BiVaR: Bidirectional Variational Discrimination for Cross-Model Reasoning Transfer

Anonymous authors
Paper under double-blind review

Abstract

Vision-language models (VLMs) inherit their language backbone from pretrained large language models (LLMs), yet still fall behind them on tasks requiring complex multi-step reasoning. Training-free reasoning transfer via model merging avoids costly multimodal supervision. However, existing merging strategies rely largely on parameter-level relations or individual response strength, which provide only indirect evidence of cross-model transferability and may therefore introduce model- or modality-specific interference during reasoning transfer. We introduce BiVaR, a distribution-guided framework that identifies transferable reasoning components through their functional consistency across models. Specifically, transferability is determined by whether a neuron's activation behavior remains compatible with the other model's activation distribution, while bidirectional evaluation retains only reasoning components with mutually preserved functional behavior across models. Reasoning transfer is then restricted to the corresponding shared-neuron subspaces, after adaptive low-rank decomposition preserves dominant cross-model directions while suppressing residual variations. Extensive experiments across multiple VLM-LLM pairs show that BiVaR consistently improves mathematical and text-only reasoning while preserving or improving general multimodal capabilities.

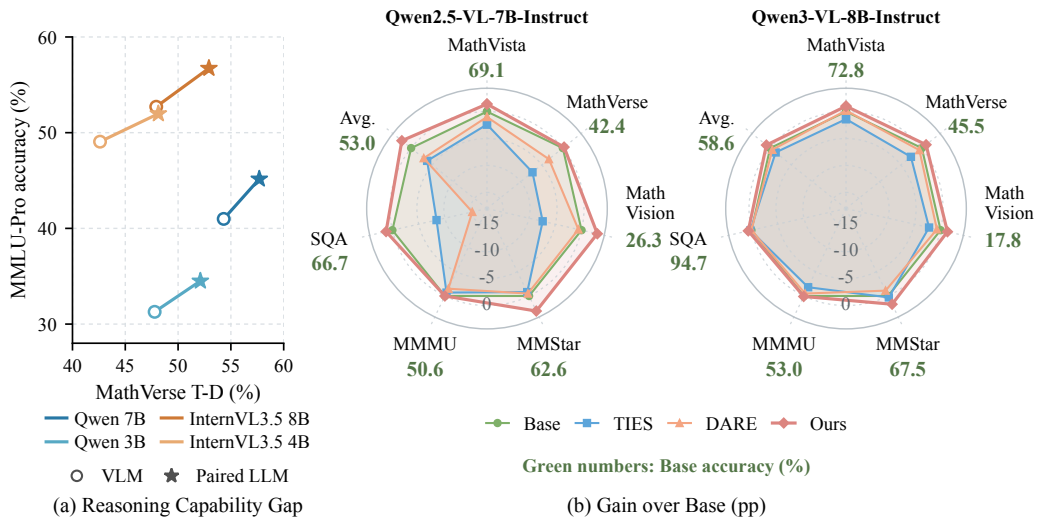
1 Introduction

Reasoning is a fundamental capability for vision-language models (VLMs), supporting multi-step inference, mathematical problem solving, and structured decision making (Qiao et al., 2025; Choudhury et al., 2025). Modern VLMs extend pretrained large language models (LLMs) with visual encoders and cross-modal projection modules (Lou et al., 2025; Jin et al., 2025), yet still lag behind their LLM backbones on long-horizon and multi-step reasoning (Rudman et al., 2025; Zhou et al., 2025), as illustrated in Fig. 1(a). Existing approaches improve reasoning through multimodal supervision, reinforcement learning (Yang et al., 2025; Liu et al., 2025), or inference-time scaling, often at substantial data and computational cost. This raises a question: *can the mature reasoning capabilities of strong LLMs be transferred directly to VLMs without expensive multimodal supervision?*

Model merging enables training-free reasoning transfer from LLMs to VLMs. While selective merging improves over direct parameter interpolation (Yadav et al., 2023) by identifying transferable components using parameter-level or importance-based criteria (Chen et al., 2025; Cui et al., 2026), such methods yield limited or inconsistent gains across reasoning benchmarks, as illustrated in Fig. 1(b). A key limitation is that these criteria remain indirect proxies for cross-model transferability: they characterize parameter relations or component importance, but do not directly determine whether these components exhibit compatible reasoning behavior across models. This limitation is further supported by our analysis in Table 6, which shows that high-response neurons are not necessarily effective carriers of reasoning transfer.

Therefore, cross-model transferability should be determined by **behavioral consistency rather than individual component responses**. We argue that activation distributions provide a more direct functional signature of such compatibility. Neurons involved in shared reasoning functions should remain distributionally consistent across models, even when their exact parameters or activation magnitudes differ. This motivates the use of distributional consistency as a direct criterion for identifying transferable reasoning components across models.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Reasoning Capability Gap (b) Gain over Base (pp)

Figure 1: Motivation of BiVaR. (a) The clear reasoning gap between paired text-only LLMs and VLMs suggests that LLMs provide a strong source of transferable reasoning capability. (b) Compared with the original VLM (Base), existing cross-model merging methods yield limited or inconsistent gains and may harm general or multimodal performance, whereas our BiVaR achieves more effective and stable improvements.

Building on this insight, we introduce BiVaR, a training-free framework that **identifies cross-model functional compatibility before performing parameter transfer**. Specifically, BiVaR reframes shared reasoning neuron identification as a cross-model distribution consistency problem, where transferability is determined by whether neuronal behavior remains compatible with the activation distribution of the other model. Bidirectional consistency is then enforced to isolate reasoning components whose functional behavior is mutually preserved across models. Finally, reasoning transfer is confined to the corresponding shared-neuron subspaces, after adaptive low-rank decomposition preserves dominant cross-model directions while suppressing residual variations.

We evaluate BiVaR across four VLM–LLM pairs on benchmarks spanning mathematical reasoning, general multimodal understanding, and text-only reasoning. Across 28 model–benchmark combinations, BiVaR achieves the best or tied-best performance in 25 cases. Compared with the corresponding base VLMs, BiVaR improves all 12 mathematical reasoning results and all four text-only reasoning results, while preserving or improving performance across all 12 general multimodal evaluations. Extensive ablations further validate the effectiveness of bidirectional distribution discrimination and adaptive low-rank transfer. Our contributions are summarized as follows:

- We formulate cross-model activation distribution consistency as a functional criterion for identifying transferable reasoning components. We propose bidirectional variational discrimination to identify shared reasoning neurons that are mutually consistent across models.
- We develop a selective transfer strategy that merges only the identified shared-neuron subspaces with adaptive low-rank decomposition, enabling effective reasoning transfer while preserving general multimodal capabilities.
- Extensive experiments across four VLM–LLM pairs and diverse reasoning and multimodal benchmarks demonstrate that BiVaR consistently improves mathematical and text-only reasoning while preserving or improving general multimodal capabilities.

2 Related Work

Reasoning in VLMs. Existing work improves VLM reasoning across different stages of the development pipeline. During pre-training, reasoning-related representations are enhanced through large-scale multimodal corpora (Han et al., 2025), synthetic data (Deng et al., 2025; Shi et al., 2024), and

improved training objectives (Lin et al., 2024). During post-training, instruction tuning (Liu et al., 2023), preference optimization (Yan et al., 2025), reinforcement learning (Shen et al., 2025; Yu et al., 2024b), and self-evolving frameworks (Liu et al., 2024; He et al., 2026) further strengthen reasoning ability. Recent studies also explore chain-of-thought supervision (Shao et al., 2024) and inference-time reasoning strategies (Gao et al., 2025). Despite these advances, improving VLM reasoning still often requires substantial multimodal data, additional optimization, or increased inference cost. `BiVaR` instead enables training-free reasoning transfer from LLMs to VLMs by identifying shared reasoning components through cross-model distribution consistency.

Model merging. Model merging combines complementary capabilities from multiple models without access to the original training data or large-scale training. Early methods rely on direct parameter averaging (Wortsman et al., 2022b;a; Rame et al., 2022), while later approaches introduce task vectors (Ilharco et al., 2022; Ortiz-Jimenez et al., 2023; Yang et al., 2024b) and importance-aware or conflict-aware strategies (Matena & Raffel, 2022; Yadav et al., 2023) to improve merging stability. Recent LLM–VLM merging methods explore selective transfer across models. IP-Merging (Hu et al., 2025) selects and aligns reasoning-related parameters through subspace similarity, while SNRF (Cui et al., 2026) identifies shared reasoning units at the neuron level for selective low-rank fusion. However, their selection criteria do not directly establish whether the identified components exhibit compatible reasoning behavior across models. `BiVaR` instead uses cross-model activation distribution consistency to identify shared reasoning components and performs selective transfer.

