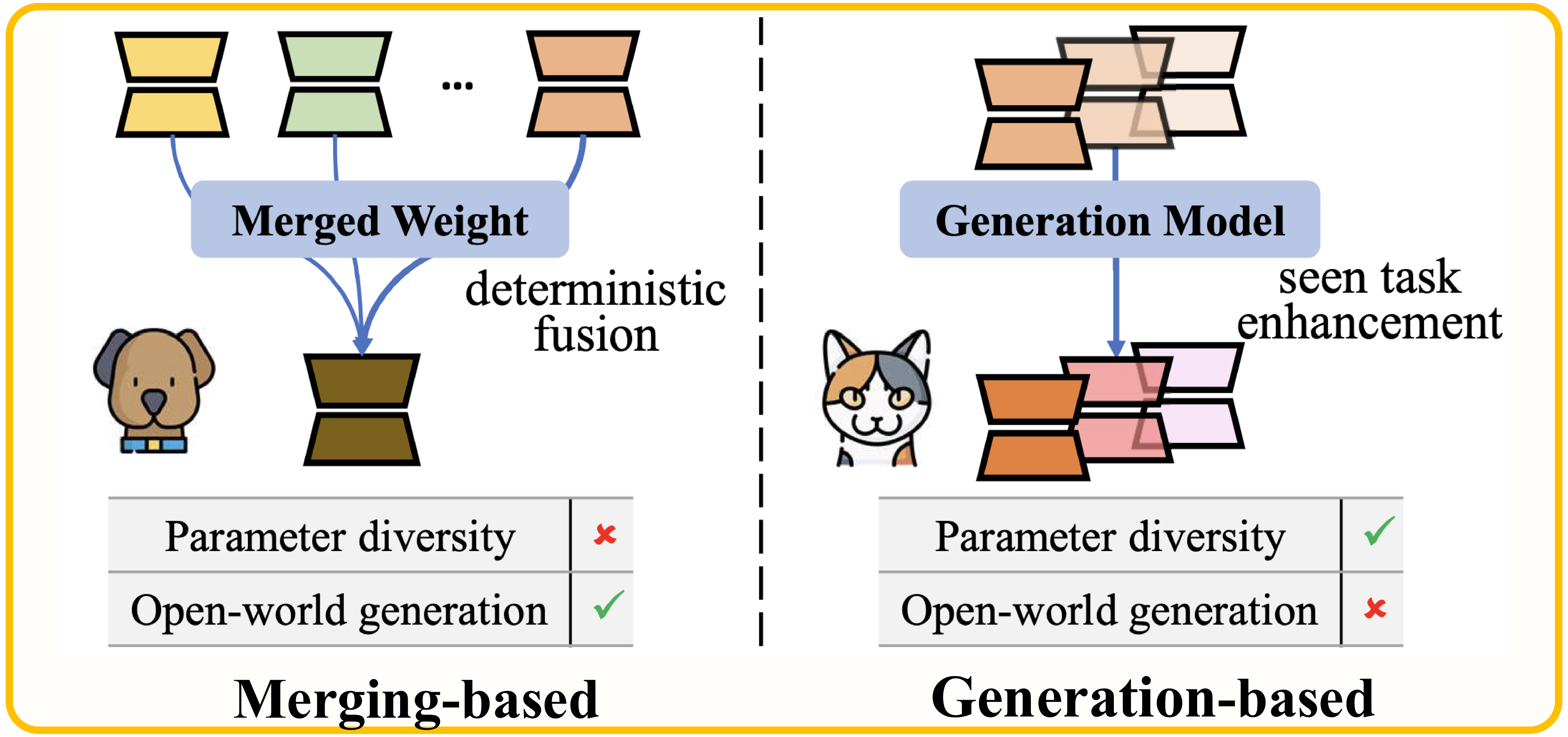
CVPR
JUNE 3-7, 2026
DENVER
COLORADO

SG-LoRA: Semantic-guided LoRA Parameters Generation

Miaoge Li, Yang Chen, Zhijie Rao, Can Jiang, Kang Wei, Jingcai Guo
The Hong Kong Polytechnic University, Hong Kong SAR & Xidian University, China & Southeast University, China

Motivation



<u>Model Merging</u>: Integrates parameter-level knowledge from multiple independently trained networks into a unified model, achieving enhanced capabilities.

<u>Struggle to explore the diversity of LoRA parameters.</u>

<u>Neural Network Parameters Generation</u>: Leverages generative models to directly synthesize new LoRA parameters.

