AIMED ELSEWHERE: UNDERSTANDING AND RECOVERING SPATIAL LOCALIZATION IN MOE VLMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Supervised fine-tuning can adapt general-purpose vision-language models to embodied tasks by improving spatial reasoning and adherence to task requirements. In the sparse mixture-of-experts backbones we study, however, these gains coexist with substantial degradation in spatial localization, while general visual-language performance remains largely preserved. Complementary interventions reveal that spatial competence remains recoverable. Rephrasing the request recovers much of the localization loss without changing the weights. Subtracting an update along the fine-tuning trajectory partially recovers performance without additional training data, while adding the same update to a healthy checkpoint of an independent run reproduces the decline. Together, these findings support a **re-aiming, not erasure** account: adaptation alters how spatial competence is expressed. These interventions establish recoverability but do not ensure recovery across task requirements. Successful localization requires both accurate spatial targets and responses in the structure required by each task. Our controlled comparisons show that supervision covering only a subset of these requirements leaves recovery incomplete. We therefore introduce Joint Spatial Re-Aiming (JSR), which guides corrective updates through joint supervision of the affected tasks and their required response structures. With 384 distinct examples and 32 updates, JSR substantially restores spatial localization while largely preserving the gains of embodied adaptation.

1 INTRODUCTION

Embodied tasks require vision-language models (VLMs) to understand a scene and provide spatial information that supports action. Understanding which object an instruction refers to and how objects relate must be complemented by accurate spatial localization, such as specifying the image coordinates of an object to grasp or a location for placement (Yuan et al., 2025). Supervised fine-tuning (SFT) adapts general-purpose VLMs to these requirements, but its benefits need not extend uniformly across capabilities. In our experiments, spatial reasoning and task adherence improve while general visual-language performance remains largely preserved. Yet spatial localization degrades substantially under the same adaptation process. Figure 1 illustrates this contrast through changes in spatial reasoning, localization accuracy, and response type. These observations raise a central question: how can embodied adaptation improve task performance while degrading the spatial localization required for action?

We investigate this phenomenon in sparse mixture-of-experts (MoE) backbones, which combine substantial model capacity with selective parameter activation, making them attractive for embodied adaptation (Yang et al., 2025; Bai et al., 2025). Spatial localization degrades in both MoE backbones we study, whereas dense variants trained on the same mixture do not exhibit the corresponding decline. Our focus is on understanding this degradation and restoring spatial localization while preserving the benefits of embodied adaptation.

The observed failures suggest that degraded localization need not imply the loss of spatial competence. When asked for a bounding box, an adapted model may instead return a point inside the requested object: spatial information remains available, but the response does not satisfy the task. Making the response structure explicit recovers much of the multi-object localization loss without changing the weights. Weight-space interventions provide complementary evidence. Using task arithmetic (Ilharco et al., 2023), we subtract an update along the fine-tuning trajectory and partially recover localization

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.

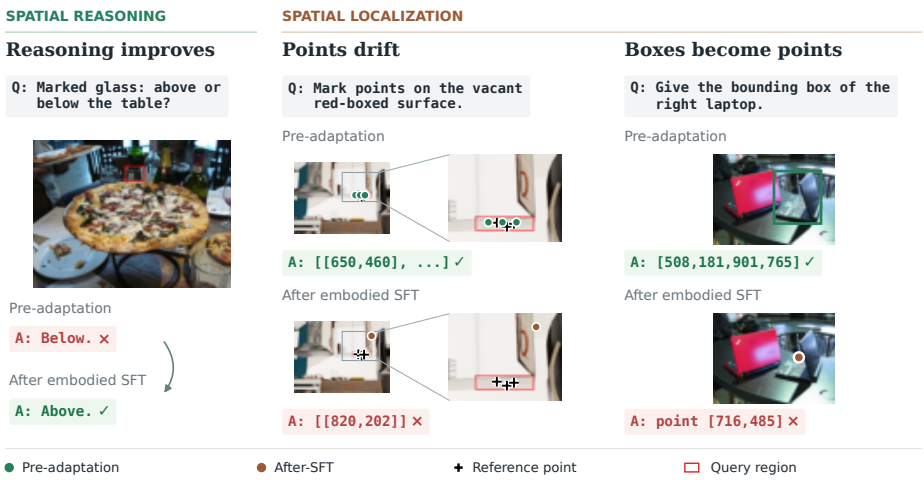


Figure 1: **Embodied adaptation improves reasoning while disrupting localization.** Selected Qwen3.5 outputs show a corrected spatial relation (left), a displaced point (middle), and a point replacing the requested box (right). The first two examples use the main evaluation; the third uses an independent response-type probe.

without additional training data. Adding the same update to a healthy checkpoint of an independent run reproduces the decline, while norm-matched random updates fail to produce comparable recovery. Together, these findings support a **re-aiming, not erasure** account: adaptation alters how spatial competence is expressed, leaving substantial capacity accessible to targeted interventions.

Recoverability alone does not ensure that all task requirements are restored together. Spatial localization requires accurate targets expressed in the response structure specified by each task. Our controlled comparisons show that supervision restricted to a subset of tasks or response structures leaves other requirements unmet. Interventions on routing, adapter configuration, and response structure provide incomplete recovery, while standard Group Relative Policy Optimization (GRPO) (Shao et al., 2024) also falls short of restoring localization to its pre-adaptation level in our experiments. These findings motivate Joint Spatial Re-Aiming (JSR), which constructs a joint supervision distribution covering the affected tasks and their required response structures. By keeping correct targets and their required expression jointly represented throughout repair, JSR guides corrective updates toward recovery across these requirements. With 384 distinct examples and 32 updates, it substantially restores spatial localization while largely preserving the gains of embodied adaptation.

Our contributions are threefold:

1. We identify selective degradation of spatial localization during embodied adaptation in two MoE backbones, alongside improvements in task adherence and spatial reasoning.
2. We provide complementary response-level and weight-space evidence for *re-aiming, not erasure*, showing that spatial performance remains recoverable and that the same trajectory update can induce degradation across runs.
3. We introduce JSR, which guides corrective updates through joint coverage of affected tasks and response structures, and demonstrate substantial localization recovery with limited supervision while largely preserving adaptation gains.

2 Related Work

Embodied adaptation and spatial grounding. General-purpose VLMs have been adapted to embodied reasoning and planning (Driess et al., 2023), as well as direct action generation in vision-language-action models (Brohan et al., 2023; Kim et al., 2024). Spatial grounding provides a complementary interface between visual understanding and action: RoboPoint predicts spatial affordances for robotic tasks (Yuan et al., 2025), while Molmo incorporates pointing into a general-purpose VLM (Deitke et al., 2024). SpatialVLM develops spatial reasoning through targeted supervision (Chen et al., 2024a), and benchmarks such as EmbSpatial and CVBench assess related capabilities (Du et al., 2024; Tong et al., 2024). Building on these efforts, we examine whether embodied

adaptation preserves existing spatial localization capabilities and how to recover them when they degrade.

Recoverable forgetting. Recent work distinguishes degraded task performance from the loss of underlying competence. Studies of spurious forgetting and format specialization highlight the role of task alignment and response conventions (Zheng et al., 2025; Wang et al., 2024; Chen et al., 2025). In continual image classification, Trivedi & Melwani (2026) show that task information can remain accessible to probes and classifier retraining, while Chaudhury (2026) recovers performance by aligning and stitching computations across checkpoints. We examine recoverability in spatial localization through complementary interventions on model responses and weights. A concurrent study investigates repetition and termination errors induced by coordinate-list fine-tuning and mitigates them through decoding controls (Zhou et al., 2026). Our focus is on degradation of existing spatial capabilities during embodied adaptation and their recovery through corrective training.

Continual adaptation of multimodal and MoE models. Capability degradation during fine-tuning has been documented in language models (Luo et al., 2023) and multimodal models (Zhai et al., 2024; Li et al., 2026). Parameter-efficient adaptation can reduce forgetting relative to full fine-tuning (Biderman et al., 2024). For MoE models, continual-learning research examines routing and expert stability (Li et al., 2025; Zhao et al., 2026; Liu et al., 2026; Hou et al., 2026; Xie et al., 2026), alongside analyses of misrouting and attention components (Yoon et al., 2026; Chen et al., 2026). Our study characterizes spatial localization degradation under embodied adaptation and evaluates interventions at the response, parameter, and supervision levels, with routing interventions included among the comparisons.

Capability recovery and joint supervision. Approaches to capability retention include parameter regularization (Kirkpatrick et al., 2017) and distribution alignment (Huang et al., 2025), while work on instruction alignment demonstrates the effectiveness of limited, high-quality supervision (Zhou et al., 2023). Task arithmetic (Ilharco et al., 2023) and instruction vectors (Jiang et al., 2024) offer complementary tools for modifying model behavior through parameter differences. We use weight-space interventions to investigate recoverability, then translate the resulting evidence into a supervision design: JSR jointly covers affected tasks and their required response structures, keeping correct targets and their required expression represented throughout repair.

3 Problem formulation

3.1 Spatial tasks and response requirements

Given an image x and an instruction q, a VLM with parameters W generates a response y. For spatial localization, the response specifies a set of points or bounding boxes associated with the objects or action locations requested by q. Successful localization requires both geometrically accurate targets and adherence to the task's response requirements, including the output type and structure. Geometric accuracy alone does not ensure that the response fulfills the instruction, and structural compliance alone does not ensure accurate localization.

To distinguish how localization degrades, we examine four diagnostic dimensions: *output validity*, whether the response can be parsed under the evaluation protocol; *task selection*, whether it provides the requested type of spatial output; *cardinality*, whether the number of predicted targets matches the reference; and *spatial precision*, whether the predicted locations agree with the annotations. These dimensions distinguish failures to express the requested output from errors in its spatial content. Cardinality is assessed separately from task selection because multiple annotated points may refer to the same object. The corresponding metrics and scoring procedures are specified in Appendix A.5.

3.2 Recovery objective

Given an adapted model with degraded spatial localization, our objective is to recover accurate spatial predictions in the structures required by the affected tasks, while preserving the benefits of embodied adaptation. A successful repair must therefore recover the affected localization capabilities while retaining the broader benefits of adaptation.

4 UNDERSTANDING RECOVERABLE SPATIAL DEGRADATION

We investigate whether degraded localization reflects a loss of spatial competence or a change in how that competence is expressed. To distinguish these possibilities, we intervene on response instructions and model weights, testing whether localization can recover without additional task training. We then examine what these interventions imply for the supervision needed to restore the affected task requirements jointly.