3 METHOD

`BiVaR` transfers reasoning capability from a text-only LLM to a VLM while minimizing interference with the multimodal knowledge acquired during vision-language alignment. We establish a variational distribution discrimination framework to measure distribution consistency without explicit density estimation (§3.2). We then apply it bidirectionally between the LLM and VLM to identify shared reasoning neurons (§3.3), followed by adaptive low-rank transfer within the corresponding weight subspaces to inject reasoning capability while limiting interference (§3.4).

3.1 PRELIMINARIES AND MOTIVATION

Given a paired LLM $\mathcal{M}^L$ and VLM $\mathcal{M}^V$, model merging transfers reasoning capability from $\mathcal{M}^L$ to $\mathcal{M}^V$ through parameter fusion while preserving the original multimodal capability of $\mathcal{M}^V$. Since modern VLMs are typically built upon pretrained LLMs, their corresponding parameters share the same dimensionality. We denote the parameters of $\mathcal{M}^V$ and $\mathcal{M}^L$ as $\mathbf{W}^V$ and $\mathbf{W}^L$, respectively. Conventional methods **directly interpolate parameters**, *e.g.*, $\mathbf{W}^M = (1-\lambda)\mathbf{W}^V + \lambda\mathbf{W}^L$, which may transfer both reasoning capability and unwanted model- or modality-specific variations.

To reduce such interference, recent methods attempt to identify transferable components before merging. Selective transfer can therefore be performed **at the neuron level** for finer-grained control. For a neuron j in a Transformer module m (*e.g.*, attention or feed-forward layers) at layer l, we regard its associated weights as a candidate transfer unit, identified by the index tuple (l, m, j). The transferability of each unit can be assessed from its activation responses over reasoning inputs, which are aggregated into a neuronal feature $\bar{z}_{l,j}^{X,m}$, where $X \in \{L, V\}$ denotes the LLM or VLM. For example, SNRF (Cui et al., 2026) identifies reasoning-related neurons separately in the LLM and VLM based on their activation magnitudes on reasoning inputs. It then defines shared neurons by intersecting these two sets.

However, individual activation strength only captures whether a neuron is highly responsive to reasoning inputs, but does not characterize its overall functional behavior across models. To address this limitation, we characterize neuron activation behaviors from a **distributional perspective**. Specifically, neuronal features collected from a model are regarded as samples from its module-level activation distribution. Neurons whose activation features are **in-distribution** with respect to the other model are regarded as shared reasoning neurons. In contrast, **out-of-distribution** activation features suggest model- or modality-specific behaviors. This distributional formulation shifts shared-neuron identification from individual response matching to distribution-level behavioral consistency, enabling bidirectional identification of shared reasoning neurons. To enable expressive distribution

estimation while avoiding explicit density modeling, §3.2 and §3.3 introduce a lightweight variational discriminator derived from the Donsker–Varadhan formula.

3.2 VARIATIONAL DISTRIBUTIONAL DISCRIMINATION

In §3.1, we recast cross-model shared-neuron identification as an in-distribution/out-of-distribution detection problem: a candidate neuronal feature is considered transferable if it remains consistent with the reference activation distribution. A common approach is to use parametric density models, which model activation features with a Gaussian distribution and enable direct density estimation (Ren et al., 2021; Lee et al., 2018). However, Gaussian assumptions fail to capture complex distributional patterns, especially in high-dimensional spaces. To capture more complex distributions, energy-based models can be adopted (Liu et al., 2020). However, their implicit density formulation requires expensive sampling-based estimation (*e.g.*, negative sampling or MCMC), making them computationally impractical for large-scale neuronal feature distributions. The effectiveness of our variational formulation over these conventional distribution-based detectors is further validated in Table 4.

Considering the activation distributions of the VLM and LLM as Q and P, respectively, the key quantity is not the absolute density $q(z)$ or $p(z)$, but their ratio $q(z)/p(z)$. A ratio close to one indicates consistency between the two distributions, while a large deviation suggests model-specific behavior. To this end, we adopt the Donsker–Varadhan variational formula (Donsker & Varadhan, 1983) which avoids explicit density modeling by transforming density ratio estimation into a tractable optimization problem over a learnable function f:

$$\mathrm{KL}(Q\|P) = \sup_f \underbrace{\left\{\mathbb{E}_{z\sim Q}[f(z)] - \log \mathbb{E}_{z\sim P}[e^{f(z)}]\right\}}_{\mathcal{J}(f)}. \tag{1}$$

Here, f acts as a variational discriminator between the candidate distribution Q and the reference distribution P. In practice, we parameterize f with a lightweight MLP, with implementation details provided in §B.2. The first term encourages larger responses for features from Q, while the second term regularizes the responses on P, allowing $\mathcal{J}(f)$ to capture the distributional discrepancy between the two distributions. The optimal discriminator is obtained by maximizing the variational objective:

$$f^* \in \arg\max_f \mathcal{J}(f), \tag{2}$$

Distributional deviation scoring. According to the optimality condition of the above variational problem, any maximizer f^* satisfies (proof in §A.1).

$$f^*(z) = \log \frac{q(z)}{p(z)} + \underbrace{\log \mathbb{E}_{z'\sim P}\left[e^{f^*(z')}\right]}_{\text{constant independent of } z}. \tag{3}$$

Since the second term is a constant independent of z, $f^*(z)$ preserves the same relative ordering as the density ratio $q(z)/p(z)$. Therefore, $f^*(z)$ provides a distributional deviation score that measures how a candidate feature differs from the reference distribution.

3.3 BIDIRECTIONAL CROSS-MODEL DISTRIBUTION DISCRIMINATION

Activation representation. Building on the variational discriminator in §3.2, we identify shared reasoning neurons from six projection modules in the Transformer, including $\mathcal{M} = \{\mathrm{q_proj}, \mathrm{k_proj}, \mathrm{v_proj}, \mathrm{o_proj}, \mathrm{up_proj}, \mathrm{down_proj}\}$. We use GSM8K (Cobbe et al., 2021) and Geo3K (Lu et al., 2021) as reasoning inputs for the LLM and VLM, respectively, and collect token-level activations for each neuron. Since sequence lengths vary across samples and modalities, we zero-pad each activation vector to a common length $T_{\max}$ and apply a fixed random projection shared across both models and all samples:

$$\tilde{z}_{i,j}^{X,m} = R^\top z_{i,j}^{X,m}, \qquad R \in \mathbb{R}^{T_{\max}\times d}, \quad X \in \{L, V\}. \tag{4}$$

We then aggregate the projected responses as $\bar{z}_j^{X,m} = \frac{1}{N_X}\sum_{i=1}^{N_X} \tilde{z}_{i,j}^{X,m}$, and independently optimize a discrimination function $f^{m,*}$ for each module $m \in \mathcal{M}$.

Bidirectional distribution discrimination. A unidirectional discrimination only verifies whether candidate neurons are consistent with the reference distribution, but does not guarantee mutual consistency between the two models. We perform distribution discrimination in two directions:

- **VLM-to-LLM direction.** We treat the LLM as the reference distribution and the VLM as the candidate distribution. Following §3.2, we optimize the module-specific discriminator $f^{m,*}_{V\to L}$ and identify VLM neurons consistent with the LLM distribution:

$$\mathcal{S}^m_{V\to L} = \left\{ (l,j) \;\middle|\; f^{m,*}_{V\to L}\left(\bar{z}^{V,m}_{l,j}\right) \le \tau^m_L \right\}. \tag{5}$$

- **LLM-to-VLM direction.** We reverse the roles of the two models, treating the VLM as the reference distribution and the LLM as the candidate distribution. The corresponding discriminator $f^{m,*}_{L\to V}$ identifies LLM neurons consistent with the VLM distribution:

$$\mathcal{S}^m_{L\to V} = \left\{ (l,j) \;\middle|\; f^{m,*}_{L\to V}\left(\bar{z}^{L,m}_{l,j}\right) \le \tau^m_V \right\}. \tag{6}$$

Adaptive thresholding and shared-neuron selection. For each direction, the decision threshold is determined from the score distribution of the corresponding reference model. Given an allowed misclassification ratio α, we define

$$\tau^m_P = \mathrm{Quantile}_{1-\alpha}\left(\left\{ f^{m,*}\left(\bar{z}^{P,m}_j\right) \right\}_{j=1}^{M_m} \right), \tag{7}$$

where $P \in \{L, V\}$ denotes the reference model and M_m is the number of neurons in module m. Finally, after aligning the corresponding layers, modules, and neuron indices, we retain neurons that are consistent in both directions:

$$\mathcal{S}^m_{\text{shared}} = \mathcal{S}^m_{V\to L} \cap \mathcal{S}^m_{L\to V}. \tag{8}$$

3.4 Adaptive Low-Rank Reasoning Transfer

We first define the cross-model difference as $\Delta\mathbf{W} = \mathbf{W}^L - \mathbf{W}^V$. We perform singular value decomposition on $\Delta\mathbf{W}$, $\Delta\mathbf{W} = U\Sigma V^\top$, where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_R)$ and $\sigma_1 \ge \cdots \ge \sigma_R$.