<u>Primarily focus on LoRA enhancement for seen task and struggle to handle open-world tasks</u>

Problem Definition

- No raw data is available for the unseen task, highlighting the need for rapid adaptability to evolving user intents ;
- Open-world task coverage, defined by a broad and unconstrained task space in which the unseen tasks may not be directly related to the seen tasks.

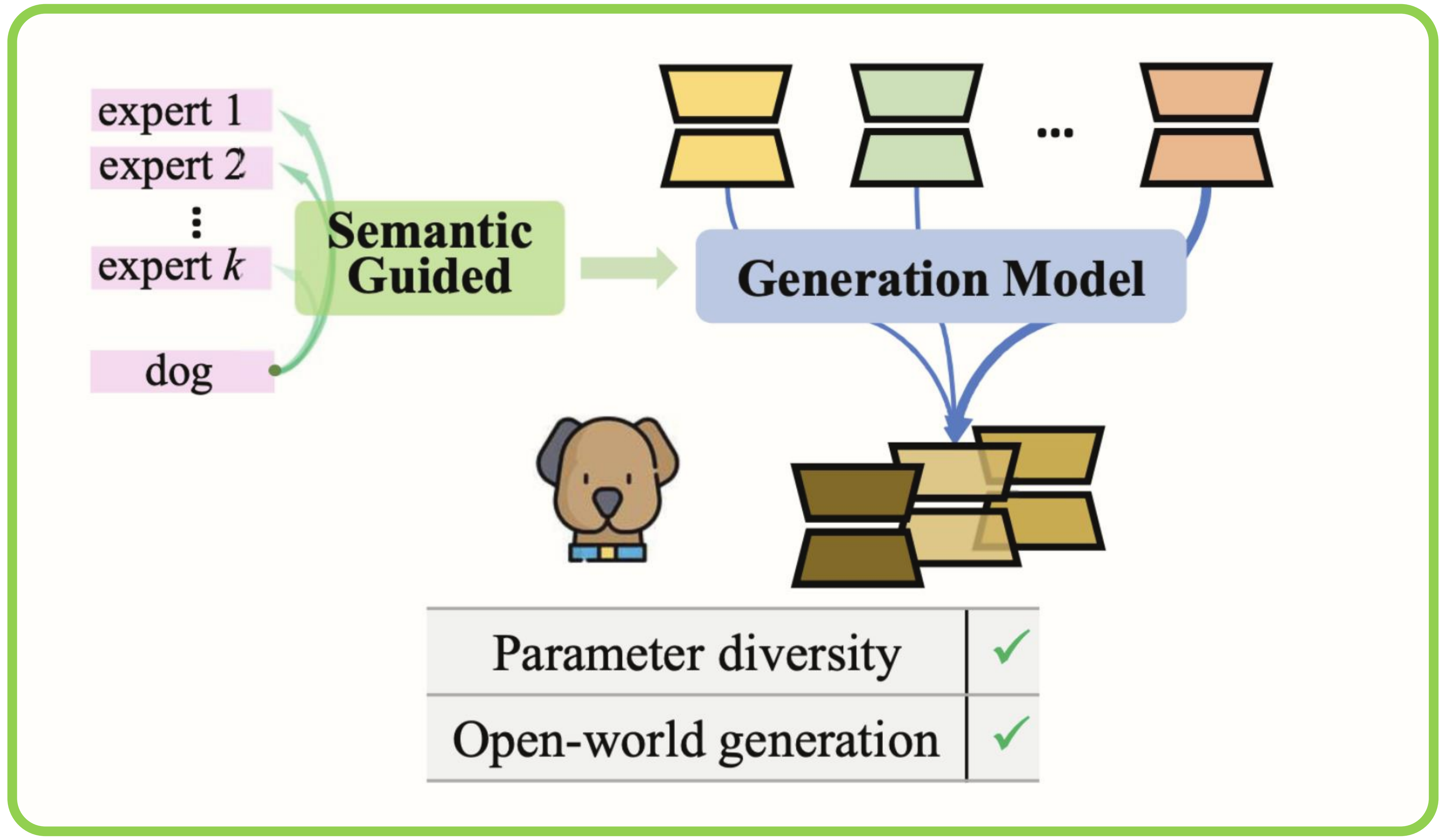
$$\mathcal{W}^* = G\left(f(\mathcal{T}^*), \mathcal{W}, f(\mathcal{T})\right)$$

Our Proposed SG-LoRA

① *Expert Repository Formation*

$$\mathcal{W}_{\text{expert}} = \{(\boldsymbol{\mu}_e, \mathbf{d}_e) \mid e \in \mathcal{E}\}, \quad \boldsymbol{\mu}_e = \frac{1}{M}\Delta \boldsymbol{W}_e$$

Compact yet expressive basis for capturing the essential characteristics of each knowledge domain.