4.1 RECOVERY THROUGH RESPONSE INTERVENTIONS

We first test whether localization can recover by changing the response instruction alone, holding model weights, images, and annotated targets fixed. On 5,000 Objects365 items, replacing a permissive instruction with an explicit response schema increases the adapted Qwen3.5-35B-A3B's box-set F_1 from 14.45 to 61.80. The base model remains nearly unchanged under the same two instructions, scoring 68.23 and 68.49, respectively. Consequently, the adapted model's deficit relative to base narrows from 53.78 to 6.69 points. This recovery without a weight update indicates that much of the observed degradation concerns how spatial predictions are elicited and expressed.

A complementary probe reveals spatial information in responses that fail to satisfy the requested output type. When asked to return bounding boxes without a prescribed task-specific template, the adapted Qwen3.5-35B-A3B instead returns points on 24.7% of the $6{,}929$ items. Of these points, 95.80% lie inside the corresponding ground-truth boxes (Appendix C.3). The responses therefore retain information about the target location while violating the requested output type, illustrating why task adherence and spatial accuracy must be examined separately.

4.2 RECOVERY THROUGH TASK ARITHMETIC

We next test whether parameter changes accumulated during adaptation can both reverse and induce localization degradation. Using task arithmetic (Ilharco et al., 2023), we denote the checkpoints immediately before and after an observed decline by W_a and W_b, and extract their parameter difference $\Delta W = W_b - W_a$. We apply this update in two directions:

$$W_{\text{recover}} = W_{\text{SFT}} - \Delta W, \qquad W_{\text{perturb}} = W_{\text{healthy}} + \Delta W.$$

Here, W_{SFT} is the adapted endpoint, and W_{healthy} is a checkpoint from an independent run of the same model before substantial degradation. Both interventions modify the stored parameters directly, without additional training.

One update acts in both directions. Subtracting ΔW from the adapted endpoint recovers 51.5% of the localization loss between W_a and that endpoint. Conversely, adding the same update to a healthy checkpoint of an independent run induces a decline equivalent to 92.4% of the loss originally observed between W_a and W_b (Figure 2). Thus, the same parameter update supports partial recovery when removed and induces degradation when transferred, without additional training in either case.

Direction-dependent recovery beyond the decline interval. Two controls clarify the role of the selected update (Appendix D.3). Random updates matched to ΔW in per-tensor norm leave referential localization performance close to the degraded endpoint, indicating that perturbation magnitude alone does not explain recovery. However, recovery is not unique to ΔW. Subtracting an update from an earlier stage of training yields greater recovery, despite having a similar norm and being nearly orthogonal to ΔW. These results show that recovery depends on the update direction but need not be confined to the interval in which degradation was observed.

4.3 FROM RECOVERABILITY TO CORRECTIVE UPDATES

Together, the response and weight interventions support a **re-aiming, not erasure** account: spatial competence remains accessible, but adaptation alters how it is expressed. This suggests a repair strategy that guides further parameter updates toward the desired task responses while retaining the benefits of adaptation. The remaining question is what supervision can support recovery across the affected tasks and their response requirements.

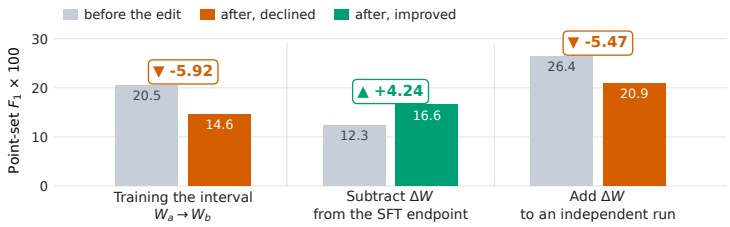


Figure 2: **The same update partially repairs localization when subtracted and induces degradation when transferred.** Scores use all 3,290 point-localization items; changes are computed before rounding.

Partial recovery motivates joint coverage. Replay and response-prefix interventions recover aspects of localization, but isolated routing, adapter, and response-structure interventions do not consistently restore all requirements (Appendices D.1, E and F). The recovery-strategy comparison in Section 6.3 evaluates parameter editing and GRPO alongside JSR. Partial recovery therefore motivates a corrective signal spanning the affected requirements.

Task coverage addresses asymmetric transfer. Recovery can transfer between tasks, but the transfer is asymmetric. Supervision for point prediction also improves multi-object box localization, whereas supervision for box prediction leaves point localization substantially below its pre-adaptation level, even when multiple response structures are covered (Section 6.4). This asymmetry motivates explicit coverage of the affected tasks rather than relying on transfer from a single repair task.

Partial recovery from separated supervision. Separating spatial targets from response structures yields partial recovery, but the benefits remain uneven. We compare supervision that preserves the requested structure while shuffling spatial targets with supervision that retains correct coordinates in a reduced response structure. The structure-preserving treatment improves box localization, indicating that response structure contributes to recovery even when spatial targets are shuffled. Applying the two supervision treatments sequentially can also recover part of the geometric performance, but does not restore the required class labels (Section 6.4 and Appendix B). These results suggest that individual supervision components can reactivate aspects of localization, motivating their joint specification to recover accurate targets and complete task responses together.

A principle for re-aiming. These findings motivate a principle for corrective adaptation: guide parameter updates with supervision that jointly represents the affected tasks and their response requirements. Weight interventions establish that localization remains recoverable through structured parameter changes, while the supervision controls show that partial recovery does not ensure recovery across requirements. We therefore construct a joint corrective signal in which accurate spatial targets and their required response structures remain represented throughout repair. The next section formalizes this principle as Joint Spatial Re-Aiming.

5 Joint Spatial Re-Aiming

Joint Spatial Re-Aiming (JSR) guides corrective adaptation through joint coverage of tasks and response structures. Given an adapted model with degraded spatial localization, JSR constructs a supervision distribution that represents each affected task and its required response structures.

5.1 Constructing joint coverage

Let $\mathcal{T}$ denote the tasks selected for repair, and let $\mathcal{S}_\tau$ denote the response structures required by task τ. For every $\tau \in \mathcal{T}$ and $s \in \mathcal{S}_\tau$, we collect examples (x, q, y) in which the target response y provides correct task content in the structure s specified by the instruction q. We denote the empirical distribution of these examples by $\mathcal{D}_{\tau,s}$. Joint coverage requires that every task–structure pair within the repair scope be represented in the supervision.

The resulting repair distribution is

$$\mathcal{D}_{\mathrm{JSR}} = \sum_{\tau \in \mathcal{T}} \sum_{s \in \mathcal{S}_\tau} \pi_{\tau,s}\, \mathcal{D}_{\tau,s}, \qquad \pi_{\tau,s} > 0, \qquad \sum_{\tau,s} \pi_{\tau,s} = 1. \tag{1}$$

Here, $\pi_{\tau,s}$ is the sampling probability assigned to task–structure pair (τ, s). Positive weights ensure that every required pair contributes to the repair objective, while their relative values determine the balance of supervision across tasks and structures. Joint coverage therefore specifies which requirements are represented; the mixture weights specify their relative emphasis.

5.2 Joint correction of the adaptation trajectory

Starting from the adapted parameters W_{SFT}, JSR learns repair parameters ϕ by minimizing the negative log-likelihood of complete target responses under the joint supervision distribution:

$$\mathcal{L}_{\mathrm{JSR}}(\phi) = -\mathbb{E}_{(x,q,y)\sim\mathcal{D}_{\mathrm{JSR}}} \sum_{t=1}^{|y|} \log p_{W_{\mathrm{SFT}},\phi}(y_t \mid x, q, y_{<t}), \tag{2}$$

Here, y contains both the correct task content and its required response structure, so the objective supervises these requirements together. The adapted parameters W_{SFT} remain fixed while ϕ is optimized.

With the mixture weights fixed, the expected gradient is a weighted sum of the gradients associated with the covered task–structure pairs. The supervision distribution therefore shapes the corrective signal throughout repair. Guided by the recoverability evidence in Section 4, JSR learns a new sequence of updates from the adapted model, with accurate task targets and their required response structures jointly represented in the objective.

We implement the repair parameters ϕ using a fresh LoRA adapter (Hu et al., 2022), trained on examples sampled from the joint supervision distribution. Coverage is defined over this distribution; individual minibatches need not contain every task–structure pair. After training, the adapter is merged into the adapted model to obtain the repaired parameters W_{rep}. The resulting model requires neither a repair-time prefix nor an external response-structure selection module at inference. Training configurations are provided in Appendix A.2.

6 Evaluation

Our evaluation addresses three questions: how embodied adaptation changes spatial localization and broader task performance; how effectively JSR restores localization while retaining adaptation gains; and how task coverage, response-structure coverage, and supervision budget affect recovery.

6.1 Experimental setting

Models and training. We study Qwen3.5-35B-A3B (Qwen Team, 2026) and Qwen3-VL-30B-A3B-Instruct (Bai et al., 2025; Qwen Team, 2025), two MoE backbones with different multimodal architectures and pretraining lineages. Both are adapted on the same embodied training mixture comprising 78 data sources, using LoRA (Hu et al., 2022) while keeping the vision encoder, multimodal projector, and router parameters frozen. Five Qwen3.5-35B-A3B runs use the same training data and differ in training seed and data order. Two dense Qwen3.6-27B variants trained on the same mixture provide an auxiliary comparison. Throughout, *base* denotes each model's checkpoint before embodied adaptation. We evaluate the Base/SFT/JSR capability profile on both MoE backbones, and conduct the detailed recovery-strategy comparisons and ablations on Qwen3.5-35B-A3B. Model-selection details are provided in Appendix A.1.

Evaluation benchmarks. Localization uses 3,290 point examples from RoboPoint, RoboRefIt, and FSD_points, 6,929 RefCOCO examples, and 5,000 Objects365 validation images. Table 1 also reports spatial reasoning, embodied understanding, and general visual-language benchmarks. Scores retain task-specific metrics on a 0–100 scale; unweighted group means summarize the profile alongside individual results. Dataset references, splits, response protocols, and scoring are provided in Appendix A.

