ICML 2025 Rebuttal of CoTo (Supplement Figures and Tables)

Figure r1

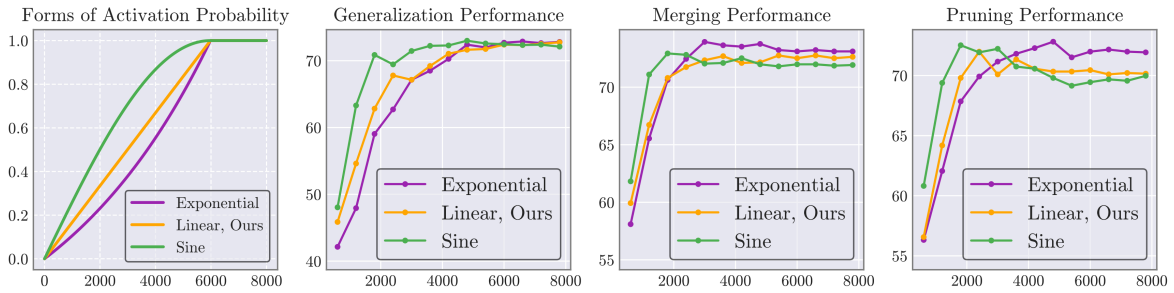

Figure r1. Generalization, merging, and pruning performance over training steps with different activation probability schedules on the *DTD* task. Merging performance is evaluated using a 1:1 LoRA merging ratio, and pruning follows a "prune every other layer" strategy.

Figure r2

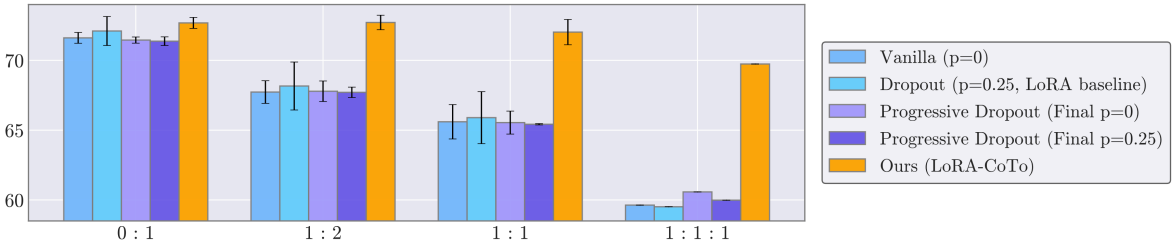

Figure r2. Merging performance under different strategies: no Dropout, Dropout, Progressive Dropout variants, and CoTo. The x-axis shows the merging ratio, where 0:1 denotes no merging and 1:1:1 indicates linear fusion of three LoRAs trained with different seeds.

Figure r3

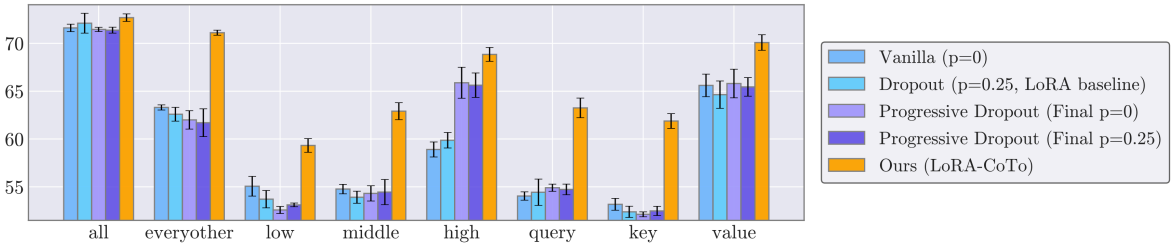

Figure r3. Pruning performance under different strategies: no Dropout, Dropout, Progressive Dropout variants, and CoTo. "All" denotes no pruning, "everyother" removes alternate LoRA layers, and other labels indicate the proportion of parameters retained.