③ *Conditional LoRA Parameter Generation*

$$\mathcal{L}_{\text{CVAE}} = \mathbb{E}_{q(\boldsymbol{z}|\boldsymbol{X},c)}\left[\|\boldsymbol{X} - \hat{\boldsymbol{X}}\|^2\right] + \lambda \cdot KL(q(\boldsymbol{z}|\boldsymbol{X},c)\|p(\boldsymbol{z}|c))$$

Encourages the decoder to reconstruct accurate LoRA parameters, while regularizes the latent space to align the task-specific prior.

② *Task Semantics Construction*

Not all experts contribute equally

Identify and prioritize the most beneficial experts

$$\left.\begin{aligned}
\alpha_k &= \frac{\exp\left(\text{sim}(\mathbf{d}^*, \mathbf{d}_k)/\tau\right)}{\sum_{k' \in \mathcal{I}_{\text{top-}k}} \exp\left(\text{sim}(\mathbf{d}^*, \mathbf{d}_{k'})/\tau\right)}, \quad k \in \mathcal{I}_{\text{top-}K} \\
\boldsymbol{\mu}^* &= \sum_{k \in \mathcal{I}_{\text{top-}K}} \alpha_k \cdot \boldsymbol{\mu}_k \\
\boldsymbol{\sigma}^{*2} &= \sum_{k=1}^{K} \alpha_k \boldsymbol{\sigma}_k^2 + \sum_{k=1}^{K} \alpha_k (\boldsymbol{\mu}_k - \boldsymbol{\mu}^*) \odot (\boldsymbol{\mu}_k - \boldsymbol{\mu}^*)
\end{aligned}\right\} c = \{\boldsymbol{\mu}^*, \boldsymbol{\sigma}^{*2}\}$$

Capture task-level correlations without exposing user-specific data.

