GuideVLA: Steering Vision-Language-Action Models with Soft Interventions for Out-of-Distribution Generalization

Anonymous ACL submission

Abstract

Vision-Language-Action (VLA) models have shown remarkable progress in grounding language instructions into executable robotic actions. However, current models exhibit significant performance degradation when confronted with previously unseen objects or linguistically rephrased commands, revealing a critical gap between in-distribution competence and the robust behavior required for real-world deployment. We propose activation steering as a simple yet effective training-free approach to improving VLA generalization. Our method computes averaged activation patterns from successful task executions and uses these as directional guides to steer the model's internal representations at inference time, enabling VLA models to better handle novel objects and diverse linguistic formulations. Experiments across diverse out-of-distribution conditions demonstrate consistent improvements: grasping success on unseen objects increases by 5.7%, task completion rates improve by 0.8%, and robustness to command rephrasing rises by 3.5% and 1.4% across grasping and success rates respectively. These gains require no architectural modifications and negligible computational overhead, making activation steering a practical and readily deployable strategy for enhancing the robustness of existing VLA systems.

1 Introduction

Activation steering has proven effective for controlling model behavior without retraining in large language models (LLMs) (Turner et al., 2024; Li et al., 2023), and has recently been extended to vision-language models (VLMs) for tasks such as hallucination mitigation and improved visual grounding (Wu et al., 2025; Balakrishnan and Phute, 2026; Sivakumar et al., 2025). Yet its applicability to Vision-Language-Action (VLA) models remains largely unexplored, despite VLA architectures inheriting the same transformer-based backbones that make steering viable in their language-only and vision-language counterparts.

Existing approaches to VLA generalization predominantly rely on scaling pretraining data and co-finetuning across diverse tasks and embodiments (Zitkovich et al., 2023; Kim et al., 2025). While effective within their training distribution, such strategies require substantial computational resources and large-scale robotic datasets, and do not directly improve robustness to novel objects and rephrased instructions at deployment time (Dey et al., 2025; Fei et al., 2025). Inference-time alternatives based on external LLMs as trajectory-level judges (Du et al., 2023; Lù et al., 2025) offer guided behavior selection but introduce significant overhead and depend on proprietary models, leaving the model's internal representations untouched.

We propose to fill this gap by applying activation steering directly to VLA models at inference time. Unlike text generation, VLA models must produce action sequences where each prediction is physically grounded and coupled to the next environment state. This introduces a novel temporal dependency into the generation process, raising research questions regarding precisely how steering vectors should be computed and applied within such a dynamic environment. Our method computes steering vectors from averaged activations of successful episodes and applies them step-by-step during the pre-grasp phase, guiding internal representations toward regions associated with successful task execution without any parameter updates.

Our contributions are as follows:

- Activation steering is applied to the VLA, demonstrating its viability as a training-free intervention for improving generalization.
- A temporally-aware steering mechanism that accounts for the sequential and physically grounded nature of action generation.

- Comprehensive evaluations across diverse out-of-distribution conditions demonstrate consistent gains in grasping accuracy, task success, and robustness to linguistic reformulations.

2 Related Work

Activation Steering for VLMs Activation steering has emerged as a practical approach to guiding language and vision-language models by exploiting the finding that behavioral concepts are encoded as approximately linear directions in transformer residual streams (Zou et al., 2023; Turner et al., 2024; Rimsky et al., 2024; Li et al., 2023). Hua et al. (2025) apply SAE-derived directions to steer LVLM residual streams and reduce hallucinations, while Laptev et al. (2025) enable multi-layer steering by tracking SAE features across consecutive layers. Both methods apply a single fixed vector regardless of input; Parekh et al. (2025) address this by training a small auxiliary network to predict input-dependent steering vectors.

VLA Models VLA models extend pretrained vision-language backbones to robotic action generation by treating actions as tokens. RT-2 and OpenVLA (Zitkovich et al., 2023; Kim et al., 2025) established and democratized this paradigm, while SpatialVLA (Qu et al., 2025) extended it with explicit 3D spatial reasoning. Despite strong in-distribution performance, VLA models remain brittle to shifts in object appearance, instruction phrasing, and their combination (Dey et al., 2025; Fei et al., 2025). INT-ACT (Fang et al., 2025) formalizes this as an *intention-action gap*: VLM backbones handle out-of-distribution inputs, but execution degrades, pointing to action-generation rather than semantics. Because VLAs generate action tokens in the same residual stream as vision and language, LLM steering transfers directly.