Figure r4

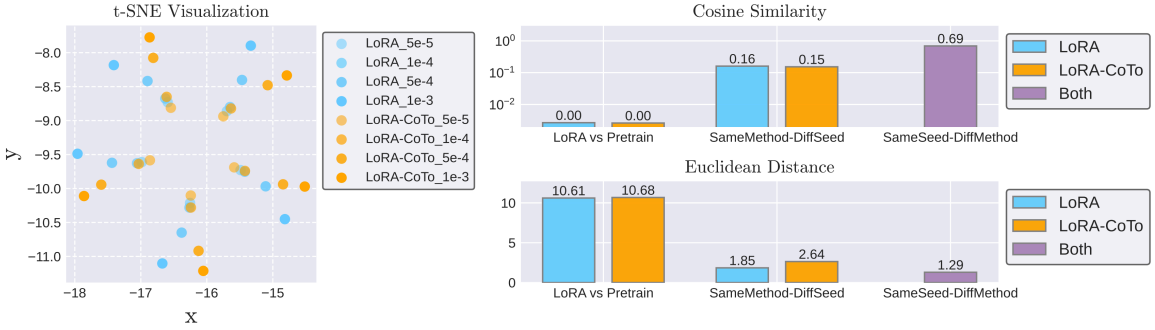

Figure r4. (Left) Visualization of LoRA weights across different seeds and learning rates. (right) Averaged cosine similarity and Euclidean distance between (i) LoRA and pre-trained weights, (ii) same method with different seeds, and (iii) same seed with different methods.

Figure r5

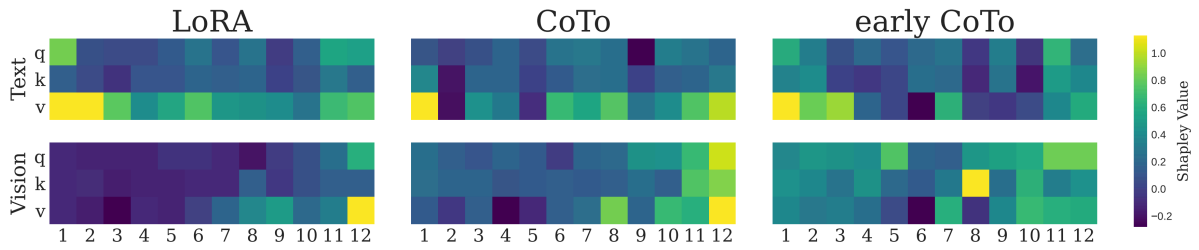

Figure r5. Approximated Shapley values of adapters across layers, with a unified colorbar for consistent comparison.

Figure r6

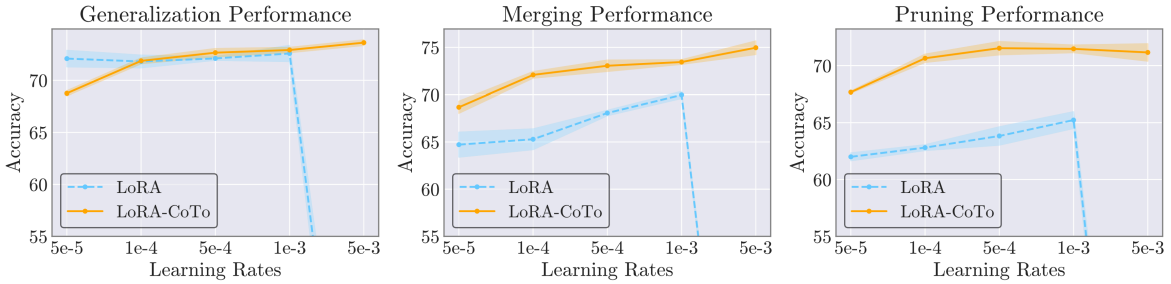

Figure r6. Generalization, merging, and pruning performance of LoRA and LoRA-CoTo across different learning rates on the *DTD* classification task. Merging is evaluated using a 1:1 LoRA merging ratio, and pruning follows a "prune every other layer" strategy.

Figure r7

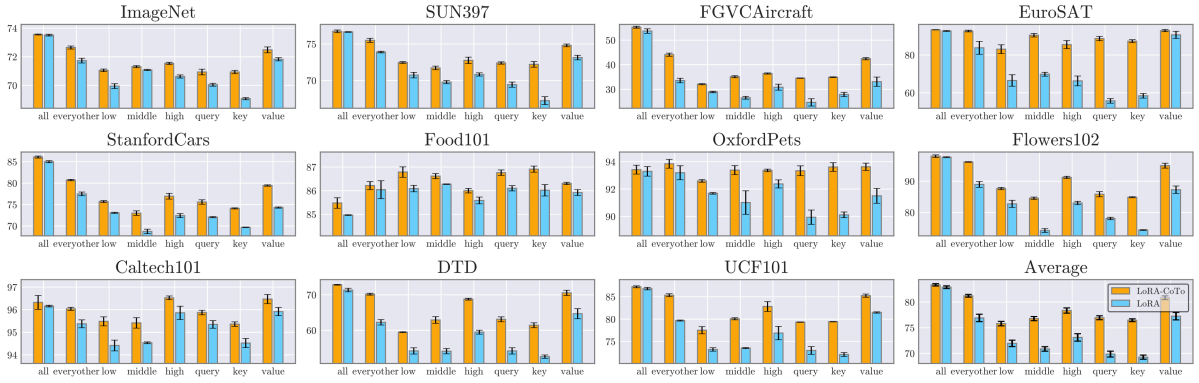

Figure r7. Structured pruning results across the complete CLIP dataset.