To retain the dominant singular directions while suppressing low-energy residual variations, we adaptively determine the transfer rank according to the spectral energy of $\Delta\mathbf{W}$. Since the Frobenius norm is unitarily invariant, its total spectral energy satisfies $\|\Delta\mathbf{W}\|_F^2 = \|\Sigma\|_F^2 = \sum_{i=1}^R \sigma_i^2$. We therefore select the minimum rank r^* that preserves at least a fraction τ_s of the total spectral energy:

$$r^* = \min\left\{ r : \frac{\sum_{i=1}^r \sigma_i^2}{\sum_{i=1}^R \sigma_i^2} \ge \tau_s \right\}. \tag{9}$$

The retained low-rank difference is reconstructed as $\Delta\mathbf{W}_{r^*} = \sum_{i=1}^{r^*} \sigma_i u_i v_i^\top$. Let $M_{\mathcal{S}_{\text{shared}}}(\cdot)$[1] denote a masking operator that retains only the weight entries associated with the shared neurons. Finally, we inject the retained reasoning update only into the identified shared-neuron subspace:

$$\mathbf{W}_{\text{merge}} = \mathbf{W}^V + \beta M_{\mathcal{S}_{\text{shared}}}\left(\Delta\mathbf{W}_{r^*}\right), \tag{10}$$

where β controls the transfer strength. This adaptive strategy allows different modules to retain different numbers of dominant transfer directions according to their spectral structures, while leaving VLM parameters outside the identified shared-neuron subspace unchanged.

4 Experiment

4.1 Setup

Model selection. We transfer reasoning from Qwen2.5-Math-7B (Yang et al., 2024a) to Qwen2.5-VL-7B-Instruct (Bai et al., 2025b), and from Qwen2.5-3B-Instruct-Math-GRPO (Williammsq,

[1]The procedure is applied independently to each module; we omit m for simplicity.

2025) to InternVL2.5-4B (Chen et al., 2024). For both Qwen3-VL-8B-Instruct (Bai et al., 2025a) and InternVL3.5-8B (Wang et al., 2025), DeepSeek-R1-0528-Qwen3-8B (Guo et al., 2025) serves as the reasoning source. Details are provided in §B.

Evaluation. We evaluate the merged models on multimodal mathematical reasoning, general multimodal understanding, and text-only reasoning. For mathematical reasoning, we use MathVista (Lu et al., 2024) with CoT, Solution, and Format evaluation, MathVerse (Zhang et al., 2024) with the Vision Dominant (V-D), Vision Intensive (V-I), and Vision Only (V-O) subsets, and MathVision (Wang et al., 2024). For general multimodal evaluation, we employ MMMU (Yue et al., 2024), MME (Fu et al., 2025), and ScienceQA (Saikh et al., 2022). We further evaluate text-only reasoning on the Chemistry, Biology, and Physics subsets of GPQA (Rein et al., 2023).

Baseline methods. We compare our method with four representative model merging approaches, including BR2V (Chen et al., 2025), DARE (Yu et al., 2024a), SNRF (Cui et al., 2026), and LED-Merging (Ma et al., 2025), covering general parameter merging, cross-modal reasoning transfer, neuron-level selective fusion, and adaptive merging strategies.

4.2 Main Results

Table 1 reports the main results across different VLM–LLM pairs, covering three evaluation categories: mathematical reasoning, general multimodal understanding, and text-only reasoning. Overall, our method delivers consistent improvements across different model families and scales.

Mathematical reasoning. Our method consistently improves mathematical reasoning across all four VLMs and achieves the best results on most benchmarks compared with existing merging baselines. On Qwen2.5-VL-7B-Instruct, MathVista and MathVision improve by 1.46 and 2.96 points, respectively. Consistent gains are also observed on Qwen3-VL-8B-Instruct, with MathVision increasing from 17.76 to 19.13, and on InternVL3.5-8B, where MathVista improves from 72.46 to 73.40. These results demonstrate effective reasoning transfer across different model families and scales. Detailed category-wise results are provided in §C.1.

General multimodal understanding. Our method maintains or improves performance on MME, MMMU, and ScienceQA in most settings, while also achieving the best results compared with existing merging baselines. For example, on Qwen2.5-VL-7B-Instruct, MME increases by 41.06 points and ScienceQA by 1.19 points, while MMMU remains unchanged at 50.56. On Qwen3-VL-8B-Instruct, all three general benchmarks are improved over the base model, with similar trends observed on the InternVL models. These results show that our transfer strategy preserves the original multimodal capability while enhancing reasoning.

Text-only reasoning. We further evaluate the merged models on GPQA, a text-only reasoning benchmark covering multiple scientific domains. Our method improves GPQA by 1.14 points on Qwen2.5-VL, 0.89 points on Qwen3-VL, 2.26 points on InternVL2.5, and 1.79 points on InternVL3.5. These consistent gains show that the transferred reasoning capability generalizes beyond multimodal mathematical tasks to cross-domain text-based reasoning.

4.3 Ablation Studies

Ablation 1: effect of bidirectional distribution discrimination. To validate the effectiveness of bidirectional distribution discrimination, we ablate the two discrimination directions on Qwen2.5-VL-7B-Instruct. V→L and L→V correspond to identifying neurons that are distributionally consistent with the reference model in each direction, while the bidirectional setting retains only neurons satisfying the consistency criterion in both directions.

As shown in Table 2, either direction alone improves several benchmarks, but the gains are inconsistent; for example, L→V decreases MathVerse from 42.43 to 41.71. This suggests that one-way consistency does not guarantee shared behavior across models. In contrast, bidirectional discrimination achieves the best performance across all benchmarks, showing that mutual distributional consistency provides a more reliable criterion for identifying shared reasoning neurons.

Ablation 2: functional role of identified shared reasoning neurons. We examine the functional importance of the identified shared neurons by zeroing them in Qwen2.5-VL-7B-Instruct and Qwen2.5-Math-7B, with random neuron zeroing as a control. As shown in Table 3, zeroing

Table 1: **Main results of cross-model reasoning transfer across different VLM–LLM pairs.** We report performance on mathematical reasoning, general multimodal understanding, and text-only reasoning benchmarks. The best results within each model group are highlighted in bold.

Method	Math Benchmarks			General Benchmarks			LLM Benchmarks
	MathVista	MathVerse	MathVision	MME	MMMU	ScienceQA	GPQA
Qwen2.5-VL-7B-Instruct							
Base	69.07	42.43	26.32	1680.69	**50.56**	66.70	29.68
BR2V	68.11	39.07	25.73	1669.73	49.11	51.54	21.67
DARE	70.10	42.51	27.94	1687.64	50.17	66.70	29.34
LED-Merging	69.48	**43.98**	27.71	1671.82	45.67	64.50	30.18
SNRF	69.23	42.43	27.12	1709.11	**50.56**	65.38	28.73
BiVaR	**70.53**	42.67	**29.28**	**1721.75**	**50.56**	**67.89**	**30.82**
Qwen3-VL-8B-Instruct							
Base	72.77	45.45	17.76	1726.94	53.00	94.65	31.69
BR2V	73.01	44.93	17.11	1701.83	52.51	94.65	25.76
DARE	72.92	45.93	17.76	1731.64	**53.79**	94.95	29.83
LED-Merging	73.14	45.96	18.13	1739.82	53.00	94.85	30.98
SNRF	72.87	46.37	17.98	1732.94	53.07	94.73	29.86
BiVaR	**73.77**	**46.47**	**19.13**	**1750.08**	53.11	**95.33**	**32.58**
InternVL2.5-4B							
Base	63.13	30.29	25.33	1671.39	48.44	97.37	28.12
BR2V	62.57	30.73	25.43	1670.21	49.83	96.97	23.69
DARE	62.94	30.65	25.33	1673.03	49.77	97.37	27.17
LED-Merging	63.97	30.63	25.74	1680.32	49.68	97.73	29.57
SNRF	62.83	30.51	25.98	1672.11	50.11	97.42	29.75
BiVaR	**65.11**	**31.97**	**26.74**	**1687.45**	**50.98**	**98.14**	**30.38**
InternVL3.5-8B							
Base	72.46	34.26	25.00	1705.93	56.33	97.17	34.60
BR2V	72.23	34.26	24.38	1707.47	56.94	97.11	32.93
DARE	72.93	34.85	25.00	1707.20	56.42	97.28	34.76
LED-Merging	72.98	35.07	25.34	1711.94	56.94	97.28	34.77
SNRF	72.76	34.93	25.10	1710.19	**57.33**	97.46	35.08
BiVaR	**73.40**	**35.18**	**25.66**	**1715.94**	57.03	**97.63**	**36.39**

Table 2: **Ablation study of bidirectional shared-neuron identification.** A checkmark indicates that the corresponding V→L or L→V discrimination direction is enabled. The bidirectional setting enables both directions and retains neurons consistently identified in both.