Experimental Results

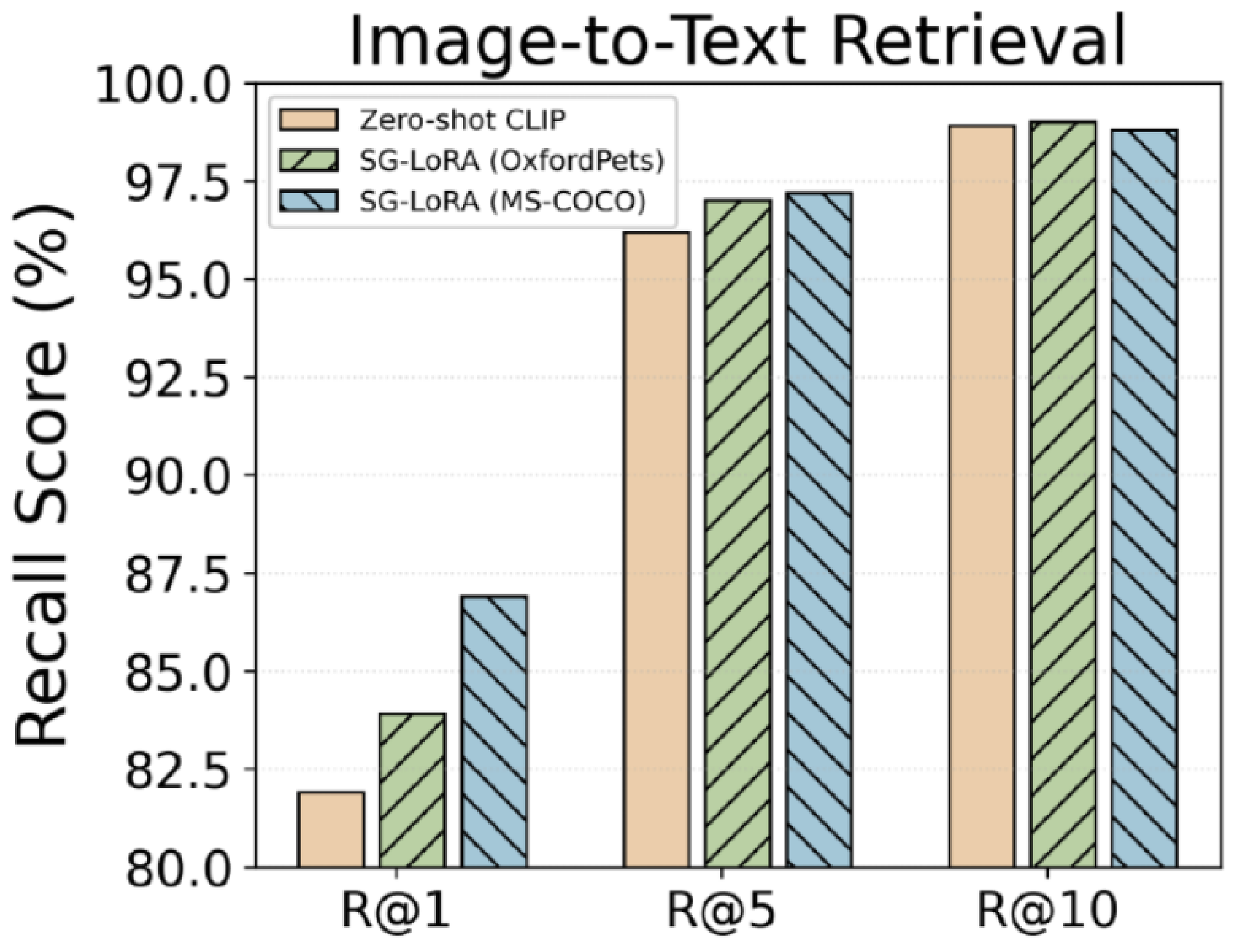
- ***In-Dataset Evaluations***

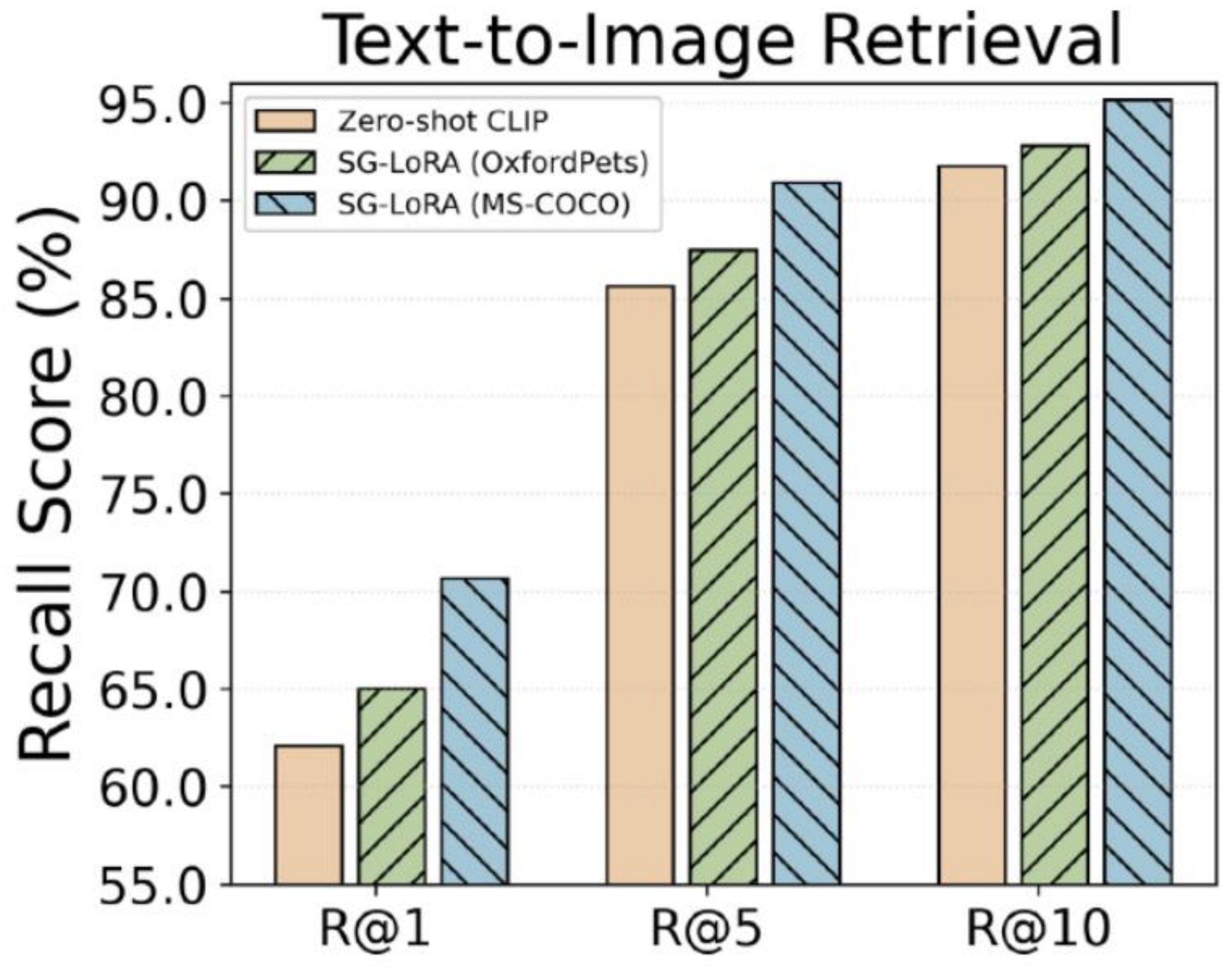
Method	MS-COCO						OxfordPets					
	I2T			T2I			I2T			T2I		
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
Zero-Shot CLIP	66.43	84.31	89.14	41.66	64.63	73.01	40.45	66.27	77.53	26.03	50.66	62.98
Oracle	72.45	88.91	93.41	53.10	76.47	83.97	55.84	81.84	89.13	40.99	70.41	80.39
Model Soups	69.37	85.96	90.95	47.38	69.54	77.97	52.54	77.80	85.59	33.51	61.77	72.93
AdapterSoup	70.70	86.57	91.09	48.64	70.51	78.79	52.59	78.52	86.09	34.05	62.70	73.93
Top-K LoRA Weighted	71.55	87.54	91.69	49.85	71.79	79.66	53.96	79.41	86.53	35.42	64.08	74.99
SG-LoRA	**74.31**	**88.78**	**92.50**	**54.42**	**75.45**	**82.18**	**57.15**	**80.40**	**88.04**	**37.62**	**67.16**	**77.44**