Group	Benchmark	Qwen3.5-35B-A3B					Qwen3-VL-30B-A3B-Instruct				
		Base	SFT	JSR	Δ_{SFT}	Δ_{JSR}	Base	SFT	JSR	Δ_{SFT}	Δ_{JSR}
Spatial localization	RoboPoint, RoboRefIt, FSD_points	26.98	12.32	45.00	−14.66	+18.02	28.87	20.46	34.48	−8.41	+5.61
	Objects365	68.23	14.45	67.57	−53.78	−0.66	66.09	46.31	63.81	−19.77	−2.27
	Objects365 (with names)	66.19	0.00	59.17	−66.19	−7.02	64.23	0.00	61.10	−64.23	−3.14
	RefCOCO	60.59	93.75	92.39	+33.16	+31.80	81.15	92.34	92.02	+11.18	+10.87
	Average	55.50	30.13	66.03	−25.37	**+10.54**	60.08	39.78	62.85	−20.31	**+2.77**
Spatial reasoning & embodied understanding	CVBench	85.60	88.51	88.78	+2.91	+3.18	75.47	89.69	87.26	+14.22	+11.79
	BLINK	40.56	66.02	65.97	+25.46	+25.41	40.77	58.92	58.44	+18.15	+17.67
	EmbSpatial	81.04	82.97	81.65	+1.93	+0.61	77.03	82.53	81.54	+5.50	+4.51
	Cosmos-Reason1	59.41	59.61	60.00	+0.20	+0.59	56.86	57.25	56.47	+0.39	−0.39
	VSI-Bench	60.66	62.12	61.69	+1.46	+1.03	60.39	59.69	58.65	−0.70	−1.74
	ERQA	50.00	48.25	50.00	−1.75	0.00	38.00	41.50	40.50	+3.50	+2.50
	RoboVQA	62.77	59.54	61.88	−3.23	−0.89	60.92	63.70	62.60	+2.78	+1.68
	Average	62.86	66.72	67.14	+3.85	**+4.28**	58.49	64.75	63.64	+6.26	**+5.15**
General visual-language	MMBench-V1.1	89.01	89.16	87.85	+0.15	−1.16	87.85	85.68	85.14	−2.17	−2.71
	MMStar	73.53	69.47	70.13	−4.06	−3.40	69.33	68.87	68.80	−0.47	−0.53
	RealWorldQA	77.65	79.74	78.95	+2.09	+1.30	74.12	73.73	72.94	−0.39	−1.18
	MMMU (validation)	59.44	62.67	60.89	+3.23	+1.45	60.89	59.67	57.78	−1.22	−3.11
	DocVQA	94.34	94.55	93.51	+0.21	−0.83	94.01	93.06	92.90	−0.95	−1.11
	Average	78.79	79.12	78.27	+0.32	−0.53	77.24	76.20	75.51	−1.04	−1.73

Table 1: Adaptation and recovery on two MoE backbones. $\Delta_{\mathrm{SFT}} = \mathrm{SFT} - \mathrm{Base}$; $\Delta_{\mathrm{JSR}} = \mathrm{JSR} - \mathrm{Base}$. Scores use a 0–100 scale; group means include all listed benchmarks. Objects365 measures geometry alone or jointly with class names.

JSR implementation. Both backbones use 384 distinct examples and a fresh LoRA adapter trained for 32 updates: 192 point examples and 192 box examples spanning three response structures, with 64 per structure. The vision encoder, projector, and router remain frozen. Further settings are in Appendices A.2 and H.

6.2 Adaptation and recovery across the capability profile

Both backbones exhibit a contrast between adaptation gains and spatial localization degradation (Table 1). SFT increases the mean across spatial reasoning and embodied understanding benchmarks by 3.85 points on Qwen3.5-35B-A3B and 6.26 points on Qwen3-VL-30B-A3B-Instruct, while the localization means decrease by 25.37 and 20.31 points, respectively. After JSR, the localization means exceed their respective base levels by 10.54 and 2.77 points, and the spatial reasoning and embodied understanding means remain 4.28 and 5.15 points above base. These group summaries describe the aggregate profile; the individual metrics distinguish the degree of recovery within each group.

The detailed Qwen3.5-35B-A3B evaluation shows benefits in both response adherence and spatial reasoning performance. Answer extraction increases from 94.69% to 100% on CVBench and from 72.01% to 100% on BLINK, indicating more consistent production of scorable responses. On EmbSpatial, where extraction is already complete for both model states, the score improves from 81.04 to 82.97. This improvement shows that the gains extend beyond increased answer extractability.

Localization changes are selective across tasks in both models. For Qwen3.5-35B-A3B, multi-object box geometry decreases by 53.78 points, whereas single-object localization accuracy increases by 33.16 points. The latter gain primarily reflects more frequent production of valid boxes, with box IoU remaining stable on examples for which both model states produce parseable outputs (Appendix C.2). The decline in aggregate point localization is consistent across all five Qwen3.5-35B-A3B runs and the Qwen3-VL-30B-A3B-Instruct run, while the two dense variants trained on the same mixture do not exhibit the corresponding decline (Appendix C.1).

JSR substantially restores localization on both backbones. On Qwen3.5-35B-A3B, point localization reaches 45.00, exceeding base by 18.02 points, while multi-object box geometry reaches 67.57, within 0.66 points of base. The paired 95% confidence interval for the geometry improvement over the adapted endpoint is $[52.21, 53.98]$. Class-sensitive box localization improves from 0.00 to 59.17, against a base score of 66.19. On Qwen3-VL-30B-A3B-Instruct, point localization rises from 20.46 to 34.48, above its base of 28.87. Multi-object geometry improves from 46.31 to 63.81, approaching

its base of 66.09, and the class-sensitive score rises from 0.00 to 61.10, against a base of 64.23. Thus, both models achieve substantial joint localization recovery. Under the primary box protocol, class-sensitive performance recovers approximately 89% and 95% of the respective SFT-induced losses, bringing labeled-box prediction close to its pre-adaptation level.

Relative to SFT, JSR changes the spatial reasoning and embodied understanding mean by $+0.42$ on Qwen3.5 and -1.12 on Qwen3-VL; general visual-language means change by -0.85 and -0.69, respectively. Single-object localization retains most of its adaptation gain, ending 31.80 and 10.87 points above base. Thus, joint localization recovery coexists with modest changes in broader group means; individual benchmark differences remain visible in Table 1.

6.3 Comparison with alternative recovery strategies

We compare recovery strategies from the same adapted Qwen3.5-35B-A3B endpoint on the same localization panel (Figure 3). Subtracting a trajectory update raises the point-localization score from 12.32 to 16.55 without additional training data. With the same 384-example repair set and 32 updates, GRPO (Shao et al., 2024) reaches 15.71, compared with 45.00 for JSR. Increasing GRPO training to 128 updates raises its score to 20.3, still below the base score of 26.98. At 32 updates, its multi-object box geometry score remains close to the adapted endpoint, at 15.60 versus 14.45, and its class-sensitive score remains zero. These comparisons show greater recovery with JSR under the evaluated configurations; the following ablations examine the contribution of joint task and response-structure coverage. Full configurations and results are provided in Appendix G.

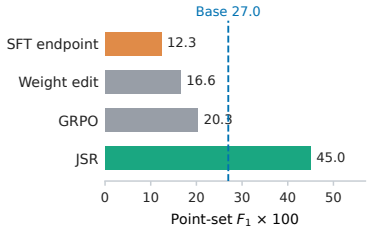


Figure 3: **Recovery strategies differ in supervision and scope.** The dashed line marks the base score.

Prefix interventions use preselected degraded items and replay changes the adaptation process itself; their separate protocols and results are reported in Appendices D.1 and E.3.

6.4 Why joint coverage matters

Task coverage. Joint task coverage supports recovery across localization tasks. At a fixed budget of 384 examples, box-only supervision spanning three response structures restores multi-object box geometry but leaves point localization at 16.26, below its base score of 26.98 (Figure 4a). Conversely, point-only supervision improves both point localization and box geometry, but provides little recovery under class-sensitive box scoring. JSR jointly recovers point localization, box geometry, and class-sensitive predictions, demonstrating the benefit of covering the affected tasks together. Complete configurations and scores are provided in Appendix B.1.

Response-structure coverage. Response-structure coverage is particularly important for recovering class-sensitive localization. The joint baseline uses the same task balance, total sample count, update budget, and pointing examples as JSR, but limits box supervision to a single response structure. Its class-sensitive score is 2.75, compared with 59.17 for JSR, while its box geometry score reaches 58.21 versus 67.57. A similar contrast appears under box-only supervision: covering three structures yields a class-sensitive score of 58.94, compared with 0.31 for one structure. These results show that recovering box geometry does not by itself ensure recovery of the associated class information, supporting explicit coverage of the required response structures.

Sequential supervision. Separating spatial targets from response structures produces partial but uneven recovery. Structure-preserving supervision with shuffled targets raises box geometry from 14.45 to 41.12. Applying the structure treatment after coordinate supervision further improves geometry to 49.43, whereas the reverse order yields 14.77. Neither sequence restores class-sensitive predictions, motivating joint supervision throughout repair (Appendix B.3).

Task balance. Varying the point-to-box allocation trades point against box recovery. Increasing the box share slightly improves box scores, whereas increasing the point share reduces class-sensitive

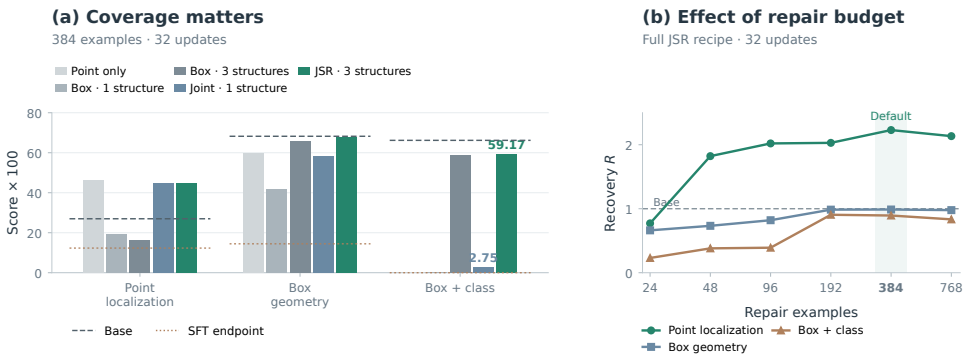


Figure 4: **Coverage and supervision budget.** (a) Task and structure coverage at 384 examples. (b) JSR recovery versus distinct repair examples; $R = 1$ marks base and shading marks the default budget. All repair configurations use 32 updates.

recovery without improving point localization. Balanced 192/192 coverage therefore provides an effective joint default (Appendix B.3).

