Label-Free Cross-Task LoRA Merging with Null-Space Compression

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

CVPR JUNE 3-7, 2026 DENVER COLORADO

Background

- **Model Merging** combines independently fine-tuned checkpoints into a single model without joint multi-task training.
- In the foundation-model era, Low-Rank Adaptation (LoRA) has become the prevalent fine-tuning paradigm, making **LoRA merging** an especially promising target.
- The idea of **entropy minimization** advanced weight-based merging via a surrogate loss that estimates merging coefficients without ground-truth labels.
- Building on AdaMerging, this idea has been widely adopted in subsequent merging methods, yielding consistent performance gains.

Motivation

- Entropy-minimization-based merging has two key limitations:

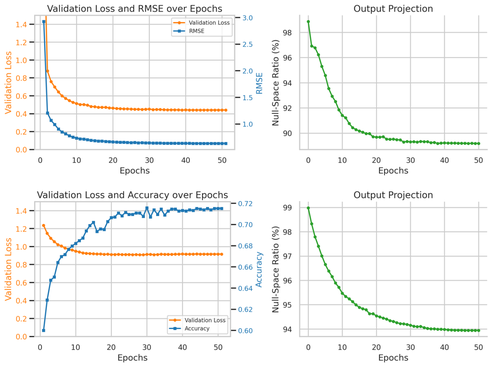
1. Does not apply to regression tasks.
2. Scales poorly to text generation in modern LLMs and VLMs.

Contribution

- We identify **null-space compression**, a structural property of LoRA adapters that strongly correlates with task performance.
- We propose **NSC Merging**: a label-free, output-agnostic, gradient-based method that uniformly handles classification, regression, and sequence generation.
- Achieves state-of-the-art results across 20 heterogeneous vision tasks, 6 NLI benchmarks, and 6 VLM tasks.

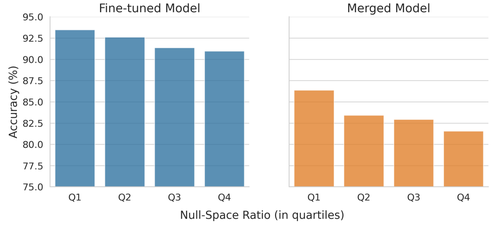
Main Method: NSC Merging

- **Null Space Compression (NSC) during LoRA Fine-tuning.** During LoRA fine-tuning, the down-projection A in ΔW = BA progressively compresses its null space. The fraction of input activations suppressed by A — the null-space ratio — strongly correlates with task performance, and we exploit it as a label-free signal for merging.



Null Space Ratio Formula

$$\omega(\boldsymbol{z}) = \frac{\left\|\mathrm{Proj}_{\mathcal{N}(\boldsymbol{A})}(\boldsymbol{z})\right\|_2}{\|\boldsymbol{z}\|_2} = \sqrt{1 - \frac{\boldsymbol{z}^\top \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} \boldsymbol{A}\boldsymbol{z}}{\|\boldsymbol{z}\|_2^2}}.$$



- **NSC Merging.** We minimize the null-space ratio as a task-agnostic surrogate loss to optimize merging coefficients.

Mean Null Space Ratio

$$\Omega_k(\boldsymbol{x};\boldsymbol{\Theta}) = \frac{1}{|\mathcal{J}|}\sum_{\ell\in\mathcal{J}} \omega_k^\ell\left(\boldsymbol{z}^\ell(\boldsymbol{x};\boldsymbol{\Theta})\right)$$

Final Optimization objective

$$\min_{\{\lambda_k^\ell\}} \frac{1}{K}\sum_{k=1}^{K}\left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{D}_k}\left[\Omega_k(\boldsymbol{x};\boldsymbol{\Theta}_{\text{merge}})\right]\right)$$

$$\text{s.t.}\quad \boldsymbol{\Theta}_{\text{merge}} = \left\{\boldsymbol{W}_0^\ell + \textstyle\sum_k \lambda_k^\ell \boldsymbol{B}_k^\ell \boldsymbol{A}_k^\ell\right\}_{\ell=1}^{L}$$