Figure r8

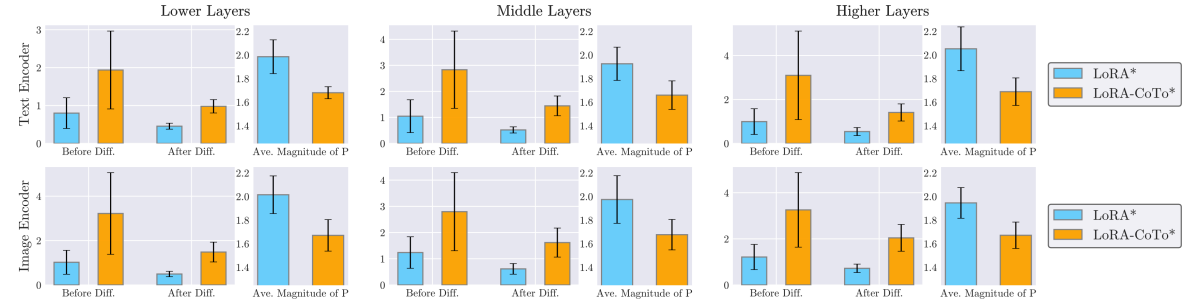

Figure r8. Analysis of the optimized alignment matrix P for LoRA* and LoRA-CoTo*. Results are reported for each encoder (text and image) and layer group (lower, middle, higher), showing the effect of P on alignment differences $\|\Delta W_f - \Delta W_m\|_2$ and its magnitude.

Table r1:

Table r1. Multi-task performance when merging LoRA fine-tuned models on six GLUE tasks using the DeBERTa-v3 backbone.

	Learning Rates	Method	CoLA	QNLI	MRPC	QQP	RTE	SST2	Average	Diff.
Without CoTo	5e-4 (default)	Task LoRA	87.44	94.31	89.46	91.05	85.56	95.18	90.50	
		Linear Fusion	69.89	49.97	68.38	65.50	47.29	55.05	59.35	
		LoRA-LEGO	73.28±3.52	74.19±4.94	33.15±2.25	80.95±3.56	61.46±3.88	69.18±10.43	65.37±2.82	
	1e-3	Task LoRA	86.48	93.87	88.48	91.16	84.12	94.95	89.84	
		Linear Fusion	69.13	49.46	68.38	63.18	47.29	50.92	58.06	
		LoRA-LEGO	73.54±3.65	79.54±5.57	69.32±15.37	78.67±7.27	51.99±3.48	54.13±1.81	67.86±2.70	
With CoTo	5e-4	Task LoRA	86.48	93.94	89.22	90.12	82.67	94.95	89.56	-0.94
		Linear Fusion	69.22	67.12	68.38	70.32	47.29	58.14	63.41	+4.06
		LoRA-LEGO	53.75±9.76	76.19±9.18	70.77±6.21	75.26±6.33	64.31±4.20	89.01±2.51	71.55±1.48	+6.18
	1e-3	Task LoRA	86.58	93.92	89.95	90.49	83.03	95.18	89.86	+0.02
		Linear Fusion	73.15	51.91	68.38	69.96	49.10	63.30	62.63	+4.57
		LoRA-LEGO	71.84±5.36	73.15±9.18	72.39±1.53	72.95±9.48	63.51±2.30	87.24±3.15	73.51±1.46	+5.65

Table r2:

Table r2. Comparison of LoRA and HiRA with and without CoTo on customized image generation tasks, evaluated using Fréchet Inception Distance (FID) and LPIPS similarity loss between generated and real images. Metrics are reported separately for object and style, with lower values indicating better similarity (though these metrics may not fully capture the semantic fidelity of object-style fusion).

	Object	Style	Average
Fréchet Inception Distance (FID) (↓)			
LoRA	31.34	40.03	35.69
LoRA-CoTo	30.02	39.64	34.83
HiRA	35.67	19.65	27.66
HiRA-CoTo	32.87	22.25	27.56
LPIPS (↓)			
LoRA	0.657	0.559	0.608
LoRA-CoTo	0.657	0.554	0.606
HiRA	0.627	0.555	0.591
HiRA-CoTo	0.636	0.559	0.597