Effect of supervision budget. We vary the number of distinct repair examples while keeping the task balance, response-structure coverage, and optimization settings fixed. Each repair set is repeated to the same number of training rows, and all configurations use 32 updates. To compare recovery across metrics, we report $R = (s_{\text{rep}} - s_{\text{SFT}})/(s_{\text{base}} - s_{\text{SFT}})$, where s denotes the score for the corresponding metric. Thus, $R = 1$ indicates recovery to the pre-adaptation level, and $R > 1$ indicates performance above that level (Figure 4b).

Recovery responds differently to supervision budget across metrics. Point localization exceeds its pre-adaptation level with 48 examples and reaches its highest observed score at 384 examples. Box geometry approaches base performance at 192 examples ($R = 0.99$), while class-sensitive recovery remains below $R = 0.4$ through 96 examples before increasing to $R = 0.91$ at 192. Increasing the budget to 384 examples improves point localization and slightly increases box geometry, although the class-sensitive score is marginally lower than at 192. At 768 examples, none of the three metrics improves over the 384-example configuration (Appendix B.2). These results support 384 examples as an effective joint-recovery configuration within the evaluated range, while showing that additional supervision does not uniformly improve every requirement.

7 Scope and Limitations

Our study covers two MoE backbones and one embodied adaptation mixture, with dense variants as auxiliary controls. Detailed ablations focus on Qwen3.5; the second repair seed tests repeatability from the same adapted checkpoint, not the complete adaptation pipeline. Generalization beyond the evaluated localization tasks and robot execution performance remain untested. Class-sensitive recovery approaches but remains below base, and broader performance changes modestly (Table 1). The underlying reasons for the different MoE and dense localization outcomes remain unresolved: our interventions characterize recoverability without fully explaining its architectural origins. Intervention protocols and reproducibility checks appear in Appendices A.4, D, F and G.

8 Conclusion

Embodied adaptation can improve task adherence and spatial reasoning while degrading spatial localization. Through complementary response and weight interventions, we show that substantial spatial competence remains recoverable, supporting a **re-aiming, not erasure** account of the observed degradation. Guided by these findings, JSR jointly supervises affected tasks and their required response structures. With 384 distinct examples and 32 updates, it substantially restores localization while largely preserving the broader benefits of adaptation. These results highlight the importance of jointly addressing spatial accuracy and task-specific response requirements when adapting VLMs for embodied use.

AI USE STATEMENT

Generative AI tools were used to assist with setting up and debugging the experimental environment, implementing and running experiments, organizing and formatting results, and improving the clarity and presentation of the manuscript and figures. The authors led the development of the core research ideas and methods and retained responsibility for the experimental methodology, protocols, and interpretation of the results.

All AI-assisted code, configurations, text, literature suggestions, and organized results were reviewed and verified by the authors. References were checked against their original sources. The authors take full responsibility for the final content, claims, results, and artifacts presented in this work.

REFERENCES

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-vl technical report, 2025. arXiv:2511.21631, 26 November 2025.

Dan Biderman, Jacob P. Portes, Jose Javier Gonzalez Ortiz, Mansheej Paul, Philip Greengard, Connor Jennings, Daniel King, Sam Havens, Vitaliy Chiley, Jonathan Frankle, Cody Blakeney, and John Patrick Cunningham. Lora learns less and forgets less. *Trans. Mach. Learn. Res.*, 2024, 2024.

Anthony Brohan et al. RT-2: Vision-language-action models transfer web knowledge to robotic control, 2023. arXiv:2307.15818.

Archie Chaudhury. Forgetting is not erasure: Recovering latent knowledge via transport keys, 2026. arXiv:2606.02860.

Anrui Chen, Ruijun Huang, Xin Zhang, Fang Dong, Hengjie Cao, Zhendong Huang, Yifeng Yang, Mengyi Chen, Jixian Zhou, Mingzhi Dong, Yujiang Wang, Jinlong Hou, Qin Lv, Robert P. Dick, Yuan Cheng, Tun Lu, Fan Yang, and Li Shang. Multi-head attention as a source of catastrophic forgetting in moe transformers, 2026. arXiv:2602.12587.

Boyuan Chen, Zhuo Xu, Sean Kirmani, Brian Ichter, Danny Driess, Pete Florence, Dorsa Sadigh, Leonidas Guibas, and Fei Xia. SpatialVLM: Endowing vision-language models with spatial reasoning capabilities, 2024a. arXiv:2401.12168.

Jinpeng Chen, Runmin Cong, Yuzhi Zhao, Hongzheng Yang, Guangneng Hu, Horace Ho Shing Ip, and Sam Kwong. SEFE: Superficial and essential forgetting eliminator for multimodal continual instruction tuning. In *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*, volume 267 of *Proceedings of Machine Learning Research*, 2025.

Lin Chen et al. Are we on the right way for evaluating large vision-language models?, 2024b. arXiv:2403.20330.

Matt Deitke et al. Molmo and PixMo: Open weights and open data for state-of-the-art vision-language models, 2024. arXiv:2409.17146.

Danny Driess et al. PaLM-E: An embodied multimodal language model, 2023. arXiv:2303.03378.

Mengfei Du, Binhao Wu, Zejun Li, Xuanjing Huang, and Zhongyu Wei. EmbSpatial-Bench: Benchmarking spatial understanding for embodied tasks with large vision-language models, 2024. arXiv:2406.05756.

Xingyu Fu, Yushi Hu, Bangzheng Li, Yu Feng, Haoyu Wang, Xudong Lin, Dan Roth, Noah A. Smith, Wei-Chiu Ma, and Ranjay Krishna. BLINK: Multimodal large language models can see but not perceive, 2024. arXiv:2404.12390.

Gemini Robotics Team et al. Gemini robotics: Bringing AI into the physical world, 2025. arXiv:2503.20020.

Zhiyan Hou, Haiyun Guo, Haokai Ma, Yandu Sun, Yonghui Yang, and Jinqiao Wang. Pass-moe: Mitigating misaligned co-drift among router and experts via pathway activation subspaces for continual learning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2026*, 2026. doi: 10.18653/v1/2026.acl-long.1474.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022*. OpenReview.net, 2022.

Yuqing Huang, Rongyang Zhang, Qimeng Wang, Chengqiang Lu, Yan Gao, Yi Wu, Yao Hu, Xuyang Zhi, Guiquan Liu, Xin Li, Hao Wang, and Enhong Chen. Selfaug: Mitigating catastrophic forgetting in retrieval-augmented generation via distribution self-alignment, 2025. arXiv:2509.03934.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In *The Eleventh International Conference on Learning Representations, ICLR 2023*, 2023.

Gangwei Jiang, Zhaoyi Li, Defu Lian, and Ying Wei. Refine large language model fine-tuning via instruction vector, 2024. arXiv:2406.12227.

Amita Kamath, Jack Hessel, and Kai-Wei Chang. What's "up" with vision-language models? investigating their struggle with spatial reasoning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, pp. 9161–9175. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.emnlp-main.568.

Sahar Kazemzadeh, Vicente Ordonez, Mark Matten, and Tamara Berg. Referitgame: Referring to objects in photographs of natural scenes. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 787–798. Association for Computational Linguistics, 2014. doi: 10.3115/v1/d14-1086.

Moo Jin Kim et al. OpenVLA: An open-source vision-language-action model, 2024. arXiv:2406.09246.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, March 2017. doi: 10.1073/pnas.1611835114.

He Li, Yuhui Zhang, Xiaohan Wang, Kaifeng Lyu, and Serena Yeung-Levy. Fine-tuning mllms without forgetting is easier than you think, 2026. arXiv:2603.14493.

Hongbo Li, Sen Lin, Lingjie Duan, Yingbin Liang, and Ness B. Shroff. Theory on mixture-of-experts in continual learning. In *The Thirteenth International Conference on Learning Representations, ICLR 2025*, 2025.

Yang Liu, Toan Nguyen, and Flora D. Salim. Cp-moe: Consistency-preserving mixture-of-experts for continual learning. In *Conference on Lifelong Learning Agents (CoLLAs)*, 2026.

Yuan Liu et al. MMBench: Is your multi-modal model an all-around player?, 2023. arXiv:2307.06281.

Yuhao Lu, Yixuan Fan, Beixing Deng, Fangfu Liu, Yali Li, and Shengjin Wang. VL-Grasp: A 6-dof interactive grasp policy for language-oriented objects in cluttered indoor scenes. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.

Yun Luo, Zhen Yang, Fandong Meng, Yafu Li, Jie Zhou, and Yue Zhang. An empirical study of catastrophic forgetting in large language models during continual fine-tuning. *CoRR*, abs/2308.08747, 2023. doi: 10.48550/ARXIV.2308.08747.

Minesh Mathew, Dimosthenis Karatzas, and C. V. Jawahar. Docvqa: A dataset for vqa on document images, 2020. arXiv:2007.00398.

NVIDIA et al. Cosmos-Reason1: From physical common sense to embodied reasoning, 2025. arXiv:2503.15558.

Qwen Team. Qwen3-vl-30b-a3b-instruct. Hugging Face model repository, `https://huggingface.co/Qwen/Qwen3-VL-30B-A3B-Instruct`, 2025. Model card and released configuration; repository created 2025-09-30; configuration (`config.json`) accessed 2026-09-02.

Qwen Team. Qwen3.5-35b-a3b. Hugging Face model repository, `https://huggingface.co/Qwen/Qwen3.5-35B-A3B`, 2026. Model card and released configuration; repository created 2026-02-24; configuration (`config.json`) accessed 2026-09-02.

Pierre Sermanet et al. RoboVQA: Multimodal long-horizon reasoning for robotics, 2023. arXiv:2311.00899.

Shuai Shao, Zeming Li, Tianyuan Zhang, Chao Peng, Gang Yu, Xiangyu Zhang, Jing Li, and Jian Sun. Objects365: A large-scale, high-quality dataset for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. arXiv:2402.03300.

Shengbang Tong, Ellis Brown, Penghao Wu, Sanghyun Woo, Manoj Middepogu, Sai Charitha Akula, Jihan Yang, Shusheng Yang, Adithya Iyer, Xichen Pan, Ziteng Wang, Rob Fergus, Yann LeCun, and Saining Xie. Cambrian-1: A fully open, vision-centric exploration of multimodal LLMs, 2024. arXiv:2406.16860.

Ayushman Trivedi and Bhavika Melwani. Catastrophic forgetting as accessibility collapse: A three-level framework for knowledge persistence in continual learning, 2026. arXiv:2606.06032.