V→L	L→V	Math Benchmarks			General Benchmarks		LLM Benchmarks
		MathVista	MathVerse	MathVision	MME	ScienceQA	GPQA
Qwen2.5-VL-7B-Instruct							
-	-	69.07	42.43	26.32	1680.69	66.70	29.68
✓		69.30	42.50	28.29	1705.92	67.20	29.91
	✓	69.07	41.71	27.63	1708.06	66.79	29.76
✓	✓	**70.53**	**42.67**	**29.28**	**1721.75**	**67.89**	**30.82**

the identified neurons causes substantially larger degradation across all reasoning benchmarks. This suggests that our criterion preferentially identifies neurons that are functionally important for reasoning, although the intervention alone does not establish reasoning specificity.

Ablation 3: hyperparameter sensitivity. We analyze the sensitivity to the attention-module threshold parameter (α in Eq. (7)) on Qwen3-VL-8B-Instruct. As shown in Fig. 2, varying α from 0.10 to 0.35 yields stable performance on MME, GPQA, and MathVision, with all settings outperform-

Table 3: **Ablation study of the functional role of identified shared reasoning neurons.** "Shared" denotes zeroing the identified shared neurons, while "Random" denotes zeroing the same number of randomly selected neurons. "Base" represents the original model without neuron intervention.

-	MathVista	MathVerse	MathVision
Qwen2.5-VL-7B-Instruct			
Base	69.07	42.43	26.32
Shared	0	0	0
Random	27.81	16.51	12.16

-	Biology	Chemistry	Physics
Qwen2.5-Math-7B			
Base	44.87	29.51	34.23
Shared	0	0	0
Random	21.68	17.87	19.09

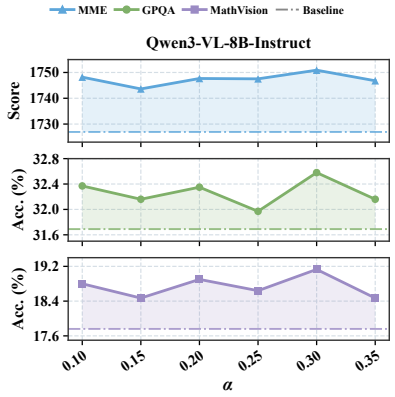


Figure 2: Sensitivity to the attention-module threshold parameter on MME, GPQA, and MathVision.

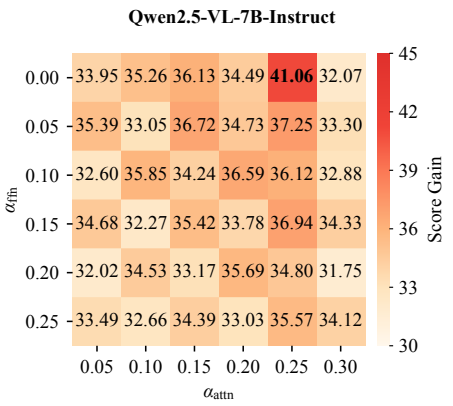


Figure 3: **Joint sensitivity to the attention- and FFN-module threshold parameters on MME.** Values denote performance gains over the base model.

ing the base model. Although the best setting varies across benchmarks, the overall performance is relatively insensitive to the exact value. Additional analyses are provided in §C.2 and §C.5.

We further examine the joint sensitivity to the attention- and FFN-module threshold parameters by varying them simultaneously. As shown in Fig. 3, performance remains stable across a broad range of parameter combinations, without a sharp optimum concentrated at a particular configuration. This suggests that BiVaR is also insensitive to the specific coupling between different module-wise threshold parameters, further demonstrating its robustness to hyperparameter choices.

Ablation 4: effect of ID/OOD detection methods. We replace our variational discriminator with conventional distribution-based detectors, including Mahalanobis Distance (MD), Relative Mahalanobis Distance (RMD), and Diagonal Gaussian (Diag. Gaussian). As shown in Table 4, all alternatives degrade reasoning and general multimodal performance, whereas our method consistently improves over the base model, supporting its effectiveness in identifying transferable reasoning neurons.

Table 4: Comparison of different ID/OOD detection methods on Qwen2.5-VL-7B-Instruct.

Method	MathVision	MME	GPQA
Qwen2.5-VL-7B-Instruct			
Base	26.32	1680.69	29.68
MD	15.79	1671.44	23.81
RMD	17.43	1672.07	24.52
Diag. Gaussian	13.82	1641.13	20.63
BiVaR	**29.28**	**1721.75**	**30.82**

4.4 Analyses

Layer-wise distribution of identified shared neurons. We further examine the layer-wise distribution of identified shared neurons using Qwen2.5-VL-7B-Instruct as an example. As shown in Fig. 4, the selected neurons are non-uniformly distributed and mainly concentrated in shallow and deeper layers. This pattern is broadly consistent with the hierarchical specialization of Transformer layers,

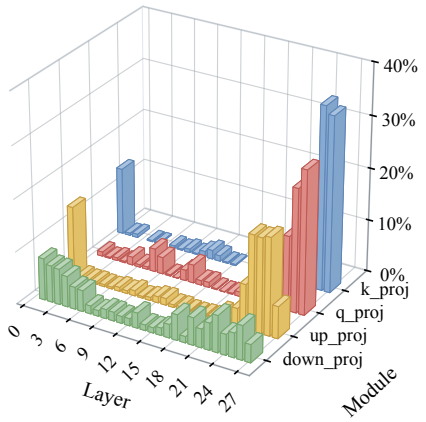


Figure 4: **Layer-wise distribution of identified shared neurons in Qwen2.5-VL-7B-Instruct.** The selected neurons exhibit a non-uniform distribution across Transformer layers and modules, with relatively higher concentrations in shallow and deeper layers.

Table 5: **Effect of SVD rank on Qwen2.5-VL-7B-Instruct and Qwen3-VL-8B-Instruct.** Performance stabilizes after retaining the dominant singular directions, suggesting that a small transfer rank is sufficient.

Rank	MathVision	MME	GPQA
Qwen2.5-VL-7B-Instruct			
2	25.99	1713.85	30.82
4	28.95	1721.01	30.82
8	29.28	1721.12	30.58
16	24.01	1717.80	29.91
32	28.95	1721.56	31.03
Qwen3-VL-8B-Instruct			
2	17.10	1727.01	31.25
4	19.13	1750.08	32.37
8	19.41	1747.63	32.37
16	17.76	1739.58	31.92
32	18.75	1746.18	32.14

where shallow layers capture lower-level representations while deeper layers support higher-level semantic processing and reasoning. The structured distribution suggests that our method preferentially identifies functionally meaningful components rather than selecting neurons arbitrarily. Results for other model pairs are provided in §C.3.

Effect of SVD rank. We analyze the effect of the retained SVD rank on Qwen2.5-VL-7B-Instruct and Qwen3-VL-8B-Instruct by varying it from 2 to 32. As shown in Table 5, performance improves after retaining the dominant singular directions but generally stabilizes at relatively small ranks, while further increasing the rank provides limited or inconsistent gains. This suggests that the transferable information is mainly concentrated in a low-dimensional subspace, supporting our adaptive low-rank transfer strategy.

Functional comparison of neuron selection criteria. As shown in Table 6, the shared subset improves MathVision and GPQA, indicating that neurons identified by both criteria contain useful transferable information. In contrast, S-only yields limited gains on MathVision and degrades GPQA, while B-only consistently improves both benchmarks. This shows that high activation alone does not reliably identify effective carriers of reasoning transfer, whereas our distributional criterion identifies neurons more suitable for cross-model reasoning transfer.

Table 6: **Functional comparison of neuron selection criteria.** We apply the same transfer update only to neurons commonly identified by `BiVaR` and SNRF (*Shared*), or uniquely identified by either method (*S-only* and *B-only*).

	Base	Shared	S-only	B-only
Qwen2.5-VL-7B-Instruct				
MathVision	26.32	26.97	26.64	27.87
GPQA	29.68	29.92	26.37	30.13

5 Conclusion

In this work, we revisit cross-model reasoning transfer from the perspective of functional compatibility. Rather than relying on structural correspondence or individual neuron responses, we argue that transferable reasoning components should exhibit consistent activation behavior across models. Extensive experiments across multiple VLM–LLM pairs show that `BiVaR` consistently improves mathematical and text-only reasoning performance while preserving general multimodal capabilities. Further analyses support the effectiveness of functional consistency for identifying reasoning components that enable reliable cross-model transfer.

AI USE STATEMENT

In this work, we used generative AI tools, primarily OpenAI's GPT-5.6 Sol and GPT-5.5, to assist with language-related tasks, including translation, linguistic and stylistic polishing, and formatting corrections. Every AI-assisted passage was manually reviewed in two separate passes to ensure its accuracy, clarity, and consistency with the authors' intended meaning.

We also used these tools for supporting, non-core programming tasks, including data processing, writing boilerplate and repetitive scripts, debugging, fixing bugs, and organizing or refactoring code. Generative AI was not used to develop the research ideas, design the proposed methodology, or implement its core components. All AI-assisted code and scripts were manually reviewed and tested in two separate passes.