- ***Cross-Dataset Evaluations***

Method	MS-COCO → OxfordPets						OxfordPets → MS-COCO					
	I2T			T2I			I2T			T2I		
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
Zero-Shot CLIP	40.45	66.27	77.53	26.03	50.66	62.98	66.43	84.31	89.14	41.66	64.63	73.01
Oracle	55.84	81.84	89.13	40.99	70.41	80.39	72.45	88.91	93.41	53.10	76.47	83.97
Model Soups	44.67	70.91	80.78	30.45	56.77	68.52	68.58	85.67	90.62	44.09	66.55	75.08
AdapterSoup	45.96	71.83	81.42	30.88	57.32	69.08	68.74	85.83	90.63	44.19	66.58	75.31
Top-K LoRA Weighted	48.13	73.43	82.73	33.34	59.53	70.89	68.75	85.77	90.67	44.60	66.76	75.25
SG-LoRA	**55.41**	**80.73**	**87.33**	**38.84**	**66.77**	**76.69**	**70.81**	**86.83**	**91.41**	**46.50**	**68.73**	**77.19**

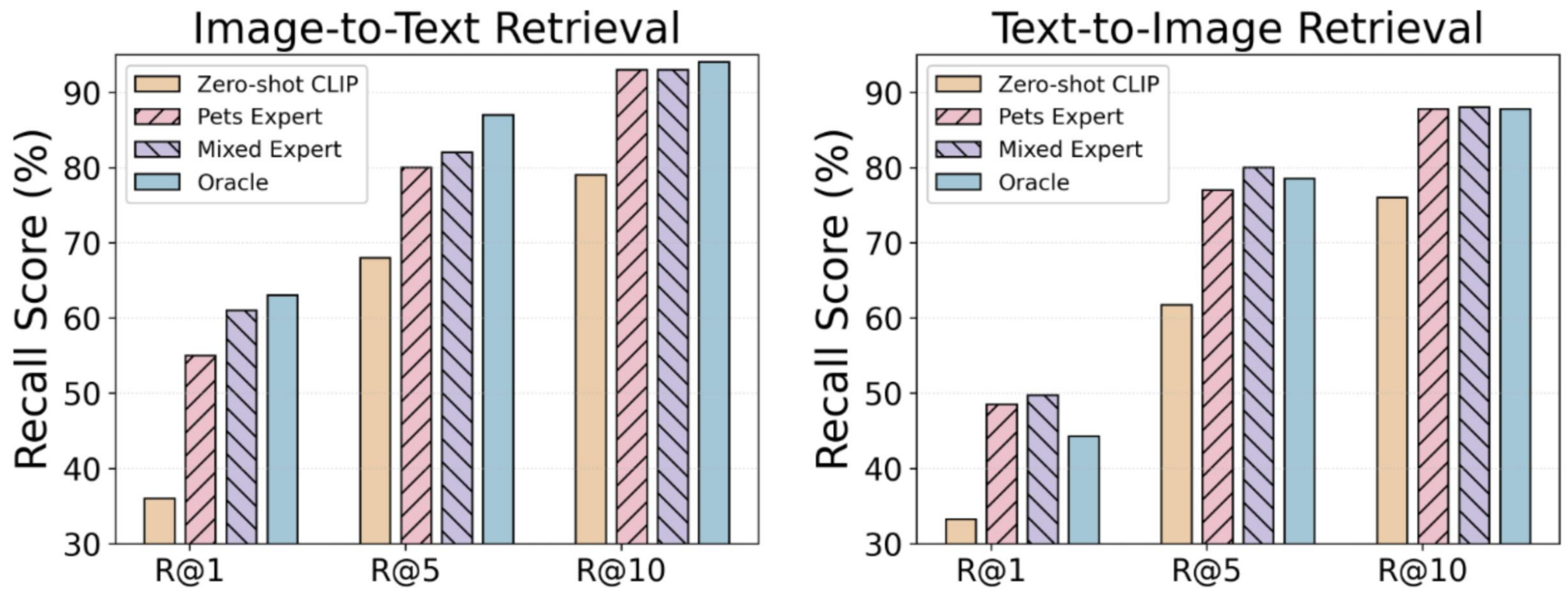
- ***Evaluation on Standard Image-Text Retrieval***