Inference-Time Guidance for VLA Policies Several recent works propose inference-time guidance for VLA policies without modifying model weights. VLS (Liu et al., 2026) steers diffusion-policy denoising via VLM-generated reward functions, VLA-Pilot (Li et al., 2025) uses MLLM as a verifier, and OmniGuide (Song et al., 2026) injects 3D energy fields from foundation models. Although training-free, these approaches depend on external models, introducing computational overhead and additional failure modes. Our work pursues a lighter alternative: steering the model's own internal representations with no external models, gradient updates, or additional parameters.

3 Methodology

3.1 Problem Formulation

A VLA model defines a policy $\pi(a_t \mid o_t, \mathbf{c})$, where o_t is the visual observation at step t, $\mathbf{c}$ is the natural language task instruction, and $a_t \in R^D$ is the predicted robot action generated by decoding K discrete action tokens produced autoregressively per environment step, so the model's internal computations are organized along two temporal axes: the environment step index s and the within-step action-token index k. We hypothesize that the hidden states produced during successful in-distribution episodes occupy a structured region of the model's representation space, and that this region encodes direction information that can partially compensate for distribution shift when injected at inference time. Let $\mathcal{D}^+$ and $\mathcal{D}^-$ denote the sets of successful and unsuccessful episodes from in-distribution tasks. For episode e, environment step s, action-token index k, and transformer layer l, we write $\mathbf{h}^{(l)}_{s,k}(e) \in R^d$ for the hidden state extracted at the last sequence position during generation of the k-th action token on step s.

3.2 Steering Vector Computation

We consider two methods for computing a steering vector $\mathbf{v}^{(l)}_{s,k}$ per triple (s, k, l): mean steering (Turner et al., 2024) and difference-in-means steering (Li et al., 2023; Zou et al., 2023). Unlike the application of steering vectors in LLMs, for VLA models, it is necessary to aggregate and incorporate the vector distribution for each temporal step s.

Mean Steering This method sets $\mathbf{v}^{(l)}_{s,k} = |\mathcal{D}^+|^{-1} \sum_{e \in \mathcal{D}^+} \mathbf{h}^{(l)}_{s,k}(e)$, the mean hidden state over successful episodes. This vector serves as a *prototype* of the representations associated with successful execution: it captures the average "look" of the model's internal state when the task is proceeding correctly, and steering toward it nudges out-of-distribution activations into a more favorable region of representation space. Because it depends only on $\mathcal{D}^+$, mean steering is insensitive to the diversity or noise of failure episodes.

Difference-in-Means Steering This method sets $\mathbf{v}^{(l)}_{s,k} = |\mathcal{D}^+|^{-1} \sum_{e \in \mathcal{D}^+} \mathbf{h}^{(l)}_{s,k}(e) - |\mathcal{D}^-|^{-1} \sum_{e \in \mathcal{D}^-} \mathbf{h}^{(l)}_{s,k}(e)$, the linear direction

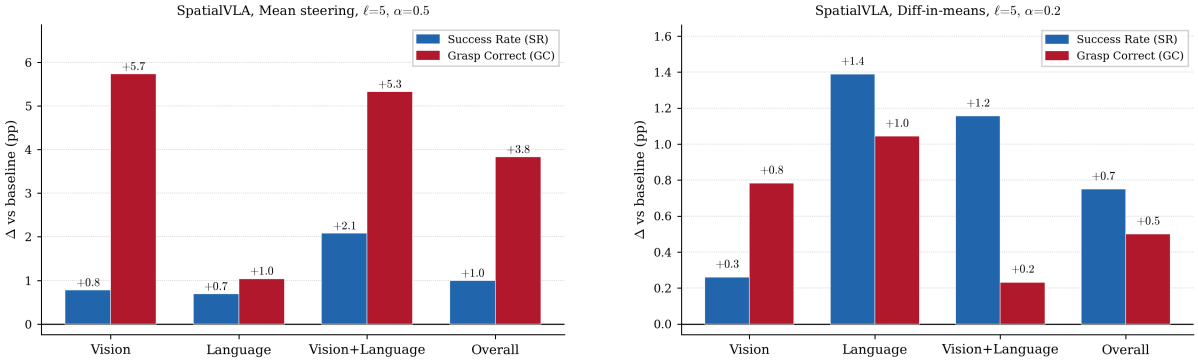

Figure 1: Per-category ΔSR and ΔGC (pp vs. baseline) for SpatialVLA at layer $\ell = 5$: Mean Steering with $\alpha = 0.5$ (left) and Diff-in-Means Steering with $\alpha = 0.2$ (right). Both configurations produce positive improvements across all four task categories on both metrics, with Mean Steering achieving stronger GC gains (up to $+5.7$ pp on Vision tasks) and Diff-in-Means delivering more balanced SR improvements (up to $+1.4$ pp on Language tasks).