The authors take full responsibility for the accuracy, originality, and integrity of the final manuscript and all associated research artifacts.

ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. In this study, no human subjects or animal experimentation was involved. All datasets used, including MME, MMStar, ScienceQA, MMMU, MathVision, MathVista and MathVerse, were sourced in compliance with relevant usage guidelines, ensuring no violation of privacy. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

REPRODUCIBILITY STATEMENT

We have made every effort to ensure that the results presented in this paper are reproducible. The proposed method is described in detail in the main text, including the construction of neuronal features, bidirectional variational discrimination, adaptive thresholding, and low-rank parameter transfer. Additional implementation details, hyperparameter settings, and model-specific configurations are provided in the Appendix. All backbone models and evaluation benchmarks used in our experiments are publicly available, including Qwen2.5-VL, Qwen3-VL, InternVL2.5, InternVL3.5, MathVista, MathVerse, MathVision, MME, MMMU, ScienceQA, and GPQA. These materials are intended to enable independent reproduction of our experimental results and facilitate further research on cross-model reasoning transfer.

REFERENCES

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*, 2025a.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025b.

Shiqi Chen, Jinghan Zhang, Tongyao Zhu, Wei Liu, Siyang Gao, Miao Xiong, Manling Li, and Junxian He. Bring reason to vision: Understanding perception and reasoning through model merging. *arXiv preprint arXiv:2505.05464*, 2025.

Zhe Chen, Weiyun Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Erfei Cui, Jinguo Zhu, Shenglong Ye, Hao Tian, Zhaoyang Liu, et al. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*, 2024.

Monjoy Narayan Choudhury, Junling Wang, Yifan Hou, and Mrinmaya Sachan. Can vision-language models solve visual math equations? In *EMNLP*, 2025.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Chenhang Cui, An Zhang, Yuxin Chen, Gelei Deng, Jingnan Zheng, Zhenkai Liang, Xiang Wang, and Tat-Seng Chua. Do llms and vlms share reasoning neurons? evidence and mechanisms of cross-modal transfer. In *CVPR Findings*, 2026.

Linger Deng, Linghao Zhu, Yuliang Liu, Yu Wang, Qunyi Xie, Jingjing Wu, Gang Zhang, Yingying Zhu, and Xiang Bai. Theorem-validated reverse chain-of-thought problem generation for geometric reasoning. In *EMNLP*, 2025.

Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain markov process expectations for large time. iv. *Communications on pure and applied mathematics*, 1983.

Chaoyou Fu, Peixian Chen, Yunhang Shen, Yulei Qin, Mengdan Zhang, Xu Lin, Jinrui Yang, Xiawu Zheng, Ke Li, Xing Sun, et al. Mme: A comprehensive evaluation benchmark for multimodal large language models. In *NeurIPS*, 2025.

Jun Gao, Yongqi Li, Ziqiang Cao, and Wenjie Li. Interleaved-modal chain-of-thought. In *CVPR*, 2025.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Xiaotian Han, Yiren Jian, Xuefeng Hu, Haogeng Liu, Yiqi Wang, Qihang Fan, Yuang Ai, Huaibo Huang, Ran He, Zhenheng Yang, et al. Infimm-webmath-40b: Advancing multimodal pretraining for enhanced mathematical reasoning. In *EMNLP Findings*, 2025.

Yicheng He, Chengsong Huang, Zongxia Li, Jiaxin Huang, and Yonghui Yang. Visplay: Self-evolving vision-language models. In *CVPR*, 2026.

Yijie Hu, Zihao Zhou, Kaizhu Huang, Xiaowei Huang, and Qiufeng Wang. Can mllms absorb math reasoning abilities from llms as free lunch? In *NeurIPS*, 2025.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022.

Hyundong Jin, Hyung Jin Chang, and Eunwoo Kim. Instruction-grounded visual projectors for continual learning of generative vision-language models. In *ICCV*, 2025.

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *NeurIPS*, 2018.

Zhenghao Lin, Zhibin Gou, Yeyun Gong, Xiao Liu, Yelong Shen, Ruochen Xu, Chen Lin, Yujiu Yang, Jian Jiao, Nan Duan, et al. Not all tokens are what you need for pretraining. In *NeurIPS*, 2024.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In *NeurIPS*, 2023.

Wei Liu, Junlong Li, Xiwen Zhang, Fan Zhou, Yu Cheng, and Junxian He. Diving into self-evolving training for multimodal reasoning. *arXiv preprint arXiv:2412.17451*, 2024.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *NeurIPS*, 2020.

Ziyu Liu, Zeyi Sun, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, and Jiaqi Wang. Visual-rft: Visual reinforcement fine-tuning. In *ICCV*, 2025.

Haoran Lou, Chunxiao Fan, Ziyan Liu, Yuexin Wu, and Xinliang Wang. Llava-sp: Enhancing visual representation with visual spatial tokens for mllms. In *ICCV*, 2025.

Pan Lu, Ran Gong, Shibiao Jiang, Liang Qiu, Siyuan Huang, Xiaodan Liang, and Song-Chun Zhu. Inter-gps: Interpretable geometry problem solving with formal language and symbolic reasoning. In *ACL-IJCNLP*, pp. 6774–6786, 2021.

Pan Lu, Hritik Bansal, Tony Xia, Jiacheng Liu, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, Kai-Wei Chang, Michel Galley, and Jianfeng Gao. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. In *ICLR*, 2024.

Qianli Ma, Dongrui Liu, Qian Chen, Linfeng Zhang, and Jing Shao. Led-merging: Mitigating safety-utility conflicts in model merging with location-election-disjoint. In *ACL*, 2025.

Michael S Matena and Colin Raffel. Merging models with fisher-weighted averaging. In *NeurIPS*, 2022.

Guillermo Ortiz-Jimenez, Alessandro Favero, and Pascal Frossard. Task arithmetic in the tangent space: Improved editing of pre-trained models. In *NeurIPS*, 2023.

Runqi Qiao, Qiuna Tan, Guanting Dong, Minhui Wu, Chong Sun, Xiaoshuai Song, Jiapeng Wang, Zhuoma Gongque, Shanglin Lei, Yifan Zhang, et al. We-math: Does your large multimodal model achieve human-like mathematical reasoning? In *ACL*, 2025.

Alexandre Rame, Matthieu Kirchmeyer, Thibaud Rahier, Alain Rakotomamonjy, Patrick Gallinari, and Matthieu Cord. Diverse weight averaging for out-of-distribution generalization. In *NeurIPS*, 2022.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. Gpqa: A graduate-level google-proof q&a benchmark. *arXiv preprint arXiv:2311.12022*, 2023.

Jie Ren, Stanislav Fort, Jeremiah Liu, Abhijit Guha Roy, Shreyas Padhy, and Balaji Lakshminarayanan. A simple fix to mahalanobis distance for improving near-ood detection. *arXiv preprint arXiv:2106.09022*, 2021.

William Rudman, Michal Golovanevsky, Amir Bar, Vedant Palit, Yann LeCun, Carsten Eickhoff, and Ritambhara Singh. Forgotten polygons: Multimodal large language models are shape-blind. In *ACL Findings*, 2025.

Tanik Saikh, Tirthankar Ghosal, Amish Mittal, Asif Ekbal, and Pushpak Bhattacharyya. Scienceqa: a novel resource for question answering on scholarly articles: T. saikh et al. *IJDL*, 2022.

Hao Shao, Shengju Qian, Han Xiao, Guanglu Song, Zhuofan Zong, Letian Wang, Yu Liu, and Hongsheng Li. Visual cot: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning. In *NeurIPS*, 2024.

Haozhan Shen, Peng Liu, Jingcheng Li, Chunxin Fang, Yibo Ma, Jiajia Liao, Qiaoli Shen, Zilun Zhang, Kangjia Zhao, Qianqian Zhang, et al. Vlm-r1: A stable and generalizable r1-style large vision-language model. *arXiv preprint arXiv:2504.07615*, 2025.

Wenhao Shi, Zhiqiang Hu, Yi Bin, Junhua Liu, Yang Yang, See Kiong Ng, Lidong Bing, and Roy Ka-Wei Lee. Math-llava: Bootstrapping mathematical reasoning for multimodal large language models. In *EMNLP Findings*, 2024.

Ke Wang, Junting Pan, Weikang Shi, Zimu Lu, Houxing Ren, Aojun Zhou, Mingjie Zhan, and Hongsheng Li. Measuring multimodal mathematical reasoning with math-vision dataset. In *NeurIPS*, 2024.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. Internvl3. 5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265*, 2025.

Williammsq. Qwen2.5-3b-instruct-math-grpo. https://huggingface.co/Williammsq/Qwen2.5-3B-Instruct-Math-GRPO, 2025. Accessed: 2025-09-08.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *ICML*, 2022a.

