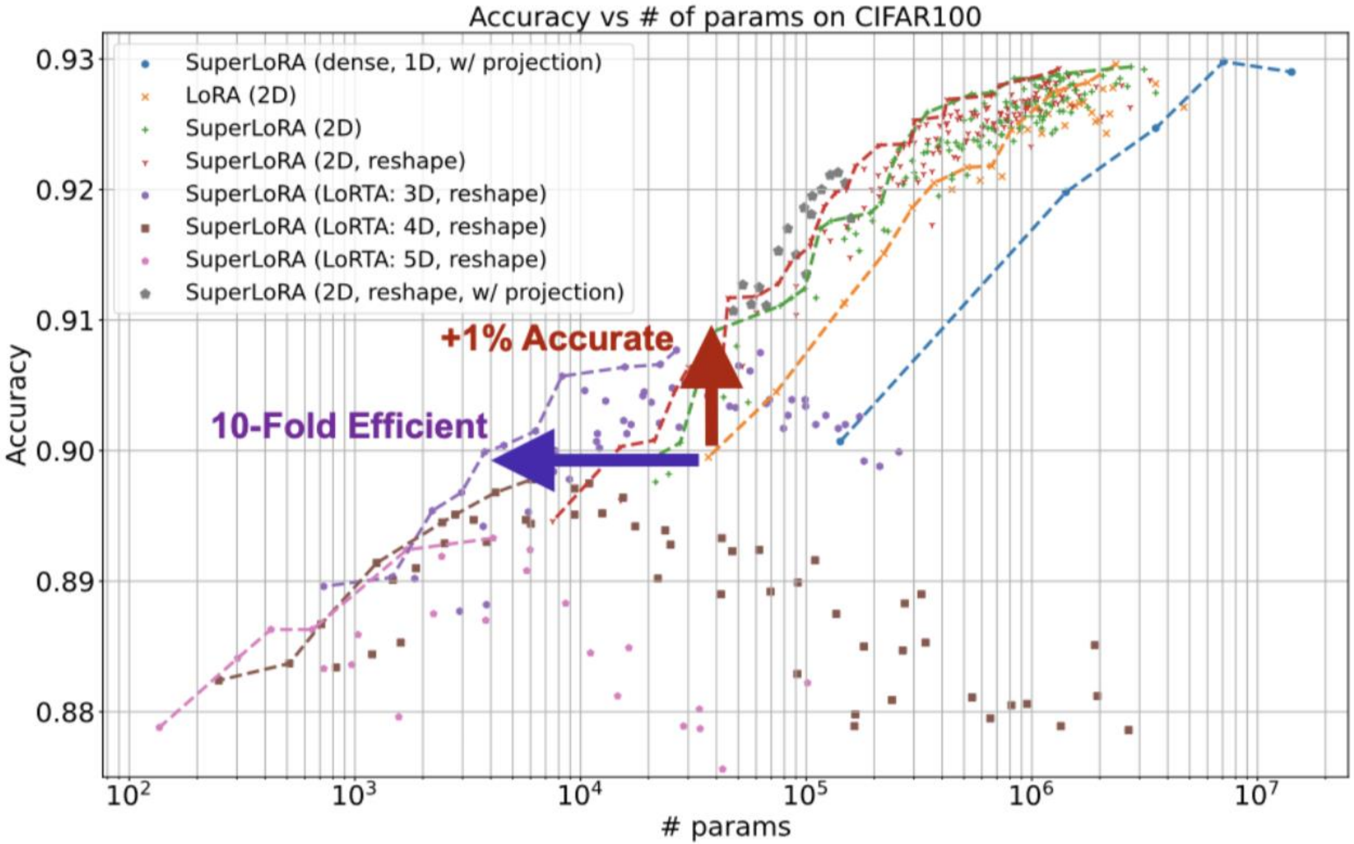
Derived LoRA variants with SuperLoRA

hyper-parameters settings	method
$\mathcal{F}=I$, weight-wise, $K=1$, $C_{g1}=I$, $M=1$, $A_{g11}\in\mathbb{R}^{d_{\text{in}}d_{\text{out}}\times 1}$	dense FT
$\mathcal{F}=I$, weight-wise, $K=1$, $C_{g1}=I$, $M=2$, $A_{g1m}\in\mathbb{R}^{d_m\times r}$	LoRA [21]
$\mathcal{F}=I$, weight-wise, $K=2$, $C_{gk}=I$, $M=2$, $A_{gkm}\in\mathbb{R}^{d_m\times r}$	LoKr [42]
$\mathcal{F}=I$, group-wise, $G=1$, $M>2$	LoTR [5]
$\mathcal{F}=I$, group-wise, $K>2$, $C_{gk}=I$, $M=2$, $A_{gkm}\in\mathbb{R}^{d_m\times r}$	LoNKr
$\mathcal{F}=I$, group-wise, $K=1$, $M>2$, $A_{gkm}\in\mathbb{R}^{d_m\times r}$	LoRTA