Yihan Wang, Si Si, Daliang Li, Michal Lukasik, Felix Yu, Cho-Jui Hsieh, Inderjit S Dhillon, and Sanjiv Kumar. Two-stage llm fine-tuning with less specialization and more generalization. In *International Conference on Learning Representations (ICLR)*, 2024.

xAI. RealWorldQA. Dataset release accompanying Grok-1.5 Vision, `https://huggingface.co/datasets/xai-org/RealworldQA`, 2024.

Zhen-Hao Xie, Jun-Tao Tang, Yu-Cheng Shi, Han-Jia Ye, De-Chuan Zhan, and Da-Wei Zhou. Same: Stabilized mixture-of-experts for multimodal continual instruction tuning. In *International Conference on Machine Learning (ICML)*, 2026.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025. arXiv:2505.09388.

Jihan Yang, Shusheng Yang, Anjali W. Gupta, Rilyn Han, Li Fei-Fei, and Saining Xie. Thinking in space: How multimodal large language models see, remember, and recall spaces, 2024. arXiv:2412.14171.

Youngsik Yoon, Siwei Wang, Wei Chen, and Jungseul Ok. When are experts misrouted? counterfactual routing analysis in mixture-of-experts language models, 2026. arXiv:2605.07260.

Licheng Yu, Patrick Poirson, Shan Yang, Alexander C. Berg, and Tamara L. Berg. *Modeling Context in Referring Expressions*, pp. 69–85. Springer International Publishing, 2016. doi: 10.1007/978-3-319-46475-6_5.

Wentao Yuan, Jiafei Duan, Valts Blukis, Wilbert Pumacay, Ranjay Krishna, Adithyavairavan Murali, Arsalan Mousavian, and Dieter Fox. Robopoint: A vision-language model for spatial affordance prediction in robotics. In *Proceedings of The 8th Conference on Robot Learning (CoRL 2024)*, volume 270 of *Proceedings of Machine Learning Research*, pp. 4005–4020. PMLR, 2025.

Xiang Yue et al. MMMU: A massive multi-discipline multimodal understanding and reasoning benchmark for expert AGI, 2023. arXiv:2311.16502.

Yuexiang Zhai, Shengbang Tong, Xiao Li, Mu Cai, Qing Qu, Yong Jae Lee, and Yi Ma. Investigating the catastrophic forgetting in multimodal large language model fine-tuning. In *Conference on Parsimony and Learning (CPAL 2024)*, volume 234 of *Proceedings of Machine Learning Research*, pp. 202–227. PMLR, 2024.

Chongyang Zhao, Mingsong Li, Haodong Lu, and Dong Gong. On token's dilemma: Dynamic moe with drift-aware token assignment for continual learning of large vision language models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Junhao Zheng, Xidi Cai, Shengjie Qiu, and Qianli Ma. Spurious forgetting in continual learning of language models. In *International Conference on Learning Representations (ICLR)*, 2025.

Chenyu Zhou, Qiliang Jiang, and Boguang Pan. Dense coordinate-list fine-tuning induces a controllable interference surface in vision-language models, 2026. arXiv:2606.14507.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, and Omer Levy. LIMA: less is more for alignment. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

A Experimental setting details

A.1 Model selection and screening

We select Qwen3.5-35B-A3B and Qwen3-VL-30B-A3B-Instruct for their pre-adaptation spatial capabilities and distinct multimodal architectures. Preliminary screening yielded point-set F_1 scores of 13.5 for Kimi-VL-A3B-Instruct on 512 items and 4.7 for MolmoE-1B-0924 on 32 items, below the screening threshold of 15.0 fixed before evaluation. These screening scores use the 0–100 display scale. The unequal screening subsets informed model selection; the experiments use the fixed panels specified below.

A.2 Adaptation and repair configurations

Both MoE backbones are adapted using LoRA (Hu et al., 2022) on the eligible linear projections, with vision-encoder, multimodal-projector, and router parameters frozen. The embodied SFT mixture contains 78 sources and 2,026,179 rows. Qwen3.5-35B-A3B's five control runs use byte-identical training data and differ in training seed and data order. Two dense Qwen3.6-27B variants use the same mixture as auxiliary comparisons. Table 2 summarizes the adaptation and repair settings.

Setting	Embodied SFT	Default JSR
LoRA rank / scaling	128 / 256	128 / 256
Updates	4,088	32
Global batch size	24	24
Maximum sequence length	16,384	16,384
Sequence packing	Enabled	Enabled
Learning rate	10^{-4} to 10^{-5}	10^{-5}
Schedule	Cosine	Constant
Warmup fraction	0.05	0

Table 2: Training settings for embodied adaptation and the default repair.

JSR initializes a fresh adapter at the adapted endpoint. Its materialized training file contains 2,048 rows drawn from 384 distinct records: 192 point examples and 192 box examples. The point subset contains 78 RoboRefIt and 114 RoboPoint examples. The box examples cover three response structures, with 64 records each: a detection list under an `objects` field, a list of objects with a `bbox` field, and a list with a `bbox_2d` field. Each target retains the additional fields required by its instruction. Qwen3-VL-specific conventions are specified in Appendix H.

A.3 Evaluation panels

The original retention panel contributes 10,219 localization items: 3,290 point-prediction and 6,929 single-object box examples. The source archive also contains 46 What'sUp items (Kamath et al., 2023), which are excluded from localization scoring.

Point localization. The panel combines 1,996 referential examples from RoboPoint (Yuan et al., 2025), 997 object-reference examples from RoboRefIt (Lu et al., 2023), and 297 robot-placement examples from the 300-item FSD_points test split. The three subsets evaluate points within an indicated region, points corresponding to a named object, and target locations for manipulation, respectively.

Box localization. Single-object localization uses 6,929 RefCOCO referring-expression items (Yu et al., 2016; Kazemzadeh et al., 2014): 6,923 from validation and 6 from testA. Multi-object localization uses 5,000 Objects365 validation images (Shao et al., 2019). A permissive response instruction provides the primary comparison; an explicit JSON-schema instruction on the same images and targets measures response sensitivity.

Broader benchmarks. The spatial and embodied evaluations contain 2,638 CVBench (Tong et al., 2024), 1,901 BLINK (Fu et al., 2024), 3,640 EmbSpatial (Du et al., 2024), 510 Cosmos-Reason1 (NVIDIA et al., 2025), 5,130 VSI-Bench (Yang et al., 2024), 400 ERQA (Gemini Robotics Team et al., 2025), and 1,921 RoboVQA items (Sermanet et al., 2023). General evaluations contain

4,876 MMBench-V1.1 prediction items (Liu et al., 2023), 1,500 MMStar (Chen et al., 2024b), 765 RealWorldQA (xAI, 2024), 900 MMMU validation (Yue et al., 2023), and 5,349 DocVQA validation items (Mathew et al., 2020). MMBench option permutations are grouped for CircularEval scoring.

A.4 Training–evaluation separation

The embodied SFT mixture is checked against the retention panel using SHA-256 image hashes and perceptual hashes. No exact image matches or perceptual-hash matches within Hamming distance four are found. The 64-example point-repair set has no shared evaluation identifiers or exact image-content matches with the retention panel. JSR draws its point examples from screened training data and selects box examples after excluding evaluation identifiers and image-content hashes.

A.5 Scoring and diagnostic statistics

Localization scores. Point localization uses mean point-set $F_1 = 2m/(|\hat{y}| + |y|)$, where m is the number of one-to-one matches within 50 units on a 0–1000 coordinate canvas. Single-object localization uses Acc@0.5, the fraction of predictions with IoU at least 0.5. Multi-object box-set F_1 uses one-to-one greedy matching at IoU at least 0.5. Its class-sensitive version additionally requires matching class names. Unparseable or wrong-type predictions receive zero. The point and single-object box scorers return perfect scores when given the panel annotations as predictions; evaluation references are read independently from the panel and matched by item identifier.

Benchmark scores and display scale. Spatial and general benchmarks retain their respective scoring metrics. MMBench-V1.1 uses CircularEval, and general visual-question answering uses exact answer extraction followed by an LLM fallback. RoboVQA uses the qwen3.8-max judge, VSI-Bench uses its numerical-answer scoring alongside answer matching, and DocVQA uses ANLS. Scores defined on the unit interval are multiplied by 100 for the main results. Group averages are unweighted means of the displayed rows and summarize the profile alongside the individual task scores. Extraction rate measures the fraction of responses from which the scorer can obtain an answer.

Diagnostic statistics. The dimensions in Section 3 are measured using emission rates, output-type frequencies, predicted cardinalities and geometric scores. Response-structure compliance measures adherence to the requested template. Conditional accuracy is point-set F_1 or box IoU on parseable responses. Fixed-count comparisons further assess precision on examples where both model states emit the same number of points.

A.6 Paired comparisons and intervals

Comparisons pair predictions by evaluation item. Common-support analyses restrict the comparison to items parsed by both model states; full-panel scores include invalid outputs as zero. The original point and single-object box analyses use 10,000 bootstrap resamples clustered by source image or video. The Qwen3.5 Objects365 repair comparisons use 6,000 paired item resamples, and the September 19 broader-benchmark report uses 8,000. Intervals accompany the comparisons for which they were computed. Training-seed repetitions are reported separately from these resampling intervals.

A.7 Inference configuration and response protocols

Localization. Qwen3.5 coordinate-retention and Objects365 evaluations use separate fixed pipelines with greedy decoding, thinking disabled and a 2,048-token generation limit. Objects365 uses the native Qwen3.5 chat template through Swift/vLLM. The Qwen3-VL point-panel evaluation and model-specific configuration are described in Appendix H.

Spatial and embodied benchmarks. The image-based spatial evaluations use the RynnScale runner and RynnBrain prompt format, with bfloat16, FlashAttention 2, a minimum image-pixel setting of 1,048,576 and a 128-token generation limit. CVBench, EmbSpatial and BLINK use greedy decoding; ERQA uses temperature 0.2, top-$p = 0.95$ and top-$k = 50$. Video evaluations use Swift/vLLM with the benchmark-provided instruction construction, temperature zero, a 128-token

generation limit and up to 32 frames. General visual-question answering uses the recorded model generations with VLMEvalKit-based answer extraction and scoring.

Response instructions are matched across model states within each benchmark and backbone. For Qwen3.5, VSI-Bench adds a numerical-response system instruction on three estimation task types; the remaining inputs are unchanged. RoboVQA uses task-specific response instructions across the three states. Qwen3-VL uses the original VSI-Bench and RoboVQA instructions. DocVQA uniformly removes the injected thinking-prefix block before ANLS scoring. These conventions are applied consistently within each reported comparison.