Mitchell Wortsman, Gabriel Ilharco, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo Lopes, Hannaneh Hajishirzi, Ali Farhadi, Hongseok Namkoong, et al. Robust fine-tuning of zero-shot models. In *CVPR*, 2022b.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. In *NeurIPS*, 2023.

Ziang Yan, Zhilin Li, Yinan He, Chenting Wang, Kunchang Li, Xinhao Li, Xiangyu Zeng, Zilei Wang, Yali Wang, Yu Qiao, et al. Task preference optimization: Improving multimodal large language models with vision task alignment. In *CVPR*, 2025.

An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jianhong Tu, Jingren Zhou, Junyang Lin, et al. Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*, 2024a.

Enneng Yang, Zhenyi Wang, Li Shen, Shiwei Liu, Guibing Guo, Xingwei Wang, and Dacheng Tao. Adamerging: Adaptive model merging for multi-task learning. In *ICLR*, 2024b.

Yi Yang, Xiaoxuan He, Hongkun Pan, Xiyan Jiang, Yan Deng, Xingtao Yang, Haoyu Lu, Dacheng Yin, Fengyun Rao, Minfeng Zhu, et al. R1-onevision: Advancing generalized multimodal reasoning through cross-modal formalization. In *ICCV*, 2025.

Le Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In *ICML*, 2024a.

Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, et al. Rlhf-v: Towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In *CVPR*, 2024b.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, et al. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. In *CVPR*, 2024.

Renrui Zhang, Dongzhi Jiang, Yichi Zhang, Haokun Lin, Ziyu Guo, Pengshuo Qiu, Aojun Zhou, Pan Lu, Kai-Wei Chang, Yu Qiao, et al. Mathverse: Does your multi-modal llm truly see the diagrams in visual math problems? In *ECCV*, 2024.

Sheng Zhou, Junbin Xiao, Qingyun Li, Yicong Li, Xun Yang, Dan Guo, Meng Wang, Tat-Seng Chua, and Angela Yao. Egotextvqa: Towards egocentric scene-text aware video question answering. In *CVPR*, 2025.

A Additional Implementation Details

A.1 Relation Between the Optimal Discrimination Function and the Log-Density Ratio

Proposition 1. *Let P and Q be two distributions defined over the same feature space, with probability densities $p(z)$ and $q(z)$, respectively, and assume that Q is absolutely continuous with respect to P. Consider the variational objective*

$$\mathcal{J}(f) = \mathbb{E}_{z\sim Q}[f(z)] - \log \mathbb{E}_{z\sim P}\left[e^{f(z)}\right]. \tag{11}$$

For an optimal function

$$f^* \in \arg\max_f \mathcal{J}(f), \tag{12}$$

we have

$$f^*(z) = \log \frac{q(z)}{p(z)} + C, \tag{13}$$

where

$$C = \log \mathbb{E}_{z'\sim P}\left[e^{f^*(z')}\right] \tag{14}$$

is independent of z. Therefore, $f^*(z)$ is equivalent to the log-density ratio up to an additive constant.

Proof. We first rewrite the objective in integral form:

$$\mathcal{J}(f) = \int q(z) f(z)\, dz - \log \int p(z) e^{f(z)}\, dz. \tag{15}$$

The first term encourages large responses on features drawn from the candidate distribution Q, whereas the second term normalizes these responses with respect to the reference distribution P. To characterize the function that optimally distinguishes the two distributions, we examine the first-order variation of $\mathcal{J}$.

Consider an arbitrary perturbation of f,

$$f_\epsilon(z) = f(z) + \epsilon\eta(z), \tag{16}$$

where $\eta(z)$ is an arbitrary test function and ϵ is a scalar. The directional derivative of the objective at $\epsilon = 0$ is

$$\frac{d}{d\epsilon}\mathcal{J}(f+\epsilon\eta)\Big|_{\epsilon=0} = \int \left[q(z) - \frac{p(z)e^{f(z)}}{\int p(z')e^{f(z')}\, dz'}\right]\eta(z)\, dz. \tag{17}$$

At the optimum $f = f^*$, the first-order variation must vanish for every admissible perturbation $\eta(z)$. Since $\eta(z)$ is arbitrary, the quantity multiplying $\eta(z)$ must vanish almost everywhere. Therefore,

$$q(z) = \frac{p(z)e^{f^*(z)}}{\int p(z')e^{f^*(z')}\, dz'}. \tag{18}$$

This condition shows that the optimal function reweights the reference density $p(z)$ such that the resulting normalized density matches the candidate density $q(z)$.

Define the normalization term

$$Z = \int p(z')e^{f^*(z')}\, dz' = \mathbb{E}_{z'\sim P}\left[e^{f^*(z')}\right]. \tag{19}$$

Then Eq. (18) can be rewritten as

$$e^{f^*(z)} = \frac{q(z)}{p(z)} Z. \tag{20}$$

Taking the logarithm on both sides gives

$$f^*(z) = \log \frac{q(z)}{p(z)} + \log Z. \tag{21}$$

Substituting the definition of Z, we obtain

$$f^*(z) = \log \frac{q(z)}{p(z)} + \log \mathbb{E}_{z' \sim P} \left[e^{f^*(z')} \right]. \tag{22}$$

The second term in Eq. (22) is independent of the particular feature z. Therefore, the optimal discrimination function preserves the same relative ordering over features as the log-density ratio $\log \frac{q(z)}{p(z)}$. Features that are more characteristic of the candidate distribution Q relative to the reference distribution P receive larger scores, whereas features that are better explained by P receive relatively smaller scores.

Finally, $\mathbb{E}_{z \sim Q}[f(z)]$ is linear in f, while

$$f \mapsto \log \mathbb{E}_{z \sim P} \left[e^{f(z)} \right] \tag{23}$$

is convex. Hence, $\mathcal{J}(f)$ is concave in f, and any stationary point is a global maximizer. This establishes the relation between the optimal discrimination function and the log-density ratio, and justifies using $f^*(z)$ as the distributional deviation score in §3.2. □

A.2 Interpretation of the Adaptive Threshold

For a reference distribution P, the threshold τ_P^m is defined as the $(1-\alpha)$-quantile of the reference score distribution:

$$\tau_P^m = \text{Quantile}_{1-\alpha} \left(\left\{ f^{m,*} \left(\tilde{z}_j^{P,m} \right) \right\}_{j=1}^{M_m} \right). \tag{24}$$

By definition of the $(1-\alpha)$-quantile, approximately an α fraction of reference features have scores larger than τ_P^m:

$$\Pr_{z \sim P} \left[f^{m,*}(z) > \tau_P^m \right] \approx \alpha. \tag{25}$$

Thus, α directly controls the tolerated fraction of reference features that are classified as distributionally deviating, providing a module-adaptive decision boundary without requiring an absolute score threshold shared across modules.

B Implementation Details

B.1 Activation Extraction and Projection

For each VLM–LLM pair, we collect neuronal activations from six Transformer modules: $\mathcal{M} = \{\text{q_proj}, \text{k_proj}, \text{v_proj}, \text{o_proj}, \text{up_proj}, \text{down_proj}\}$. We use GSM8K as the reasoning input for the LLM and Geo3K for the VLM, and randomly sample $N = 100$ examples from each side for activation extraction.

Since the token lengths of different reasoning samples vary, we normalize the activation vectors to a common length of $T_{\max} = 256$ before random projection. Specifically, activation vectors shorter than 256 tokens are zero-padded to length 256, while those exceeding 256 tokens are truncated to their first 256 positions. In practice, the vast majority of the sampled reasoning sequences do not exceed this limit, and truncation is therefore rarely required. We adopt a fixed maximum length rather than padding all samples to the longest observed sequence, since a small number of unusually long inputs would otherwise introduce substantial redundant padding and unnecessarily enlarge the subsequent projection space.

After length normalization, all activation vectors are projected from 256 dimensions to a common feature space of dimension $d = 64$ using a fixed random projection matrix. Importantly, the same projection matrix is shared across all samples and both models within each VLM–LLM pair, ensuring that their neuronal responses are represented in a consistent feature space. Finally, the projected responses are averaged over the 100 reasoning samples to obtain the neuronal features used for subsequent bidirectional distribution discrimination.

Table 7: Implementation settings of the variational discriminator.

Setting	Value
Input dimension	64
Optimizer	AdamW
Learning rate	5×10^{-4}
Weight decay	1×10^{-4}
Optimization steps	300
Optimization precision	FP32
Optimization mode	Full batch

B.2 Variational Discriminator Training

We independently instantiate a variational discriminator for each module m. The two discrimination directions are optimized separately, since reversing the reference and candidate distributions defines a different variational objective and therefore requires an independently optimized discrimination function.

Before discriminator training, we normalize the neuronal features using the feature-wise mean and standard deviation computed from the reference distribution. The same reference statistics are then applied to the candidate features, ensuring that both distributions are represented under a common normalization without introducing candidate-dependent statistics. The common optimization settings used for all model pairs are summarized in Table 7.