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

[5] Daniel Bershatsky, and et al.. LoTR: Low tensor rank weight adaptation. 2024
[21] Hu, Edward J., et al. "LoRA: Low-rank adaptation of large language models.” *ICLR 2022*.
[42] Shin-Ying Yeh , Navigating text-to-image customization: From lyCORIS fine-tuning to model evaluation. ICLR 2024

Experiments - Image Classification

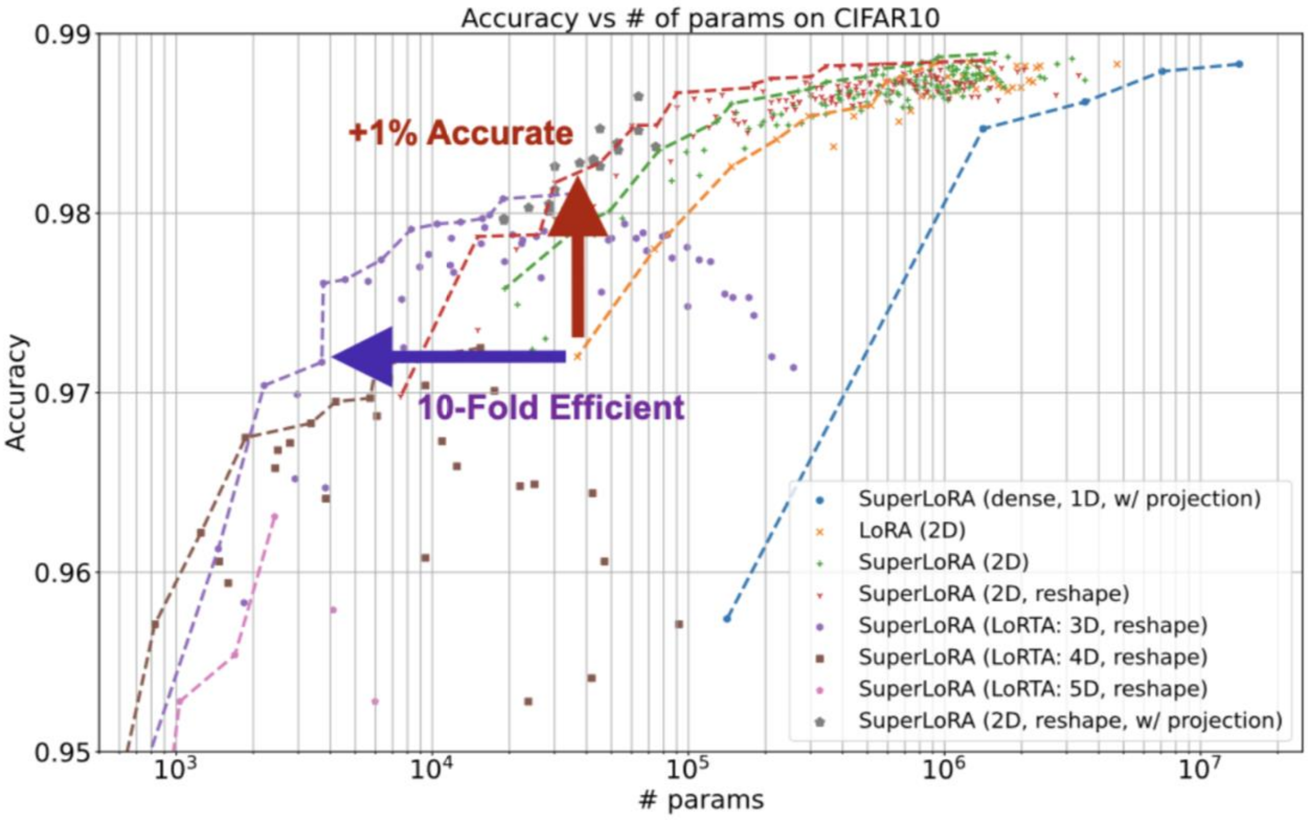


Pareto frontier lines included

Settings:

ImageNet21k -> CIFAR100
ViT-Base (86.6M # params)

Experiments - Image Classification



Settings:

ImageNet1k -> CIFAR10
ViT-Base (86.6M # params)