Ablation Study

- ***Comparison on expert repository configuration***



Comparison on expert repository configuration: single-source experts vs. mixed-source experts.

Expert strategy	*Egyptian Mau* **I2T**			*Egyptian Mau* **T2I**		
	R@1	**R@5**	**R@10**	**R@1**	**R@5**	**R@10**
w/o *Cat* expert	36.08	62.89	71.13	15.21	31.70	44.07
w/ *Cat* expert	37.11	63.92	72.16	15.21	35.05	46.91

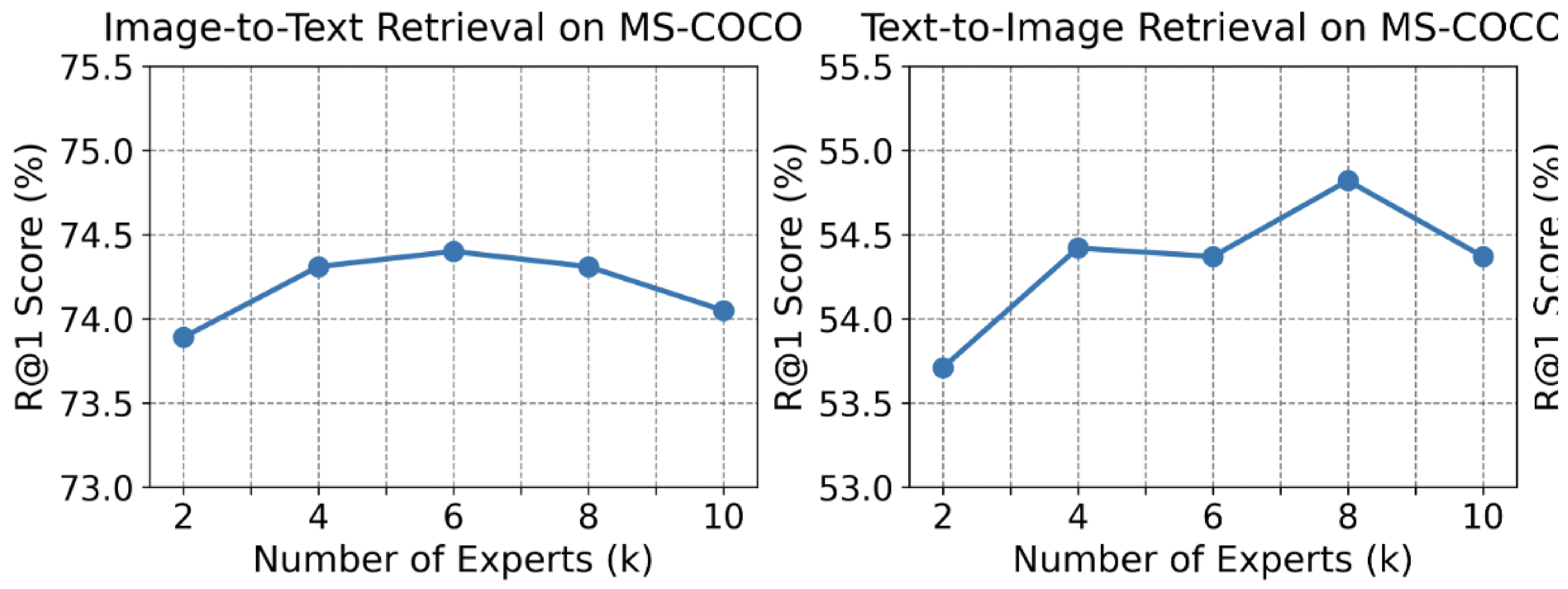
Ablation on expert repository strategy for cross-dataset evaluation. We evaluate how the MS-COCO Cat expert affects retrieval on unseen OxfordPets cat tasks.

