

PRISM



Spectral Pruning and Reconstruction for Parameter-Efficient Model Merging of MLLMs

Merging diverse LoRA expertise without retraining

Wuxuan Shi, Haotian Chen, He Li, Qiang Yang, Mang Ye

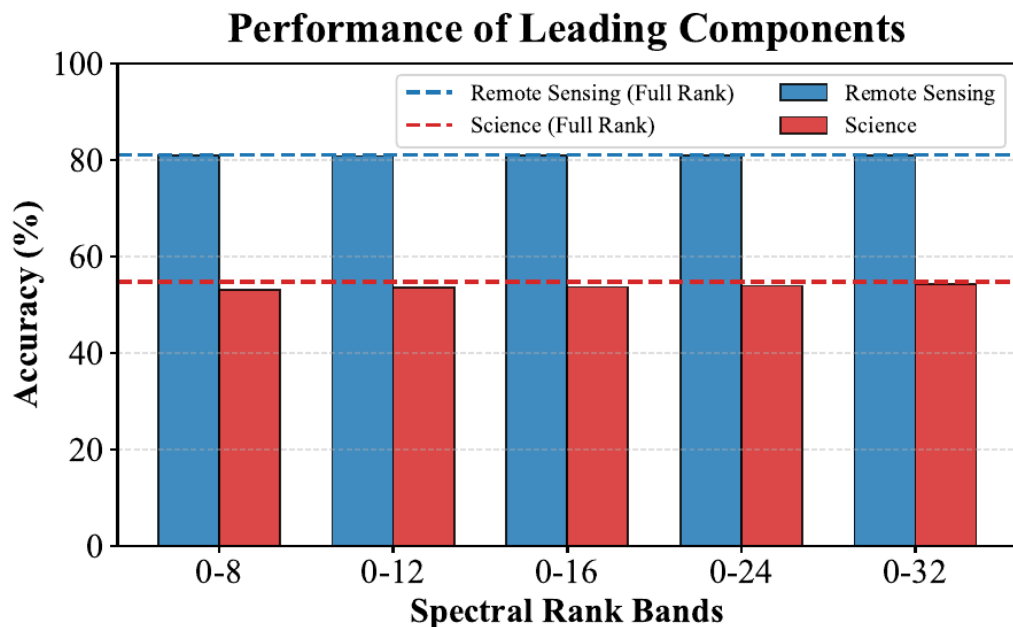
Wuhan University

The Hong Kong Polytechnic University



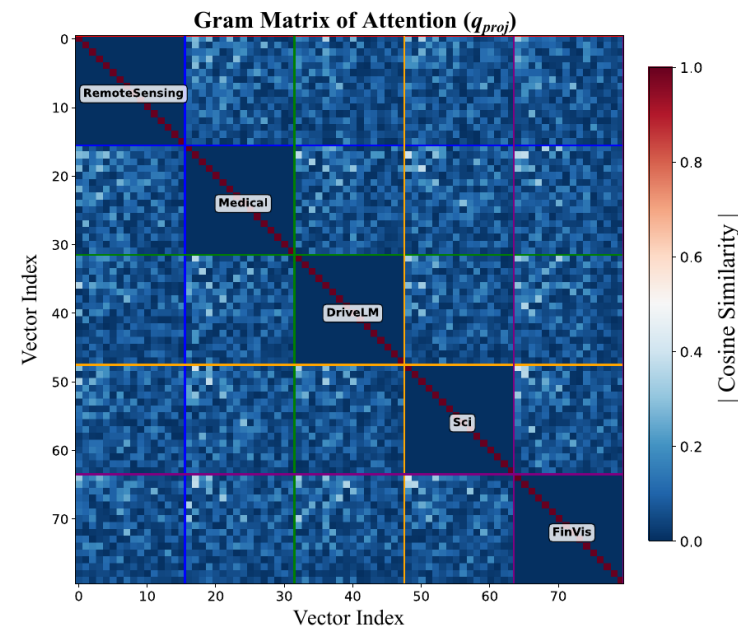
Two spectral observations motivate PRISM

Goal: combine task-specific experts in one LoRA adapter



Concentrated task knowledge

A few leading components retain most task-relevant information.