The discrimination function is implemented as a lightweight multilayer perceptron consisting of two hidden layers with a hidden dimension of 64. Each hidden layer is followed by a SiLU activation, and the final linear layer maps the hidden representation to a scalar discrimination score. We optimize each discriminator for 300 steps using AdamW with a learning rate of 5×10^{-4} and a weight decay of 10^{-4}. The optimization is performed in full batch over all neuronal features of the corresponding module across Transformer layers.

For each direction, the decision threshold is calibrated using the score distribution of the corresponding reference model as described in §3.3. The threshold-related hyperparameter α is selected separately for different model pairs, and the specific values are reported in §B.4. Model forward passes and activation extraction are performed in FP16, while the discriminator is optimized in FP32 for numerical stability.

B.3 Shared-Neuron Subspace Construction and Merging

After bidirectional discrimination, we construct the transferable parameter subspace using the neurons retained in both directions. The neuron indices are matched according to their Transformer layer and projection module. For $\{\mathrm{q_proj}, \mathrm{k_proj}, \mathrm{v_proj}, \mathrm{up_proj}\}$, each identified neuron corresponds to an output channel of the associated linear projection and therefore indexes a row of its weight matrix. For $\{\mathrm{o_proj}, \mathrm{down_proj}\}$, neurons are defined on the input channels and correspondingly index columns of the weight matrix. This row/column convention is kept consistent between activation profiling and parameter transfer.

For each layer and projection module, we first compute the parameter difference between the paired LLM and VLM and perform SVD on the full difference matrix. Rather than using a fixed rank across different layers and modules, we retain the minimum number of singular directions whose cumulative spectral energy reaches a predefined ratio τ_s. Consequently, the effective transfer rank can vary according to the spectral structure of each parameter difference.

The reconstructed low-rank difference is masked before being applied to the VLM, ensuring that parameters outside the identified shared-neuron subspace remain unchanged. The magnitude of the transferred update is further controlled by the coefficient β. Thus, τ_s determines how much of the dominant spectral information is retained, while β controls the overall strength with which the retained difference is injected into the VLM.

Table 8: **Model-specific hyperparameter settings used in the main experiments.** α denotes the reference rejection ratio for distribution discrimination, τ_s denotes the retained spectral-energy ratio, and β controls the reasoning-transfer strength.

VLM–LLM Pair	α_{attn}	α_{ffn}	τ_s	β
Qwen2.5-VL-7B-Instruct / Qwen2.5-Math-7B	0.25	0.00	0.90	0.50
Qwen3-VL-8B-Instruct / DeepSeek-R1-0528-Qwen3-8B	0.30	0.00	0.85	0.20
InternVL2.5-4B / Qwen2.5-3B-Instruct-Math-GRPO	0.15	0.05	0.90	0.80
InternVL3.5-8B / DeepSeek-R1-0528-Qwen3-8B	0.10	0.00	0.95	0.50

The values of τ_s and β used for each model pair are summarized in §B.4. We further analyze their sensitivity in §C.

B.4 Model-Specific Settings

The implementation details shared across all model pairs have been described in §B.1, §B.2, and §B.3, including the number of reasoning samples, activation length normalization, random projection dimension, discriminator architecture, and optimization settings. Here, we summarize only the hyperparameters that are adjusted for individual VLM–LLM pairs.

Specifically, α controls the reference rejection ratio used for distribution-consistency discrimination, while τ_s determines the spectral-energy ratio retained during adaptive low-rank decomposition and β controls the strength of the transferred parameter update. These hyperparameters are selected independently for each model pair, and the settings used in our main experiments are summarized in Table 8.

C Additional Experimental Results

C.1 Detailed Breakdown of Mathematical Reasoning Results

In the main paper, we report the aggregated performance on MathVista, MathVerse, and MathVision for a compact comparison across different model pairs and merging baselines. Here, we provide the detailed breakdown over the corresponding subcategories to further examine where the improvements of `BiVaR` arise.

As shown in Table 9, `BiVaR` achieves consistent improvements across multiple fine-grained reasoning categories rather than relying on gains from a single subset. On MathVista, the improvements are observed across CoT, Format, and Solution evaluation settings, while on MathVerse, `BiVaR` maintains competitive or improved performance across V-D, V-I, and V-O. Similar gains are also observed on MathVision. These detailed results are consistent with the aggregated results reported in the main paper and indicate that the effectiveness of `BiVaR` extends across diverse forms of visual mathematical reasoning.

C.2 Additional Hyperparameter Sensitivity Results

In the main paper, we report the sensitivity to the threshold parameter for attention-module on Qwen3-VL-8B-Instruct. Here, we provide additional results on Qwen2.5-VL-7B-Instruct, InternVL2.5-4B, and InternVL3.5-8B to further examine whether the observed robustness generalizes across different VLM architectures and model scales.

As shown in Fig. 5, varying α from 0.10 to 0.35 leads to only moderate performance fluctuations across MME, GPQA, and MathVision. For all three models, most configurations remain comparable to or outperform their corresponding base models, while no single value consistently dominates across all benchmarks. These results are consistent with the observation on Qwen3-VL-8B-Instruct in the main paper, indicating that `BiVaR` is not sensitive to a narrowly tuned attention-module rejection ratio and remains stable over a reasonably broad range of settings.

Table 9: Main results of cross-model reasoning transfer across different VLM–LLM pairs. We report performance on mathematical reasoning benchmarks. The best results within each model group are highlighted in bold.

Method	MathVista				MathVerse				MathVision
	CoT	Format	Solution	Overall	V-D	V-I	V-O	Overall	
Qwen2.5-VL-7B-Instruct									
Base	69.3	69.2	68.7	69.07	41.88	43.02	42.39	42.43	26.32
SNRF	69.4	69.2	69.1	69.23	42.01	43.02	42.26	42.43	27.12
BiVaR	**71.1**	**69.5**	**71.0**	**70.53**	**42.26**	**43.15**	**42.64**	**42.67**	**29.28**
Qwen3-VL-8B-Instruct									
Base	76.0	67.2	75.1	72.77	46.83	45.69	43.83	45.45	17.76
SNRF	76.1	67.2	75.3	72.87	47.84	**46.19**	**45.05**	46.37	17.98
BiVaR	**76.5**	**68.4**	**76.4**	**73.77**	**48.73**	45.63	**45.05**	**46.47**	**19.13**
InternVL-2.5-4B									
Base	61.3	65.7	62.4	63.13	31.98	32.61	26.27	30.29	25.33
SNRF	60.9	65.2	62.4	62.83	32.02	32.67	26.84	30.51	25.98
BiVaR	**63.2**	**68.0**	**64.1**	**65.11**	**32.29**	**33.74**	**29.88**	**31.97**	**26.74**
InternVL-3.5-8B									
Base	73.1	70.8	73.5	72.46	38.71	37.56	26.52	34.26	25.00
SNRF	73.3	71.2	73.8	72.76	39.04	**37.99**	27.76	34.93	25.10
BiVaR	**74.0**	**71.7**	**74.5**	**73.40**	**39.72**	37.94	**27.88**	**35.18**	**25.66**

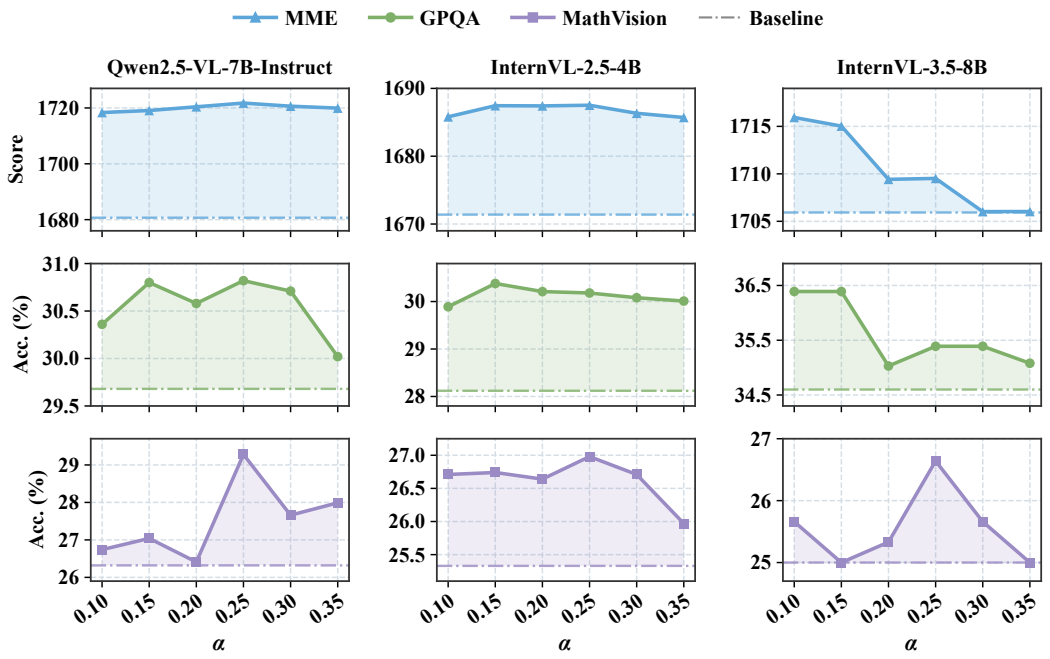


Figure 5: **Sensitivity analysis of the attention-module threshold parameter.** Performance on MME, GPQA, and MathVision; dashed lines indicate the corresponding base-model scores.