Pareto frontier lines included

THANKS

Email: koike@merl.com

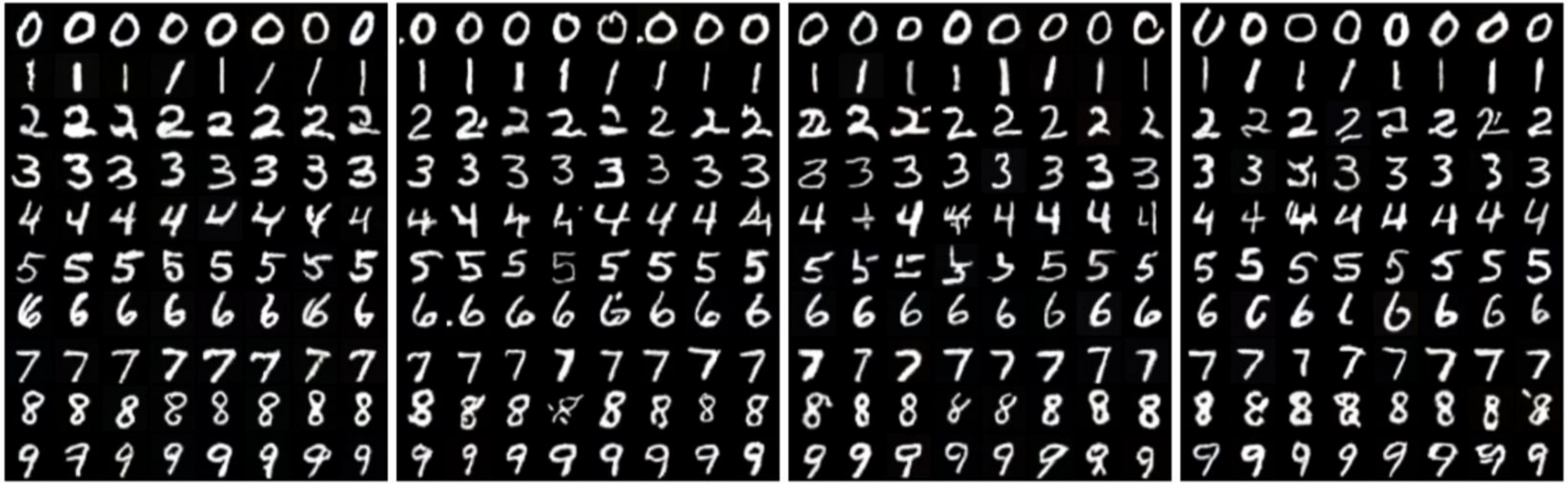
Visualization – Image Generation (SVHN → MNIST)

Diffusion model adaptation

Generated images
after pre-training on SVHN datasets

Expected generated images
after fine-tuning on MNIST datasets

Visualization – Image Generation



(a) dense
#param 565,248
IS 0.0184

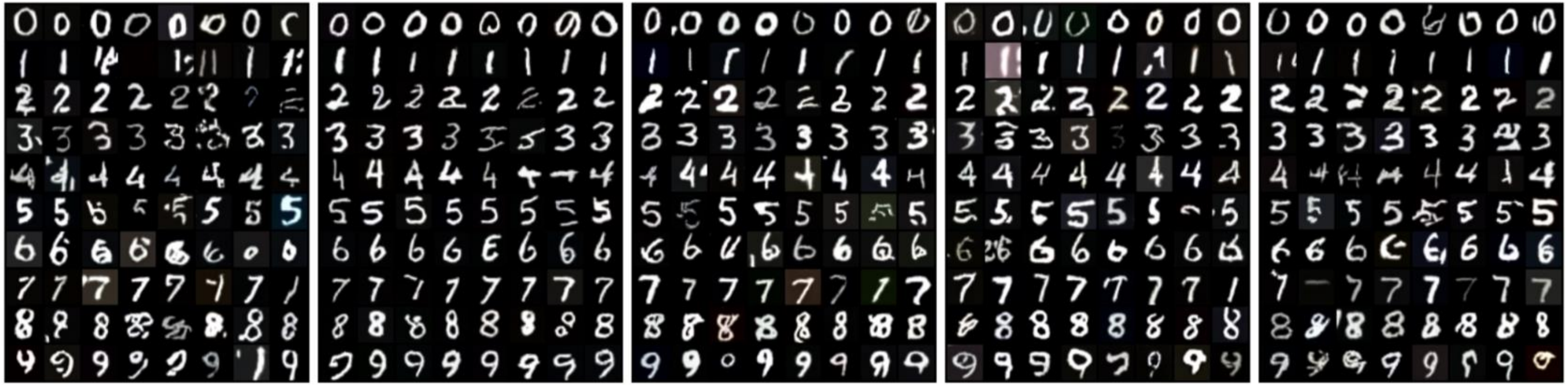
(b) LoRA $r = 8$
#param 75,776
IS 0.03025

(c) SuperLoRA
#param 70,720
IS 0.0305

(d) SuperLoRA
#param 73,728
IS 0.0263

Figure 19. Visualization of generated images under high-parameter level (> 70,000).

