Preference-Aligned LoRA Merging: Preserving Subspace Coverage and Addressing Directional Anisotropy

Wooseong Jeong*, Wonyoung Lee*, and Kuk-Jin Yoon
{stk14570, wylee, kjyoon}@kaist.ac.kr Visual Intelligence Lab., KAIST, Korea

CVPR JUNE 3-7, 2026 DENVER COLORADO

Background

- **Model Merging.** Integrate knowledge from separately fine-tuned models into one without multi-task training.
- Low-Rank Adaptation (LoRA) has emerged as the dominant fine-tuning approach, making LoRA merging a particularly compelling direction for further exploration.
- **LoRA merging** combines task-specific LoRA adapters into a single multi-task model while reducing the need for costly re-training.
- Existing methods overlook two key aspects of LoRA me rging: subspace coverage and directional anisotropy.

Motivation

- Two overlooked aspects of LoRA merging:
  - **Subspace coverage.** Naive interpolation collapses task -specific LoRA directions and erodes representational capacity.
  - **Directional anisotropy.** Task losses are unequally sens itive to different LoRA directions, distorting multi-task trade-offs.

Contributions

- Identify **subspace coverage** and **anisotropy** as two key aspects of LoRA merging.
- Propose **TARA** (Task-Rank Anisotropy Alignment) — preserves LoRA subspaces while reweighting directions by sensitivity.
- **SOTA** on 8 vision tasks and 6 NLI benchmarks, strong generalization to unseen tasks.

Subspace Coverage & Anisotropy

- **Notation.** Base W_0, for task i, LoRA adapter

$$\Delta W_i = B_i A_i^\top = \sum_{j=1}^{r_i} b_{ij} a_{ij}^\top$$

r_i = LoRA rank of task i (number of rank-1 components $b_{ij}a_{ij}^\top$).

- **Subspace coverage via effective rank:**

$$erank(X) = exp(-\sum_k p_k \log p_k), \qquad p_k = \frac{\sigma_k^2}{\sum_j \sigma_j^2}$$

Per rank-1 component $b_{ij}a_{ij}^\top$ contributes one singular direction.
Higher *erank* = more rank-1 directions retained.

- Three stacks compared. (per-task sum / **ΔX-stack**, LoRA-agnostic / **Rank-1 stack**, LoRA-aware).
- **Anisotropy via task-loss Jacobian:**

$$\Delta f \approx J\varphi, \qquad J_{i,k} = \langle \nabla f_i(W), S_k \rangle_F$$

$J_{i,k}$ measures task *i*'s loss sensitivity along the *k*-th rank-1 direction.

- **Proposition 1 (Anisotropy Bounds).**

$$\sigma_{min}(J)\|\varphi\|_2 \le \|\Delta f\|_2 \le \sigma_{max}(J)\|\varphi\|_2$$

- Large $\kappa(J) = \frac{\sigma_{max}}{\sigma_{min}} \Rightarrow$ equal-norm φ produce disproportionate loss changes.
- **Directional sensitivity misalignment:**

$$\xi(\rho_1, \rho_2) = 1 - \frac{|\langle h(\rho_1), h(\rho_2)\rangle|}{\|h(\rho_1)\| \, \|h(\rho_2)\|}$$

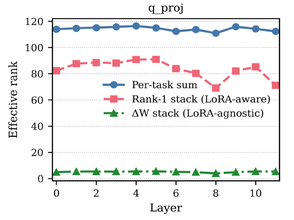

Figure 1. Effective rank across layers

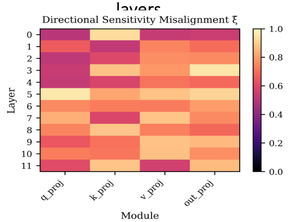

Figure 2. Directional-sensitivity misalignment.

TARA = Task-Rank Anisotropy Alignment. Reweight LoRA directions with selection weights φ, optimized under preference ρ via an entropy-minimization pseudo-loss.

- **Variant A — Per-rank LoRA direction selection.** One scalar φ_{ij} per rank-1 factor; no SVD step; per-adapter anisotropy control.

$$W_A(\varphi) = W_0 + \sum_{i=1}^{N}\sum_{j=1}^{r_i} \varphi_{ij} b_{ij} a_{ij}^\top$$

- **Variant B — Shared singular-direction selection.** Concatenate adapters, SVD, partition right-singular vectors by task; shared $\{u_k\}$ preserves coverage, per-task v_{ki} + weights φ_{ik} reduce interference.

$$X = [\Delta W_1, \dots, \Delta W_N] = U\Sigma V^\top$$
$$v_k = [v_{k1}^\top, \dots, v_{kN}^\top]^\top, \qquad S_{i,k} = u_k v_{ki}^\top$$
$$W_B(\varphi) = W_0 + \sum_i \sum_k \varphi_{ik}\sigma_k u_k v_{ki}^\top$$

- **Smooth Tchebycheff scalarization.** Anchors z_i = entropy when only adapter i is applied:

$$\Psi(\varphi, \rho) = \alpha \log\left(\sum_i exp\left(\frac{\rho_i(f_i(W(\varphi)) - z_i)}{\alpha}\right)\right)$$

- Small α → worst-task focus; large α → even spread.