B Repair-set composition and ablations

This appendix provides the complete configurations and scores for the coverage, budget and composition experiments. All runs start from the same adapted Qwen3.5-35B-A3B checkpoint. Point localization uses the 3,290-item panel, and box localization uses the 5,000-item Objects365 panel under the permissive instruction. Scores are reported on a 0–100 scale; base and SFT rows provide the recovery references.

B.1 Task and response-structure coverage

The configurations in Table 3 vary the allocation of 384 examples across tasks and response structures while retaining the same 32-update training schedule. The joint baseline and JSR use the same 192 pointing examples and 192 box examples; their box subsets cover one and three response structures, respectively. The two box-only configurations provide a corresponding structure-coverage comparison without pointing supervision. Class-sensitive box scores require both geometry and class agreement, complementing the geometry-only metric.

Supervision	Point ex.	Box ex.	Box structures	Pointing	Box geometry	Box geometry + class
Base	—	—	—	26.98	68.23	66.19
SFT endpoint	—	—	—	12.32	14.45	0.00
Pointing only	384	0	0	46.03	59.66	0.02
Box only	0	384	1	19.38	41.99	0.31
Box only	0	384	3	16.26	65.64	58.94
Joint baseline	192	192	1	44.78	58.21	2.75
JSR	192	192	3	45.00	67.57	59.17

Table 3: Task and response-structure coverage on Qwen3.5-35B-A3B. All repair configurations use 384 examples and 32 updates. Geometry + class requires both a geometric match and a correct class name.

B.2 Supervision budget

Table 4 reports the absolute scores used to compute the recovery ratios in the main-text budget analysis. Each repair set contains equal numbers of point and box examples, with the box examples distributed equally across three structures. Repeating each set to the same number of training rows holds the update count at 32 while varying the number of distinct examples from 24 to 768. Recovery ratios use the base and SFT reference scores separately for each metric.

B.3 Task balance and sequential supervision

The task-balance comparison in Table 5a holds the total at 384 examples and retains all three box-response structures. The 96/288, 192/192 and 288/96 point-to-box allocations expose the trade-off between point localization and box recovery. Increasing the box allocation slightly improves box scores, whereas the balanced allocation gives the highest point score among these configurations. Reducing the box allocation affects class-sensitive recovery more strongly than geometry alone.

For the sequential comparison in Table 5b, both treatments are derived from the same 384 box examples. The structure condition retains the requested output fields while shuffling coordinate

Repair examples	Pointing	Box geometry	Box geometry + class
Base	26.98	68.23	66.19
SFT endpoint	12.32	14.45	0.00
24	23.63	50.06	15.26
48	39.03	53.86	25.20
96	41.94	58.57	25.87
192	42.07	67.46	59.94
384	45.00	67.57	59.17
768	43.61	67.05	55.18

Table 4: Scores underlying the supervision-budget sweep in Figure 4b. All configurations use balanced task coverage, three box-response structures and 32 updates.

targets. The coordinate condition retains correct targets in a reduced response structure. Each individual treatment uses 32 updates; sequential runs apply the two treatments in either order for 64 updates in total.

(a) Task balance

Point ex.	Box ex.	Pointing	Box geom.	Box geometry + class
96	288	41.42	68.03	60.12
192	192	45.00	67.57	59.17
288	96	44.48	64.22	44.99

(b) Sequential corrections

Treatment	Upd.	Pointing	Box geometry
SFT endpoint	—	12.32	14.45
Structure	32	21.77	41.12
Coordinates	32	13.90	11.65
Structure → coord.	64	12.76	14.77
Coord. → structure	64	11.34	49.43

Table 5: Repair-set composition on Qwen3.5-35B-A3B. (a) Task balance at a fixed total of 384 examples, with three box-response structures covered. (b) Treatments derived from the same 384 box examples, applied individually for 32 updates or sequentially for 64 updates.

Structure-preserving supervision provides partial geometric recovery. Applying it after the coordinate treatment further raises box geometry to 49.43, whereas the reverse order yields 14.77. This order dependence shows that sequential composition can improve an individual requirement without achieving joint recovery of geometry and class-sensitive outputs. The results motivate maintaining jointly correct targets and response structures throughout repair.

C Detailed spatial diagnostics

C.1 Replication across runs and models

Table 6 reports endpoint changes for five Qwen3.5-35B-A3B runs and two dense variants. Each comparison includes both full-panel scores and scores on items parsed by both checkpoints. All five MoE runs decline, whereas both dense variants improve. The MoE runs share the training mixture and recipe and vary training seed and data order; the dense variants provide an auxiliary comparison.

Table 7 decomposes the Qwen3-VL-30B-A3B-Instruct point-localization results by task. The largest decline occurs in object reference, while referential localization and robot placement also decrease. This complements the Qwen3.5 pattern, whose largest endpoint losses occur in referential localization and robot placement.

C.2 Localization accuracy and response validity

Qwen3.5-35B-A3B exhibits different changes in point and single-object box localization. On referential localization, emission remains close to its base level while conditional point-set F_1 decreases substantially. Robot placement shows a similar pattern (Table 8). These results identify spatial errors among parseable responses as a substantial component of the endpoint loss.

For single-object boxes, emission increases from 0.6440 to 1.0000, while the paired IoU change on common support is -0.0005, with a 95% interval of $[-0.0056, 0.0047]$. The increase in full-panel Acc@0.5 from 60.59 to 93.75 therefore accompanies a large increase in valid box responses and

Run	ΔC (CI95)	Aggregate Δ, unparseable $= 0$ (CI95)	Δparse	n_{common}
Qwen3.5-35B-A3B, five control runs				
seed 42	-0.0643 [-0.0754, -0.0528]	-0.0821 [-0.0938, -0.0720]	-0.0690	2,940
seed 43	-0.0621 [-0.0747, -0.0529]	-0.0612 [-0.0738, -0.0520]	$+0.0043$	3,167
seed 44	-0.1486 [-0.1612, -0.1385]	-0.1466 [-0.1597, -0.1375]	$+0.0012$	3,159
seed 45, data order of seed 42	-0.0995 [-0.1114, -0.0899]	-0.1060 [-0.1180, -0.0967]	-0.0298	3,064
seed 42, second data order	-0.0359 [-0.0488, -0.0266]	-0.0340 [-0.0469, -0.0243]	$+0.0125$	3,193
Qwen3.6-27B dense, directional contrast				
variant A	$+0.0127$ [+0.0037, +0.0244]	$+0.0656$ [+0.0573, +0.0778]	—	2,314
variant B	$+0.0111$ [+0.0014, +0.0221]	$+0.0593$ [+0.0496, +0.0703]	—	2,247

Table 6: Endpoint (step 4,088) effects against the respective base model on the frozen 3,290-item pointing panel, same estimands and scoring chain for every row. ΔC is the paired change on the items both checkpoints parse; the aggregate change scores unparseable answers as zero; Δparse is the change in emission rate. Media-cluster bootstrap, B=10,000. The last two control runs vary the training seed and the data order separately from the first three. The dense rows come from a different pretraining lineage.

Task	Items	Base	SFT	JSR
Object reference	997	36.76	17.47	51.65
Referential localization	1,996	25.89	23.04	28.28
Robot placement	297	22.38	13.15	18.45

Table 7: Qwen3-VL-30B-A3B-Instruct point-localization scores under the same configuration as Table 1. Scores are point-set F_1 on a 0–100 scale, with invalid outputs scored zero.

little change in common-support geometry. The dense variants show the same qualitative distinction: conditional IoU remains near its base level while box emission increases.

		all items		*items* both *checkpoints answer*		
model / skill / checkpoint	n	emission	acc. when emitted	change	95% interval	n
Qwen3.5-35B-A3B: point localization						
referential, base	1996	0.9815	**0.3228**	—		
referential, endpoint	1996	0.9729	**0.1064**	-0.2151	$[-0.2303, -0.1998]$	1,907
placement, base	297	1.0000	**0.2333**	—		
placement, endpoint	297	0.9966	**0.0203**	-0.2122	$[-0.2414, -0.1820]$	296
Qwen3.5-35B-A3B: single-object box localization						
box, base	6929	0.6440	**0.8732**	—		
box, endpoint	6929	1.0000	**0.8651**	$\mathbf{-0.0005}$	$[-0.0056, +0.0047]$	4,462

Table 8: Response validity and conditional localization accuracy on Qwen3.5-35B-A3B. Emission is the fraction of parseable responses; conditional accuracy is point-set F_1 or box IoU on those responses. The common-support block reports paired changes on the same parsed items, with 10,000 media-cluster bootstrap resamples. Scores use the 0–1 scale.

C.3 Response type, cardinality and spatial precision

Task selection. A separate Qwen3.5-35B-A3B probe asks for a bounding box in JSON without prescribing a task-specific template. The adapted model substitutes a point on 24.7% of the 6,929 items and produces a parseable box on 39.8%. Of the substituted points, 95.80% lie inside the corresponding ground-truth box and 90.61% inside its central quarter. These responses retain information about the requested location while supplying a different type of spatial output.

The corresponding independent probe on Qwen3-VL-30B-A3B-Instruct returns points on 6,865 of 6,929 items (99.08%). Of these, 6,597 (96.10%) lie inside the requested ground-truth boxes. The probe requests a bounding box in JSON without a thinking-wrapper requirement and tests output-type selection separately from the point-localization evaluation in Table 1.

Output cardinality. On the 3,240 pointing items with multiple annotated points, the fraction answered with a single point increases from 41.5% at base to 92.7% in Qwen3.5's seed-44 endpoint.

The corresponding endpoint rates in seeds 42 and 43 are 39.4% and 40.9%. Thus, cardinality changes contribute to the loss differently across runs. They are assessed separately from task selection because multiple annotations may describe the same object.

Spatial precision. On 1,861 examples with three annotated points for which both the pre-decline checkpoint and adapted endpoint emit one point, precision falls from 0.3337 to 0.2719. The paired decline is 0.0618, with a 95% interval of $[0.0409, 0.0832]$. Spatial matching can therefore deteriorate at a fixed output count. Together, the validity, task-selection, cardinality and precision analyses distinguish the components of the observed localization changes.

D Additional parameter and prefix interventions

This appendix specifies the response-prefix and parameter interventions supporting the recoverability analysis. These experiments examine changes to a fixed adapted model separately from supervised repair.

D.1 Response-prefix interventions

For Qwen3.5-35B-A3B, we force a canonical response opening, such as `<point>`, and let the adapted model continue generation. The opening follows the response structure identified by the intervention procedure; longer conditions extend into reference content using a fixed first-divergence procedure. Items are selected using the healthy and adapted predictions before the intervention continuations are generated. Each forced condition is compared with a fresh unforced continuation on the same selected items.