- ***Modalities of Semantic Prior Condition***

Condition	**Metrics**		**Dataset**
	I2T R@1	T2I R@1	
Visual	73.16	52.70	MS-COCO
Textual	74.31	54.42	
Visual	86.30	70.12	Flickr30K
Textual	86.90	70.66	

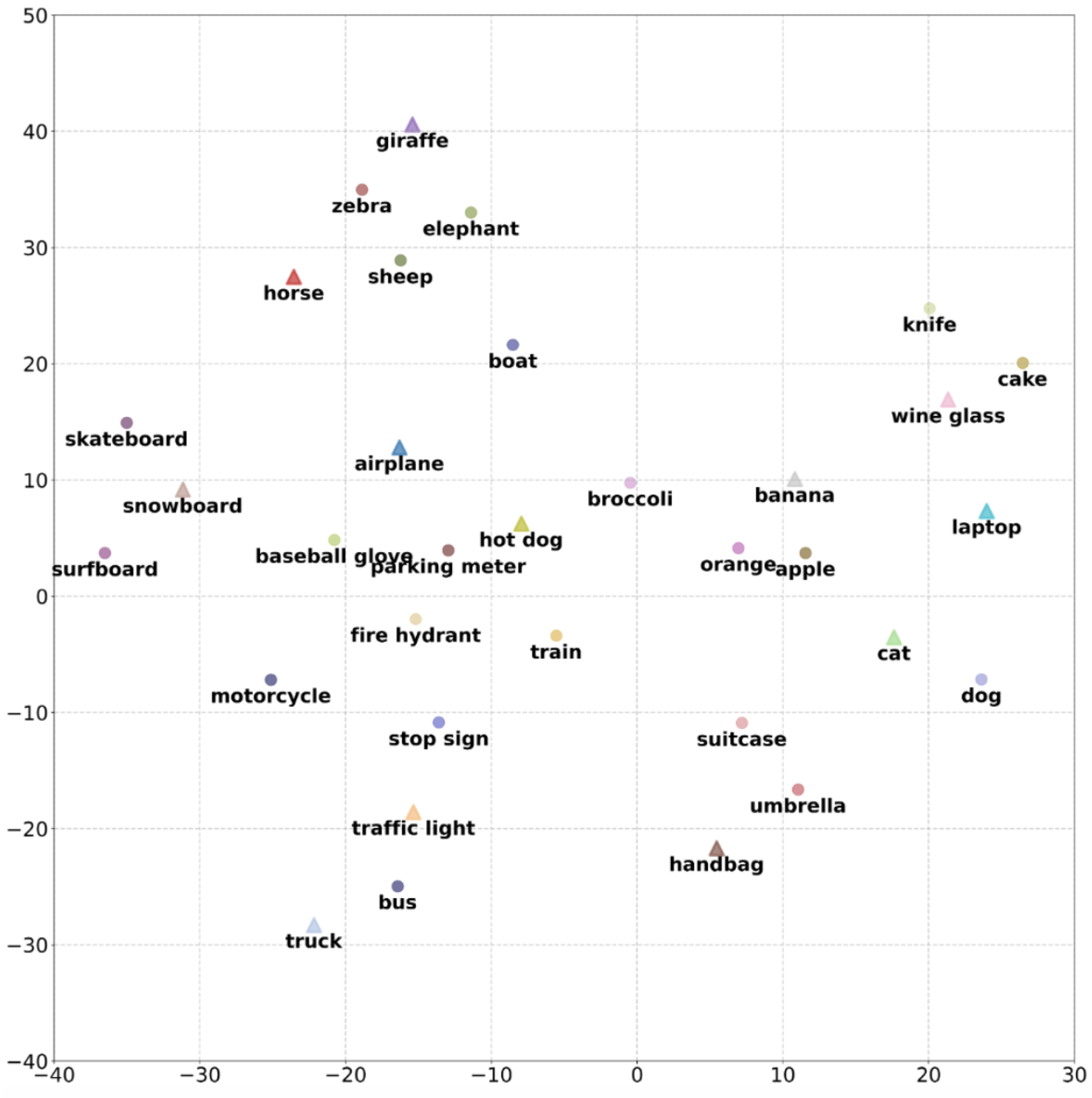
Ablation study on modalities of semantic prior condition.