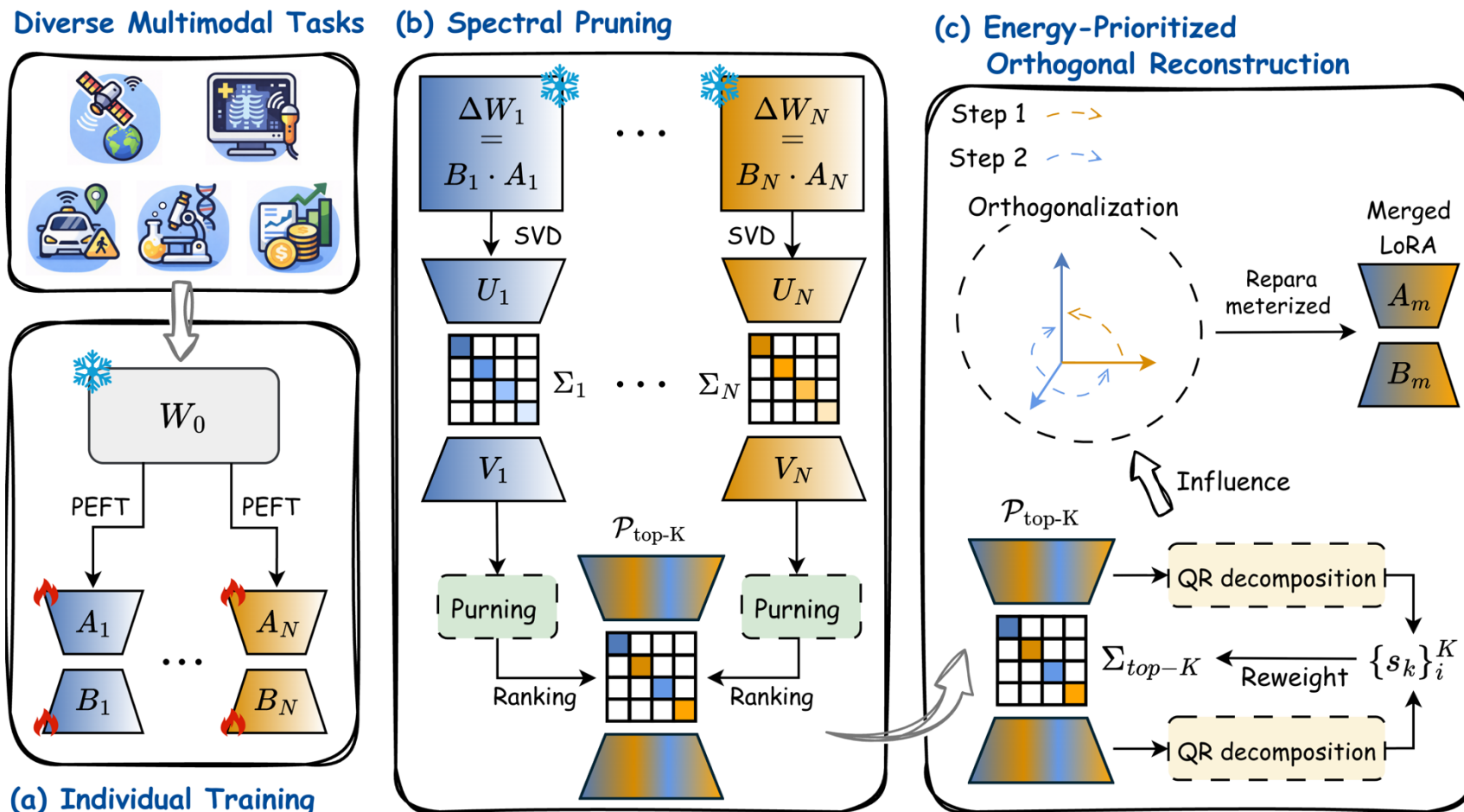


Structured overlap across experts

Strong directions can interfere when merged directly.