Let $\bar{s}_H$ denote the healthy-reference score, $\bar{s}_0$ the unforced-continuation score, and $\bar{s}_L$ the score with a forced prefix. We measure recovery as

$$R_L = \frac{\bar{s}_L - \bar{s}_0}{\bar{s}_H - \bar{s}_0}.$$

This denominator measures the gap to the healthy reference within the intervention experiment.

On 692 preselected degraded items, using checkpoint 1,090 as the healthy reference, forcing the response opening recovers 58% of the gap at checkpoint 1,224 and 49% at the adapted endpoint. Recovery reaches 74% and 51%, respectively, on items where the forced opening differs from the model's own opening. The corresponding endpoint recovery is 6.7% when the opening is unchanged. Extending the prefix into content does not provide additional recovery.

D.2 Weight editing and cross-run transfer

For Qwen3.5-35B-A3B, the selected update is $\Delta W = W_{1224} - W_{1090}$ from training seed 44. Subtraction is applied to that run's adapted endpoint. Addition transfers the same update to seed 43 at step 409. Both operations directly edit the merged weights. The full-panel comparison in Figure 2 uses all 3,290 point-localization items. The controls below additionally report referential-localization and robot-placement scores separately.

D.3 Direction and trajectory controls

Table 9 reports the selected update, five alternative trajectory segments and a random-direction sweep. Random updates are matched to ΔW separately for each parameter tensor using its Frobenius norm. The scale sweep changes the magnitude of one random direction; additional seeds test independently sampled directions at full scale.

The trajectory controls vary the interval from which the update is extracted. Two alternative segments recover as much or more referential localization than the selected interval, despite being nearly orthogonal to its update. Other segments yield smaller gains or remain close to the endpoint. These comparisons identify structured recovery directions without restricting recovery to the observed decline interval.

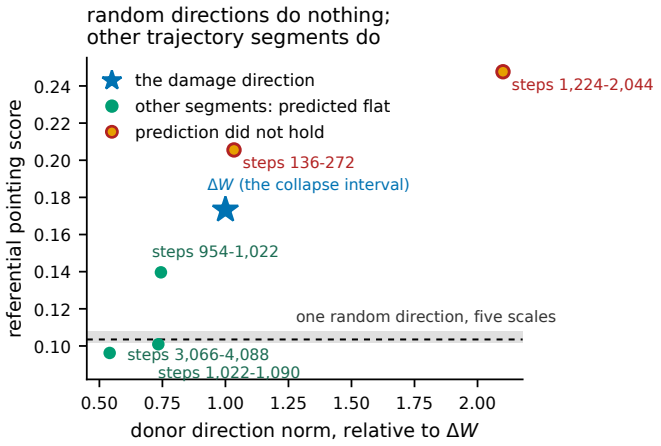


Figure 5: Referential localization after subtracting trajectory updates, plotted against update norm relative to ΔW. The gray band shows the norm-matched random-direction scale sweep. Red outlines identify conditions whose registered intervals did not hold. Scores use the 0–1 scale; the highest-norm trajectory update exceeds the random sweep range.

subtracted direction	$\|\Delta\|$ ratio	cos w/ ΔW	referential	placement	pre-registered interval
$\Delta W = W_{1224} - W_{1090}$	1.000	1.0000	0.1733	0.0859	reference
segment steps 1,224 to 2,044	2.101	-0.0122	0.2476	0.1189	**did not hold**
segment steps 136 to 272	1.034	-0.0030	0.2055	0.0286	**did not hold**
segment steps 954 to 1,022	0.744	-0.0147	0.1396	0.0354	held†
segment steps 1,022 to 1,090	0.734	+0.0336	0.1009	0.0288	held
segment steps 3,066 to 4,088	0.541	-0.0105	0.0962	0.0219	held
random, norm-matched ($\times 5$)	matched	≈ 0	0.1019–0.1082	0.0118–0.0253	10 of 10 held
no surgery (degraded endpoint)	—	—	0.1035	0.0202	—

Table 9: Trajectory and random-direction subtraction controls. Norms are relative to the selected update; scores use the 0–1 scale. Under the original rules, 13 of 15 registered interval predictions held. The two exceptions show stronger recovery from alternative trajectory segments. The † marks a later boundary-sensitivity observation; it does not replace the registered outcome.

Independent random directions. Three random seeds give referential-localization scores of 0.1076, 0.1034 and 0.1096, and robot-placement scores of 0.0202, 0.0202 and 0.0236. These remain close to the endpoint scores of 0.1035 and 0.0202, compared with 0.1733 and 0.0859 after subtracting ΔW. The three draws perturb the same checkpoint; the cross-run addition uses only the selected trajectory update.

Registered comparisons. Under the original decision rules, 13 of 15 interval predictions held. The two exceptions correspond to stronger-than-predicted recovery from alternative trajectory segments. A subsequent sensitivity analysis marks one additional condition as boundary-sensitive while retaining the original registered outcome. The table records both readings.

E Point-localization repair and replay controls

This appendix examines how task-specific supervision restores point localization and how replay preserves it during adaptation. These controls complement the joint-coverage experiments by varying the repair data's composition and size, and the stage at which supervision is introduced. We use *dose* to denote the number of distinct repair examples and report the recovery ratio R defined in Section 6.4.

E.1 Task identity and source coverage

We compare three repair conditions on the adapted Qwen3.5-35B-A3B endpoint, using 384 examples and a fresh LoRA adapter trained for 32 updates at a learning rate of 10^{-5}. Correct point-localization supervision drawn from all three sources yields $R = 2.30$, exceeding the base level. Using RoboRefIt alone also produces substantial recovery ($R = 1.61$, 95% interval $[1.531, 1.697]$), while unrelated-task supervision yields $R = -0.06$ ($[-0.106, -0.009]$). Thus, supervision of the affected task drives recovery, and broader source coverage further improves its extent (Figure 6a).

E.2 Repair dose and attained performance

Small repair sets already yield substantial recovery. With 16 examples, two Qwen3.5-35B-A3B runs exceed their base point-localization scores ($R = 1.87$ and 2.32 on the full panel). With 64 examples, referential localization reaches 0.3746 and 0.3671 in these runs, and 0.3773 in Qwen3-VL-30B-A3B-Instruct. Despite different initial levels of degradation, the three attained scores differ by less than 0.011. For Qwen3-VL, referential localization improves by 0.131 even though it had not declined in this repair experiment. These results suggest that task-specific supervision can establish a stronger level of performance rather than simply return the model to its earlier score.

The Qwen3-VL dose study is a separate point-only intervention, with its own base and adapted references: 0.3617 and 0.1817 for object-reference localization, and 0.2421 and 0.2467 for referential localization. Its 64-example results fall within the two intervals specified before evaluation: 0.3773 for referential localization (predicted 0.35–0.40) and 0.5748 for object-reference localization (predicted 0.56–0.61). The subsequent doses of 16, 256 and 1,024 examples satisfy eight of nine predictions across the three point-localization types. Object-reference localization reaches 0.4930 at 16 examples and increases to 0.5870 and 0.6196 at 256 and 1,024 examples, respectively (Figure 6b). The 16-example result lies below its predicted interval of $[0.5481, 0.6022]$, while the larger doses support the predicted recovery levels.

E.3 Replay during adaptation

Adding a small fraction of point-localization data during adaptation preserves much of the capability. On Qwen3.5-35B-A3B, two continuations from the same checkpoint differ only in whether 1.007% replay data is included. At step 1,224, replay attains 0.4481 on the point-localization panel, compared with 0.1855 for the control; the paired improvement is 0.2625 (95% interval $[0.249, 0.276]$).

A corresponding Qwen3-VL experiment tests predictions specified before training. At step 408, the control's lowest-scoring checkpoint, replay attains 0.2668 versus 0.0267 for object-reference localization, 0.3749 versus 0.1088 for referential localization, and 0.2277 versus 0.0034 for robot placement. Replay exceeds the base level on the latter two types and retains 73% of it on object-reference localization. Two of the three predicted lower bounds are met; object-reference localization shows substantial preservation but falls below the predicted bound of 0.3390. Together, these intermediate-checkpoint comparisons show that a small replay fraction can substantially reduce the decline during adaptation.

E.4 Effects on box localization

Point-localization supervision also affects box localization. On Qwen3.5-35B-A3B, replay retains a box-localization accuracy of 0.9319, close to the matched continuation control's 0.9352. For repair applied after adaptation, accuracy rises from 0.9043 to 0.9200 as the dose increases from 16 to 1,024 examples, compared with 0.9375 before repair (Figure 7). These results complement the joint-coverage analysis: point-only supervision produces strong recovery on its target task, while joint supervision explicitly addresses recovery across tasks and response structures.

F Diagnostic tests for candidate explanations

We examine routing, expert updates, adapter configuration, and response structure on Qwen3.5-35B-A3B to identify which interventions support localization recovery. These diagnostics complement the

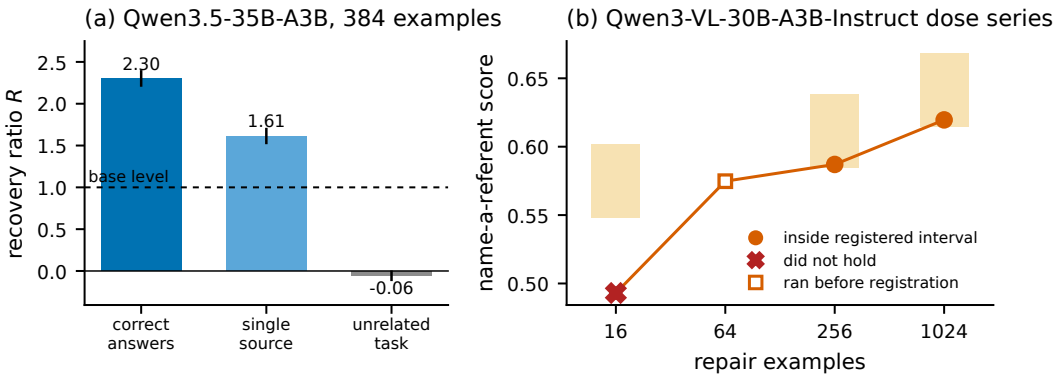


Figure 6: **Task-specific supervision supports recovery across repair budgets.** (a) Qwen3.5-35B-A3B repair conditions with 384 examples and 32 updates, evaluated on all 3,290 point-localization items. The single-source condition uses RoboRefIt alone; $R = 1$ marks the base level. (b) Qwen3-VL-30B-A3B-Instruct object-reference localization across repair doses. Shaded boxes show intervals specified before the corresponding runs; the open square marks the initial 64-example result used to inform these predictions.