- ***Number of experts***

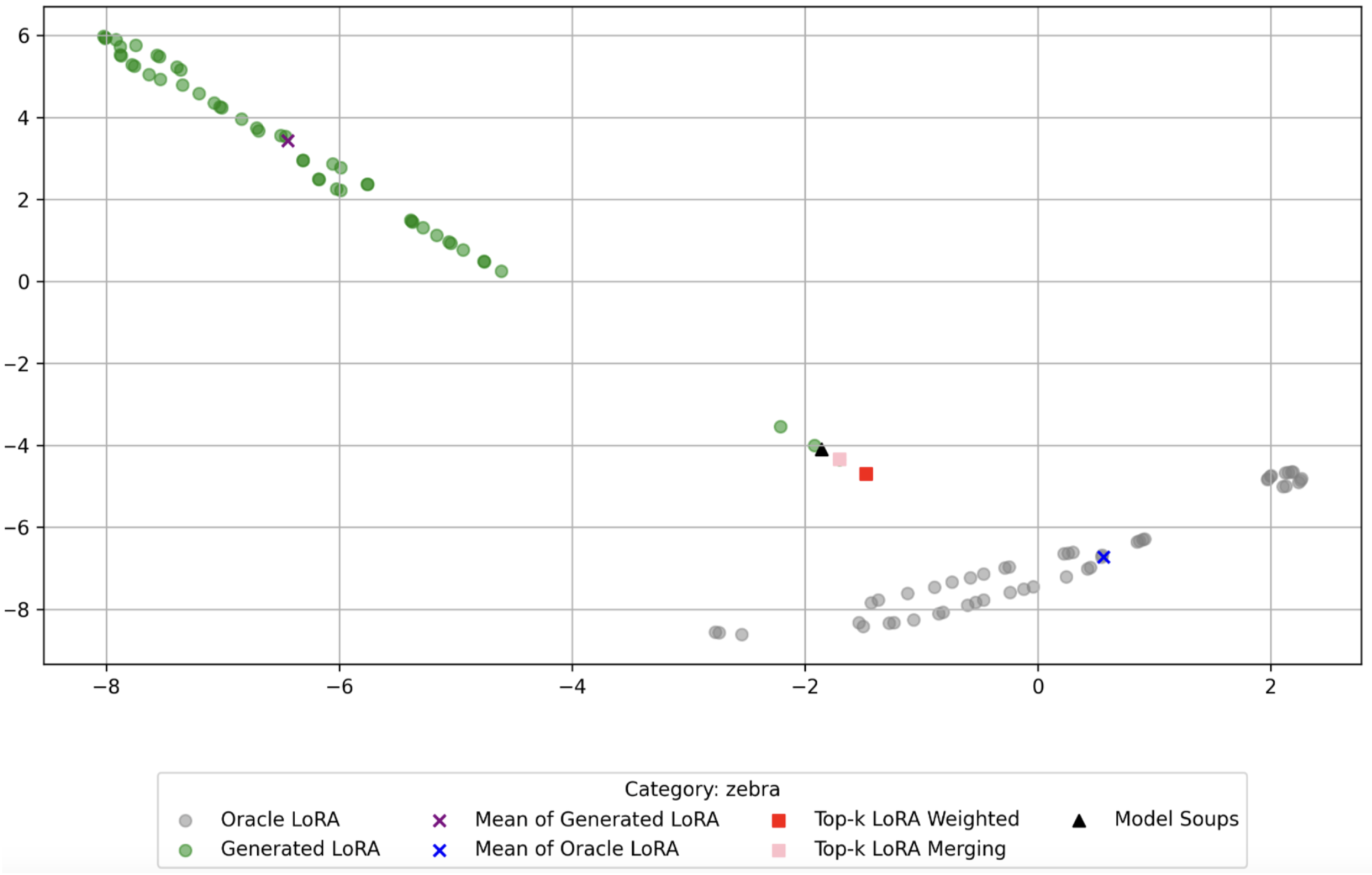


Ablaton study on the number of experts.

Visualization Analysis



Semantically similar LoRA parameters tend to cluster closely together.



The distribution of LoRAs generated by SG-LoRA (in green) exhibits diversity.