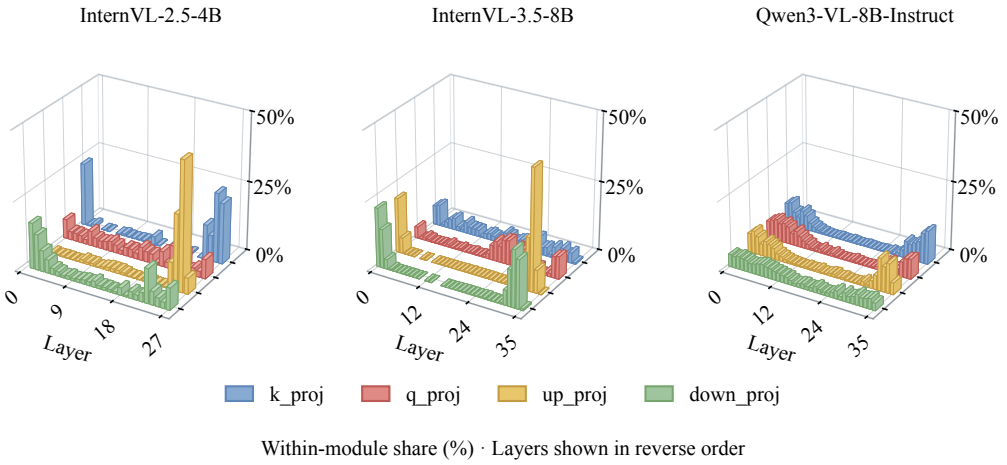


Figure 6: **Layer-wise distribution of identified shared neurons on additional model pairs.** The bars show the within-module proportion of selected neurons across Transformer layers.

C.3 Additional Layer-wise Distribution Results

In the main paper, we analyze the layer-wise distribution of identified shared neurons using Qwen2.5-VL-7B-Instruct as a representative example. Here, we provide additional results for Qwen3-VL-8B-Instruct, InternVL2.5-4B, and InternVL3.5-8B.

As shown in Fig. 6, the identified shared neurons exhibit clear non-uniform layer-wise distributions across all three models. Despite differences in model architecture and scale, the selected neurons tend to appear more frequently in particular stages of the Transformer rather than being evenly distributed across layers. In particular, a noticeable concentration can be observed in relatively shallow and deeper layers, broadly consistent with the pattern reported for Qwen2.5-VL-7B-Instruct in the main paper. This cross-model consistency further suggests that the proposed identification criterion captures structurally organized and functionally meaningful reasoning-related components rather than arbitrary neuron subsets.

C.4 Case Study

Fig. 7 provides a qualitative comparison of the reasoning behaviors produced by different transfer strategies. All three models correctly perceive the key visual elements, including the $55°$, $75°$, and $x°$ angles, indicating that the main difficulty does not lie in visual perception but in reasoning over their geometric relations. The original VLM introduces unnecessary intermediate variables and incorrect angle relations, eventually producing an inconsistent reasoning chain and a wrong answer. SNRF exhibits a similar failure: although the relevant angles are correctly recognized, it incorrectly constructs a pentagon-based relation and applies an inappropriate angle-sum constraint, suggesting that its transferred neurons do not substantially strengthen the underlying reasoning process.

This difference is closely related to how transferable neurons are identified. SNRF mainly selects neurons according to their response strength to reasoning inputs and matches them through positional correspondence. Such highly activated neurons can be correlated with reasoning inputs without necessarily carrying compatible reasoning functions across the LLM and VLM. Consequently, transferring these components does not guarantee improved reasoning behavior. In contrast, `BiVaR` selects neurons according to their cross-model distributional consistency, favoring components that exhibit compatible functional responses in both models. As illustrated in the example, the resulting model directly identifies the relevant triangle relation and correctly derives $55 + 75 + x = 180$, yielding $x = 50°$. Together with the neuron-ablation results, this qualitative example further supports that the neurons identified by `BiVaR` are more closely associated with the reasoning computation required for solving the task.

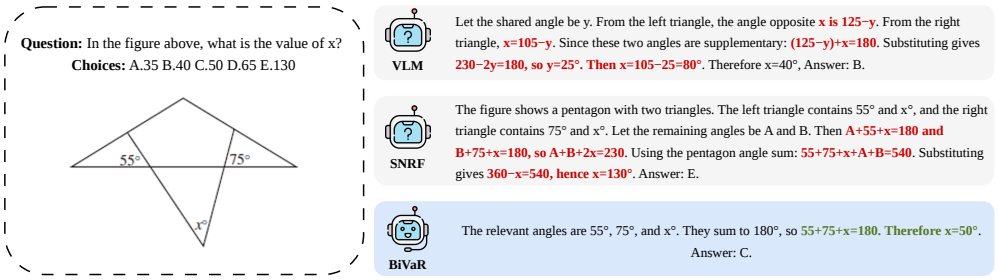


Figure 7: Qualitative comparison of reasoning transfer. All models correctly perceive the key angles, while the original VLM and SNRF fail to establish the correct geometric relations. In contrast, `BiVaR` produces a concise and correct reasoning chain, indicating more effective transfer of reasoning-related functionality.

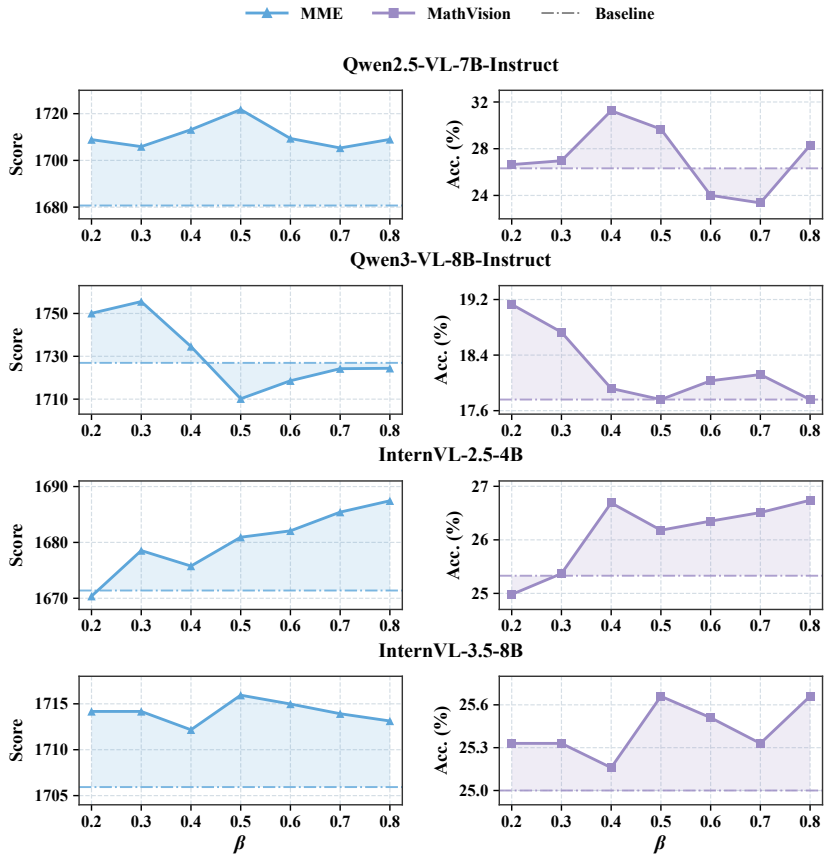


Figure 8: Sensitivity to the transfer strength β across different VLM–LLM pairs on MathVision and MME.

C.5 Sensitivity to Transfer Strength

We further study the effect of the transfer strength β, which controls the magnitude of the reasoning update injected into the VLM. Figure 8 reports the sensitivity results across all VLM–LLM pairs. We take Qwen2.5-VL-7B-Instruct as a representative example for analysis. For this model, MME remains consistently above the base model across all tested values, indicating that the transferred

update does not substantially disrupt general multimodal capability. In contrast, MathVision is more sensitive to the transfer strength: moderate values of β yield clear improvements, while overly strong transfer can degrade reasoning performance due to increased cross-model interference. These results suggest that an appropriate transfer strength is important for balancing reasoning enhancement and preservation of the original VLM capability. Accordingly, we use $\beta = 0.5$ for this model pair.