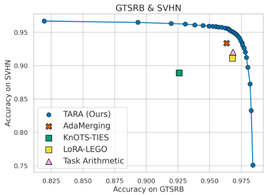

Figure 3. Two-Task Trade-Off

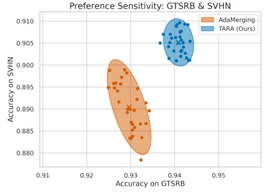

Figure 4. Preference Sensitivity

Experimental Results

- On **8 vision tasks** with CLIP ViT-B/32,TARA-B reaches **76.3%** avg normalized accuracy, outperforming the best vanilla baseline Iso-C and the best LoRA-aware baseline KnOTS-TIES.

Method	Dataset								
	Cars	DTD	EuroSAT	GTSRB	MNIST	RESISC45	SUN397	SVHN	Avg
	Per-task absolute accuracies (%)								
Finetuned	74.0	58.3	99.0	92.7	99.3	88.4	64.5	96.2	84.1
	Per-task accuracies of merged models, normalized to finetuned (%)								
Vanilla Merging									
RegMean [40]	80.2	71.3	37.9	47.3	43.1	70.5	93.9	43.0	60.9
TA [32]	82.1	74.3	48.7	41.8	53.4	71.5	96.6	42.0	63.8
TIES [99]	81.0	72.5	53.8	37.4	69.0	65.3	94.8	45.3	64.9
DARE-TIES [107]	81.6	74.5	51.0	37.2	59.2	66.6	96.0	38.8	63.1
EMR-Merging [31]	82.6	73.9	48.2	40.2	54.6	70.6	95.7	45.9	64.0
FR-Merging [115]	79.9	71.0	24.2	34.2	46.6	65.4	92.2	47.6	57.6
Iso-C [60]	82.0	82.1	55.2	73.0	68.9	80.3	97.7	49.6	73.6
Iso-CTS [60]	80.8	**82.3**	56.2	**74.8**	68.5	79.6	97.4	48.5	73.5
AdaMerging [101]	79.5	73.5	70.9	39.7	63.0	69.0	97.8	66.6	70.0
LoRA-aware Merging									
SVD [84]	81.8	73.5	47.7	38.3	52.8	70.2	96.4	40.4	62.6
Linear [59]	81.4	74.5	49.5	43.0	55.3	70.3	96.8	39.0	63.7
KnOTS-TIES [79]	82.7	73.7	49.3	48.9	68.9	70.9	95.5	53.8	68.0
KnOTS-DARE-TIES [79]	81.8	75.9	50.7	40.3	53.2	70.2	97.9	41.0	63.9
LoRA-LEGO [114]	81.1	73.0	54.4	40.3	68.6	71.5	97.3	37.1	62.9
RobustMerge [109]	80.1	76.8	54.8	38.2	55.4	69.5	96.5	35.2	63.3
TARA-Variant A	82.2	76.0	74.9	43.5	76.3	70.2	98.0	**70.8**	74.0
TARA-Variant B	**86.2**	78.4	**76.8**	42.9	**82.7**	**75.4**	**98.6**	69.7	**76.3**

- On **6 NLI tasks** with LLaMA-3 8B, TARA-B reaches **80.3%** avg normalized accuracy and ranks first on every task.

Method	Dataset						
	MNLI	QNLI	SNLI	RTE	SICK	SCITAIL	Avg
	Per-task absolute accuracies (%)						
Finetuned	90.8	95.3	92.1	84.8	91.3	87.3	90.2
	Per-task accuracies of merged models, normalized to finetuned (%)						
Vanilla Merging							
TA [32]	67.3	87.3	41.8	95.7	77.9	76.9	74.6
TIES [99]	61.3	91.5	38.2	85.5	76.8	70.5	70.6
DARE-TIES [107]	42.0	72.5	44.1	77.8	76.9	86.8	66.7
EMR-Merging [31]	75.3	87.5	41.2	88.0	70.3	73.9	72.7
Iso-C [60]	35.0	57.8	40.5	64.1	51.1	**100.2**	58.1
Iso-CTS [60]	37.1	57.1	39.8	59.4	49.8	95.3	56.4
FR-Merging [115]	65.7	86.9	41.9	93.2	78.3	77.6	73.9
AdaMerging [101]	47.5	92.9	41.3	102.6	93.8	94.2	78.7
LoRA-aware Merging							
SVD [84]	67.7	87.2	41.7	95.7	77.6	76.6	74.4
Linear [59]	63.1	86.2	39.5	90.6	77.5	74.5	71.9
KnOTS-TIES [79]	41.1	83.4	**56.6**	87.2	87.9	94.8	75.2
KnOTS-DARE-TIES [79]	**76.4**	88.2	39.9	99.2	79.6	75.6	76.5
LoRA-LEGO [114]	62.5	84.8	41.4	100.0	87.0	83.7	76.6
RobustMerge [109]	66.4	87.2	41.8	94.0	77.7	77.3	74.1
TARA-Variant A	51.7	92.6	41.4	102.6	95.3	94.4	79.7
TARA-Variant B	46.8	**94.1**	41.4	**103.4**	**98.1**	97.8	**80.3**

- **Two-task trade-off:** TARA traces a smooth Pareto curve dominating baseline operating points.
- **Preference sensitivity:** under randomized ρ, TARA's accuracy cluster is much tighter than AdaMerging's.