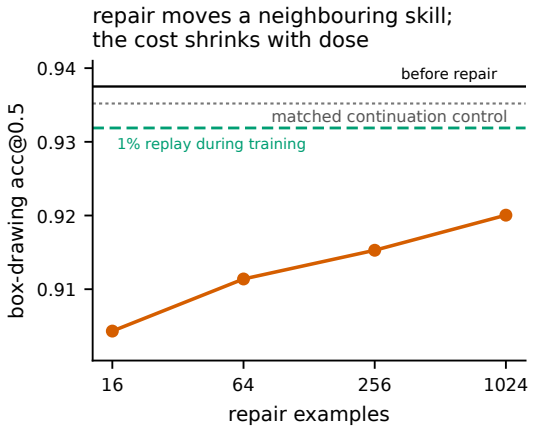


Figure 7: Box-localization accuracy at IoU ≥ 0.5 on all 6,929 box items following point-localization repair on Qwen3.5-35B-A3B. Larger repair doses retain more box accuracy. Replay during adaptation remains close to its matched continuation control (0.9319 versus 0.9352).

response and parameter interventions in Section 4; Table 11 summarizes the routing, expert, adapter, and prompt comparisons.

Routing interventions. Although router parameters remain frozen during adaptation, routing assignments can change as the input representations evolve. We therefore test both replaying base-model routes during training and fixing routing assignments at inference. Route replay recovers 16.7% of the decline at the evaluated checkpoint, below the specified 30% recovery target. At inference, replacing the adapted model's own routes with base routes recovers part of the localization loss in one run, but the effect reverses in an independent training seed. Route pinning also reduces performance on the adaptation task. Moreover, two runs exhibit similar routing drift at an early checkpoint despite different levels of localization loss. Thus, the tested routing interventions provide some recovery, but do not consistently preserve localization across runs.

Distribution of expert updates. Parameter-update energy is broadly distributed across experts. The pooled Gini coefficients are 0.009 at the early checkpoint and 0.013 at the endpoint. An effective expert count of 255.9 out of 256 and a top-eight share of 0.033, close to the uniform share of 0.031, likewise indicate broadly distributed update magnitudes. This diagnostic provides no clear basis for targeting a small subset of experts by update magnitude alone.

Adapter coverage and capacity. With matched microbatches, all-linear, expert-only, and no-expert adapter configurations attain localization scores within 0.01 of one another. Reducing expert rank from 128 to 4 changes the score by -0.006 (95% interval $[-0.012, 0.000]$), and balancing the loss contributions of response-structure and content tokens changes it by -0.001. The tested adjustments therefore yield similar performance without restoring localization, motivating a corrective supervision signal beyond changes to adapter placement or capacity.

Output-structure randomization. We vary templates and key sets in training examples with explicit output requirements while preserving their coordinate targets. This improves compliance with unseen templates, yet point localization remains below base in all four matched training blocks. On items jointly parseable under the base, control, and randomized conditions, the paired endpoint changes relative to base are -0.0859, -0.1081, -0.1111, and -0.1051, with intervals excluding zero. Improved structure compliance can thus coexist with degraded localization, supporting the joint assessment of response requirements and spatial accuracy.

Prompt wording. The adapted model's output compliance decreases from 1.000 to 0.927 when a literal task marker is removed, and to 0.573 when the request is rewritten as prose (Table 10). Nevertheless, it remains more compliant than the base model under all three prompt forms. Adaptation therefore improves adherence to the tested response requirements, while the size of the gain depends on the instruction form. The larger change under prose rewriting indicates sensitivity beyond a single task marker.

Table 10: Output compliance under three prompt forms on Qwen3.5-35B-A3B. Values are proportions on the 0–1 scale.

Checkpoint	With marker	Marker deleted	Prose
Base	0.268	0.640	0.011
SFT endpoint	1.000	0.927	0.573

G GRPO AND SUPERVISED REPAIR

We compare group-relative policy optimization (GRPO) (Shao et al., 2024) with supervised repair from the same adapted Qwen3.5-35B-A3B endpoint. The experiments examine the contribution of a verifiable reward at a small repair budget and recovery under the 384-example configuration used for JSR. Point-localization results use the same 3,290-item panel, with recovery ratio R defined in Section 6.4.

G.1 REWARD-GUIDED RECOVERY WITH 64 EXAMPLES

The initial comparison uses the same 64 point-localization examples and 32 updates for supervised repair, GRPO, and a reward-shuffled GRPO control. GRPO uses a fresh LoRA adapter with rank 128 and scaling factor 256, sampling 16 responses per prompt at temperature 0.7. A verifiable geometric reward evaluates the predicted coordinates against the annotated targets; the control shuffles rewards within each response group.

GRPO restores much of the lost point-localization performance, raising the score from 12.32 to 26.15, close to the base score of 26.98. Its recovery ratio is 0.943, compared with -0.056 for the shuffled-reward control (Table 12). Their paired difference of 0.999 (95% interval $[0.926, 1.072]$) supports the contribution of the geometric reward to recovery. Direct supervised repair reaches 42.77, exceeding the base level under the same example and update counts.

A pre-training sampling probe found zero within-group reward variance for 39 of the 64 prompts. Increasing sampling temperature and group size produced nonzero variance for about a third of these prompts. This observation illustrates a constraint on the group-relative learning signal: sampled responses must differ in reward to provide a task-reward advantage. Supervised repair supplies the target response directly, providing a corrective signal without this sampling requirement.

Test on Qwen3.5-35B-A3B	Value	Outcome
Routing		
Base-route replay during training, against a registered gate of $\geq 30\%$ recovery at step 409	16.7%	did not hold
Pinning to base routes minus pinning to own routes, step 409	$+0.0044$ $[-0.0050, +0.0141]$	unresolved
Route reassignment at the endpoint: content term, and its share of the control run's damage	$+0.0141$ $[+0.0050, +0.0234]$; 21.9% $[7.6, 38.0]$	minority
The same two quantities on seed 43, a registered replication	-0.0270 $[-0.0358, -0.0183]$; -44% $[-61, -28]$	reversed sign
Effect of pinning itself on parse rate at the endpoint, that run / seed 43	$+6.84$ pts $[+5.80, +7.90]$ / -0.06 pts	parse-rate effect
Plasticity cost of pinning on the new task (6,879 rows), registered non-inferiority	-0.0214 $[-0.0277, -0.0152]$	did not hold
Routing drift from base routes at step 409, seeds 42 / 43, whose damage states there are opposite	0.663 / 0.671; endpoint 0.73 / 0.73	similar drift
Experts		
Update-energy Gini pooled over layers, step 409 / endpoint	0.009 / 0.013; per layer 0.03–0.20	not concentrated
Effective expert count, and top-eight share against uniform	255.9 of 256; 0.033 against 0.031	not concentrated
Adapter		
LoRA coverage on matched micro-batches: all-linear / expert-only / no-expert	0.1907 / 0.2001 / 0.1995	within 0.01
Expert rank $128 \rightarrow 4$, against all-linear	-0.006 $[-0.012, +0.000]$	unresolved
Equal loss mass on template and content tokens, against all-linear	-0.001	unresolved
Prompt		
Compliance change under marker deletion / prose rewrite	-0.073 / -0.427	wording-sensitive

Table 11: Diagnostic interventions on Qwen3.5-35B-A3B, grouped by the component each probes. Bracketed ranges are 95% intervals; registered comparisons are assessed against their specified thresholds. Score changes use the 0–1 scale unless marked as percentages or percentage points.

Repair condition	R	95% interval
Supervised repair	$\mathbf{2.077}$	$[1.995, 2.161]$
GRPO	0.943	$[0.875, 1.012]$
GRPO with shuffled rewards	-0.056	$[-0.106, -0.005]$

Table 12: Point-localization recovery on Qwen3.5-35B-A3B with 64 examples and 32 updates. GRPO approaches the base level ($R = 1$), while supervised repair exceeds it. Intervals use 10,000 bootstrap resamples clustered by media identity.

G.2 Comparison at the JSR repair budget

We also evaluate GRPO on the same 384-example repair set used by JSR. With 32 updates and a KL penalty coefficient of $\beta = 0.04$, GRPO raises point localization to 15.71. Extending training to 128 updates increases the score to 20.3, while removing the KL penalty at 128 updates produces 14.8. These configurations provide partial recovery, with the highest score remaining below the base level of 26.98.

At 32 updates, multi-object box geometry increases from 14.45 to 15.60, class-sensitive box scoring remains zero, and single-object accuracy changes from 93.75 to 93.94. Together with the 64-example reward control, these results establish GRPO as a viable source of localization recovery, while the

evaluated supervised approach achieves stronger recovery. The coverage ablations in Section 6.4 examine how joint task and response-structure supervision supports the broader recovery achieved by JSR.

H SUPPLEMENTARY QWEN3-VL CONFIGURATION AND EVALUATION

Repair configuration. For Qwen3-VL-30B-A3B-Instruct, JSR starts from the embodied SFT endpoint and follows the task balance and box-response structures specified in Appendix A.2. Point-training instructions explicitly specify the required response structure. A fresh LoRA adapter is trained for 32 updates with rank 128, scaling parameter 256, learning rate 10^{-5}, global batch size 24, and maximum sequence length 16,384.

Evaluation configuration. Base, SFT, and JSR use the same inference configuration on the 3,290-item point-localization panel, with thinking enabled and a 2,048-token generation limit. RefCOCO uses its single-object localization prompts, which do not require a thinking wrapper. Multi-object localization uses 5,000 Objects365 images under both permissive and explicit-schema instructions. Broader benchmarks follow the response protocols in Appendix A.7; ERQA evaluates all 400 items under the same protocol across model states, and MMBench-V1.1 uses CircularEval. The capability profile is reported in Table 1.

Recovery under an explicit response schema. Complementing the permissive-instruction results in the main table, JSR also restores multi-object localization under an explicit response schema. Box geometry improves from 54.16 after SFT to 63.30, close to the base score of 64.68. The class-sensitive score increases from 0.02 to 59.97, exceeding the corresponding base score of 46.03. These results show that joint localization recovery extends across the two evaluated instruction forms. All scores use the 0–100 scale and the scoring definitions in Appendix A.5.