- **Caching Gram-inverse.** The additional cost of layer-wise null-space-ratio computation can be further reduced by caching the Gram-inverse of the LoRA down-projection factor.

$$\omega_k(\boldsymbol{z}) = \sqrt{1 - \frac{\boldsymbol{z}^\top \boldsymbol{A}_k^\top (\boldsymbol{A}_k \boldsymbol{A}_k^\top)^{-1} \boldsymbol{A}_k \boldsymbol{z}}{\|\boldsymbol{z}\|_2^2}}.$$

Algorithm 1 NSC Merging

Require: Pretrained $\{\boldsymbol{W}_0^\ell\}_{\ell=1}^L$; LoRA $\{(\boldsymbol{B}_k^\ell, \boldsymbol{A}_k^\ell)\}$ for $k{=}1{:}K$, unlabeled validation set $\{\mathcal{D}_k\}$; target layers $\mathcal{J}_{\text{tgt}} \subseteq \mathcal{J}$; steps T; batch size b; learning rate η
Ensure: Coefficients $\{\lambda_k^\ell\}$, merged model $\boldsymbol{\Theta}_{\text{merge}}$
Step 1: Prepare adapter Gram-inverse
for $k = 1$ to K **do**
  for $\ell \in \mathcal{J}_{\text{tgt}}$ **do**
    Compute and cache $\left(\boldsymbol{A}_k^\ell \boldsymbol{A}_k^{\ell\top}\right)^{-1}$
  end for
end for
Step 2: Optimize coefficients
Initialize $\lambda_k^\ell \quad \forall\, k, \ell \in \mathcal{J}$
for $t = 1$ to T **do**
  $\boldsymbol{\Theta}_{\text{merge}} \leftarrow \left\{\boldsymbol{W}_0^\ell + \sum_k \lambda_k^\ell \boldsymbol{B}_k^\ell \boldsymbol{A}_k^\ell\right\}_{\ell=1}^L$
  for $k = 1$ to K **do**
    Sample $\{\boldsymbol{x}_i\}_{i=1}^b \sim \mathcal{D}_k$ ▷ Unlabeled mini-batch
    $\widehat{\mathcal{L}}_k \leftarrow \frac{1}{b}\sum_{i=1}^b \Omega_k(\boldsymbol{x}_i;\boldsymbol{\Theta}_{\text{merge}})$
  end for
  $\widehat{\mathcal{L}} \leftarrow \frac{1}{K}\sum_{k=1}^K \widehat{\mathcal{L}}_k$
  $\lambda_k^\ell \leftarrow \lambda_k^\ell - \eta\, \frac{\partial \widehat{\mathcal{L}}}{\partial \lambda_k^\ell} \quad \forall\, k, \ell \in \mathcal{J}$
end for
return $\{\lambda_k^\ell\}, \boldsymbol{\Theta}_{\text{merge}}$

Experimental Results

- **20 Heterogeneous Vision Tasks.** Evaluated on 20 tasks spanning NYUD-v2 (4), PASCAL-Context (5), and Taskonomy (11), spanning dense predictions including semantic segmentation, depth, and normals. We use a ViT-B backbone with task-specific decoders.

Method	NYUD-v2 (4)				PASCAL-Context (5)					Taskonomy (11)											Avg
	Depth RMSE ↓	Semseg mIoU ↑	Normal Mean ↓	Normal L1 ↑	Semseg mIoU ↑	Parts mIoU ↑	Saliency mIoU ↑	Normal Mean ↓	Edge L1 ↑	DE L1 ↓	DZ L1 ↓	EO L1 ↓	ET L1 ↓	K2 L1 ↓	K3 L1 ↓	N L1 ↓	C RMSE ↓	R L1 ↓	S2 L1 ↓	S2.5 L1 ↓	
Finetuned performance																					
Finetuned	0.657	37.66	25.98	0.051	70.07	54.77	80.00	18.36	0.046	0.016	0.016	0.101	0.171	0.162	0.082	0.217	0.710	0.136	0.170	0.144	–
Merged models, normalized to finetuned (%)																					
TA	24.0	1.3	54.9	105.2	4.7	20.6	36.4	62.5	104.3	100.1	99.8	101.7	104.6	106.4	102.2	100.2	103.6	106.2	107.4	98.8	77.2
TIES	28.5	1.3	55.2	105.1	4.7	20.6	34.6	66.1	104.3	102.5	102.0	100.6	103.7	105.4	99.5	100.1	102.4	105.2	104.3	99.5	77.3
KnOTS-TIES	19.4	1.5	54.5	105.2	4.7	20.6	38.3	61.2	104.3	100.1	99.6	101.3	102.8	105.2	102.4	100.1	102.7	104.9	104.8	97.9	76.6
RobustMerge	37.2	76.5	63.4	100.1	83.1	73.7	85.7	73.7	100.6	100.0	101.0	102.2	100.6	100.5	100.6	100.0	100.2	99.7	99.1	99.8	89.9
NSC (Ours)	45.9	85.1	69.6	100.4	86.7	76.7	88.7	80.9	101.7	100.1	101.0	102.2	100.8	100.8	100.7	100.0	100.4	100.0	99.6	99.6	**92.0**