PRISM merges experts in the spectral domain

Select high-energy components, then resolve their overlap



Energy-prioritized orthogonal reconstruction

Largest singular values receive priority

Order the retained components from highest to lowest energy.

QR separates overlapping directions in both singular-vector spaces.

Residual scales suppress components that repeat earlier subspaces.

No training data and no learned merging parameters

Ordered QR

$$U_{\text{cat}} = Q_U R_U, \quad V_{\text{cat}} = Q_V R_V$$

Surviving contribution

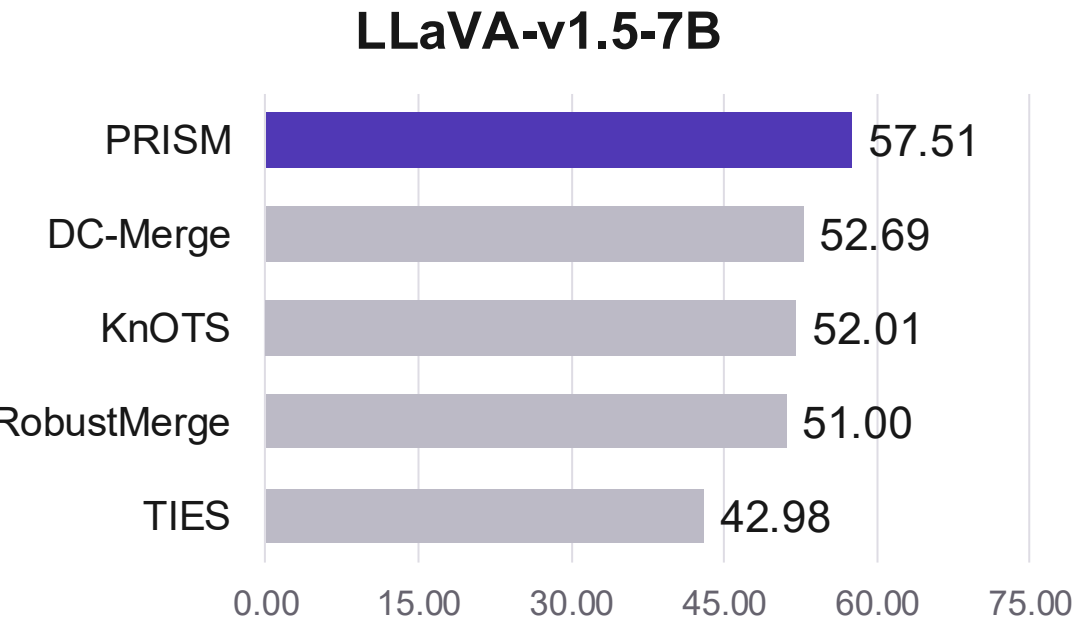
$$s_k = |(R_U)_{kk}| \cdot |(R_V)_{kk}|$$