Visualization – Image Generation



(a) LoNKr
#param 5,112
IS 0.036

(b) SuperLoRA
#param 10,752
IS 0.0294

(c) LoRTA
#param 8,160
IS 0.0272

(d) LoRTA
#param 11,100
IS 0.036

(e) SuperLoRA w/ p
#param 8,512
IS 0.0273

Figure 20. Visualization of generated images under middle-parameter level ([5,000, 20,000]).

Visualization – Image Generation

(a) LoNKr
#param 1,080
IS 0.0471

(b) LoRTA
#param 1,060
IS 0.0565

(c) SuperLoRA w/ p
#param 832
IS 0.0607

(d) LoRTA
#param 76
IS 0.1131

(e) LoRTA w/ p
#param 31
IS 0.0871

Figure 21. Visualization of generated images under low-parameter level (1,000) and extremely-low level (< 100).

MITSUBISHI ELECTRIC
Changes for the Better

Geometric Analysis

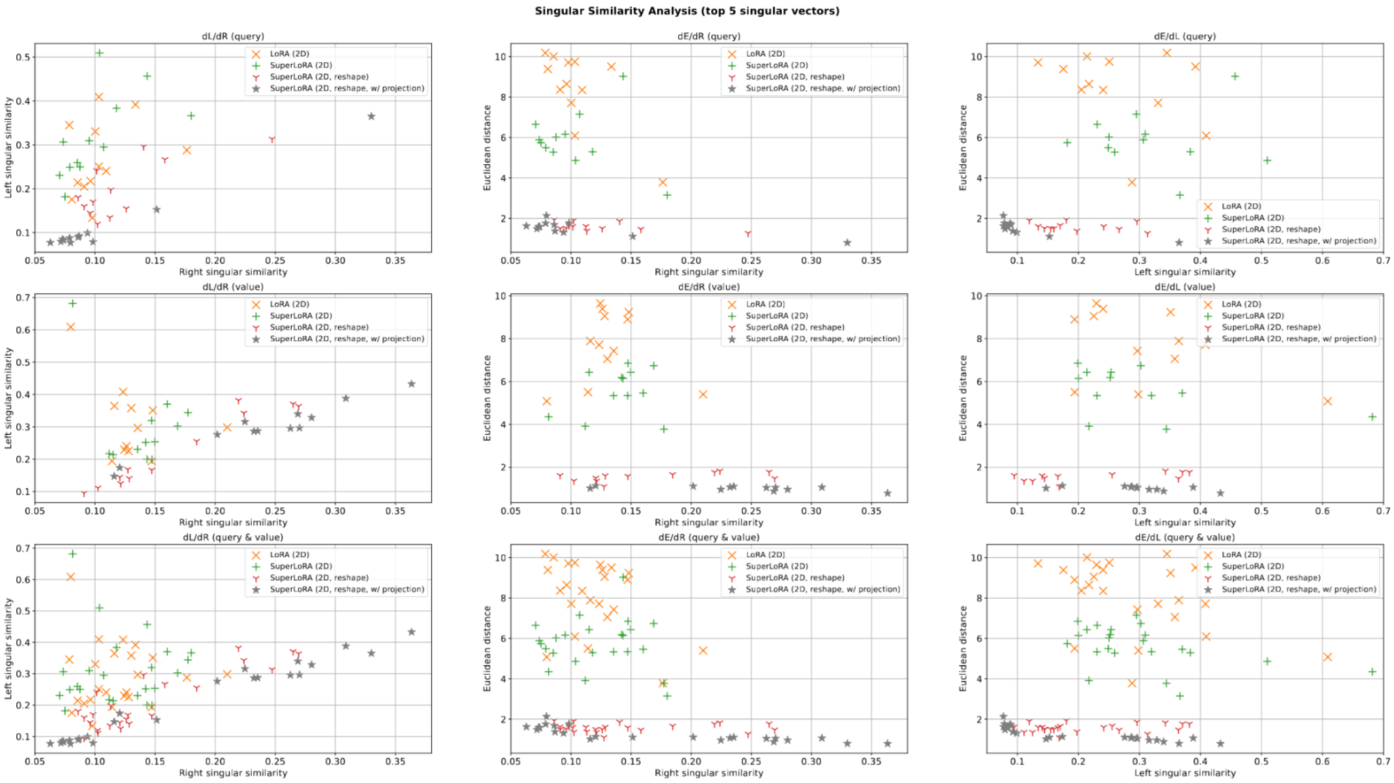


Figure 11. Geometric similarity analysis (top 5 principal singular vectors).

To converge to dense FT:
dE → 0
dL/dR → 1

1. dE dropped for SuperLoRA
2. dL: LoRA > SuperLoRA
3. dR: SuperLoRA > LoRA
4. dR: value > query