- **6 VLM Tasks.** Evaluated on 6 VQA and image-captioning tasks using LoRA fine-tuned LLaVA-1.5-7B.

Method	IconQA	VizWiz$_{\text{val}}$	ChartQA	DocVQA$_{\text{val}}$	COCO	Flickr30k	Avg
Per-task absolute score							
Zero-Shot	17.9	55.2	18.2	24.3	109.5	79.2	–
Finetuned	67.8	69.3	39.0	40.7	130.5	91.3	–
Avg. generated tokens	2.1	2.8	4.7	6.7	12.5	14.2	7.17
Merged models (normalized to fine-tuned baselines, %)							
TA	56.8	83.2	73.0	85.2	92.9	97.6	81.4
TIES	61.2	81.6	76.1	82.3	92.9	96.7	81.8
KnOTS-TIES	49.8	84.1	70.4	85.5	92.3	97.2	79.9
RobustMerge	82.3	58.2	71.1	82.1	93.6	96.9	80.7
AdaMerging (Single Token)	64.7	76.0	76.4	82.1	87.5	98.4	80.9
AdaMerging (Full Token)	68.7	76.2	77.9	85.8	91.1	94.4	82.4
NSC (Ours)	59.7	82.9	78.1	87.1	91.7	96.8	**82.7**

- **6 NLI Tasks.**

Method	MNLI	QNLI	SNLI	RTE	SICK	SciTail	Avg
Per-task absolute accuracy (%)							
Finetuned	90.8	94.9	91.8	87.0	90.9	94.8	91.7
Merged models (normalized to finetuned baselines, %)							
TA	92.8	86.8	93.3	93.6	83.8	95.0	90.9
TIES	94.3	88.8	90.8	89.8	86.6	94.4	90.8
KnOTS-TIES	92.0	82.0	94.9	92.1	80.2	95.3	89.4
RobustMerge	94.3	88.1	93.7	93.6	83.0	94.5	91.2
AdaMerging	94.3	84.8	92.5	92.1	89.2	84.8	89.6
NSC (Ours)	94.9	88.3	92.8	91.3	91.2	95.1	**92.3**

- **Computational Cost**

Method	Requirements Prep.	Opt.	Val.	Prep. (min)	Opt. (min)	Val. (min)	*Total* (min)	*GPU Mem.* (GB)
Gradient-Free								
TA [30]	✗	✗	✓	–	–	16.6	16.6	**14.4**
TIES [90]	✓	✗	✓	4.5	–	16.6	20.1	**14.4**
KnOTS-TIES [71]	✓	✗	✓	6.8	–	16.6	23.4	**14.4**
Gradient-Based								
AdaMerging [91] (Single Token)	✗	✓	✗	–	13.3	–	**13.3**	18.0
AdaMerging (Full Token)	✗	✓	✗	–	104.0	–	104.0	18.0
NSC (Ours)	✓	✓	✗	≈0.0	13.3	–	**13.3**	17.9

Summary

The null-space ratio of LoRA adapters is a label-free, output-agnostic signal that generalizes across classification, regression, and sequence generation. We propose **NSC Merging**, the first model merging algorithm to exploit this phenomenon.