Merged update

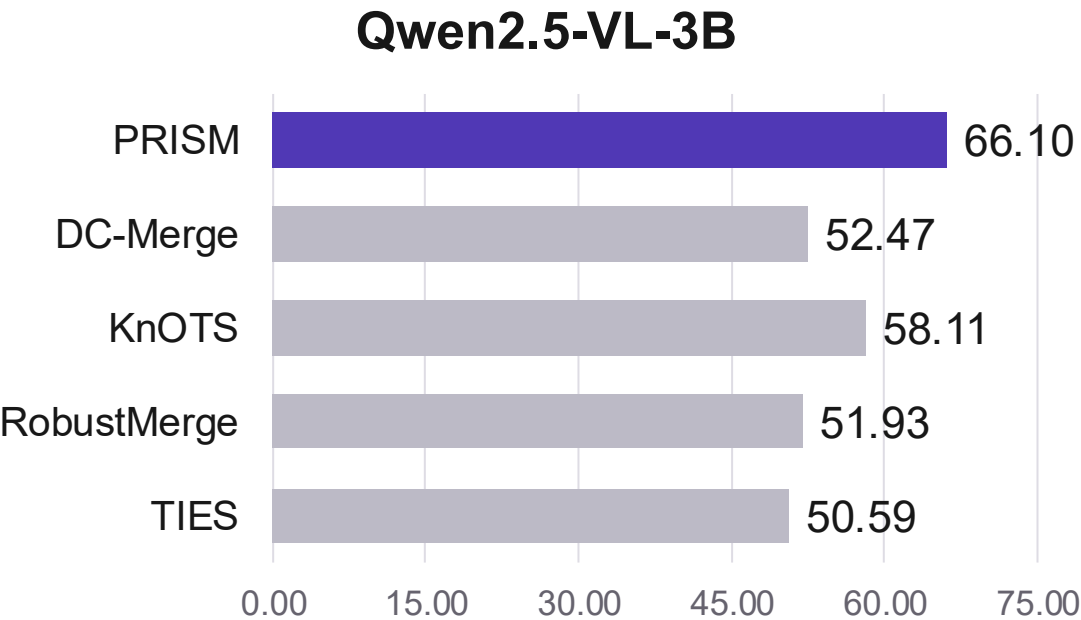
$$\Delta W_{\text{merge}} = Q_U \Sigma_{\text{new}} Q_V^{\top}$$

PRISM improves merging across both backbones

Multi-domain average accuracy (%), selected baselines at equal merged rank



+4.82 pp
over DC-Merge



+7.99 pp
over KnOTS

Compression and ablation

Compressed rank budgets

Backbone	Merged / expert rank	Avg. acc. (%)
LLaVA 7B	32 / 128	55.45
Qwen 3B	8 / 16	65.14

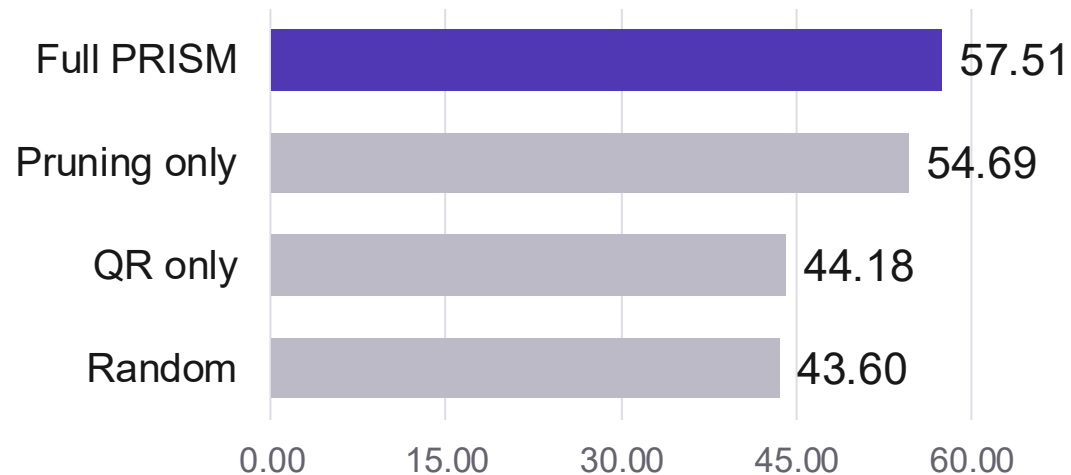
25% rank on LLaVA

50% rank on Qwen

Both outperform every full-rank multi-domain baseline.

Both stages contribute

LLaVA multi-domain accuracy (%)



+11.09 pp from spectral pruning

+2.82 pp after reconstruction

Spectral structure supports better LoRA merging

Preserve dominant task signals

Global singular-value selection uses a prescribed rank budget.

Reduce overlap between experts

Energy-ordered reconstruction produces one merged LoRA adapter.

Current scope: LoRA experts trained from a shared backbone.

Thank you

github.com/ShiWuxuan/PRISM