Method	α	*Vision*		*Language*		*Vision+Language*		**Overall**	
		SR	GC	SR	GC	SR	GC	SR	GC
Baseline	—	16.4	33.9	28.1	46.2	16.4	38.4	21.3	40.5
Mean steering	0.5	17.2 (+0.8)	39.6 (+5.7)	28.8 (+0.7)	47.2 (+1.0)	18.5 (+2.1)	43.7 (+5.3)	22.2 (+0.9)	44.3 (+3.8)
Diff-in-means	0.2	16.7 (+0.3)	34.7 (+0.8)	29.5 (+1.4)	47.2 (+1.0)	17.6 (+1.2)	38.6 (+0.2)	22.0 (+0.7)	41.0 (+0.5)

Table 1: SpatialVLA activation steering at layer $\ell = 5$. SR = Success Rate (%), GC = Grasp Correct (%). Deltas are relative to baseline (+ improvement, − degradation).

in activation space separating the *failure manifold* from the *success manifold* of the model's latent representations. This is analogous to the Contrastive Activation Addition framework of Rimsky et al. (2024), adapted here to the step- and token-level structure of action generation.

3.3 Inference-Time Steering

At inference time, the precomputed vector is injected additively into the model's hidden states before the remaining transformer layers process them: $\tilde{\mathbf{h}}^{(l)}_{s,k} = \mathbf{h}^{(l)}_{s,k} + \alpha \cdot \mathbf{v}^{(l)}_{s,k} \cdot \mathbf{1}[s < S^*(e)]$, where $\alpha \in R$ is a scalar multiplier that controls the strength of the intervention, and $S^*(e)$ denotes the environment step at which episode e first achieves a successful grasp. The indicator $\mathbf{1}[s < S^*(e)]$ formally restricts the intervention to the pre-grasp phase: steering is active for all steps preceding the grasp event and is automatically disabled thereafter. This design is intentional: our goal is to investigate whether a lightweight intervention confined to the pre-grasp sub-sequence is sufficient to propagate improvements through to full-trajectory task success, and more broadly, to demonstrate that activation steering can serve as a viable tool for improving robotic behavior without any modification to model parameters or training procedures.

4 Experiments

4.1 Setup

We evaluate activation steering on **SpatialVLA**, fine-tuned on the BridgeV2 dataset and evaluated in the SIMPLER simulation environment (Li et al., 2024). We focus on SpatialVLA because, as an autoregressive VLA, it generates action tokens through the same residual stream as visual and language tokens, making it directly amenable to activation-level interventions without architectural modifications.

The evaluation suite covers 51 manipulation tasks drawn from the INT-ACT benchmark (Fang et al., 2025), grouped into four categories: *Original* (in-distribution), *Vision* (novel objects and backgrounds), *Language* (rephrased or negated instructions), and *Vision+Language* (combined visual and linguistic perturbations).

Both steering strategies are applied *only during the pre-grasp phase* of each trajectory and are derived exclusively from *Original*-category episodes, with no model parameters updated at any stage. We additionally consider two variants of the source episode set: *all successful grasps* and *full success only* (marked ⋆ in Table A1), retaining only episodes where the robot both grasped correctly

and completed the full task.

Steering is applied at a single transformer layer with a scalar multiplier $\alpha \in \{0.2, 0.5\}$ across layers $\{5, 10, 15\}$. We report two metrics: *Success Rate* (SR), measuring end-to-end task completion, and *Grasp Correct* (GC), measuring whether the robot grasps the correct object. Each configuration is evaluated over 24×51 episodes; grasp detection relies on the SIMPLER environment's signal.

4.2 Main Results

Table 1 and Figure 1 present results for SpatialVLA steered at layer $\ell = 5$ under two configurations that both yield consistent, broad improvements across all task categories.

Mean Steering produces strong, cross-category gains. At $\ell = 5$, $\alpha = 0.5$, Mean Steering improves SpatialVLA across every category on both metrics: Overall SR rises by $+0.9$ pp and Overall GC by $+3.8$ pp relative to the baseline (SR = 21.3%, GC = 40.5%), with the largest gains on Vision+Language tasks ($+2.1$ pp SR, $+5.3$ pp GC) and Vision tasks ($+5.7$ pp GC). The magnitude of the GC improvement on Vision and Vision+Language categories is particularly noteworthy: a $+5.7$ pp gain on Vision GC indicates that the steering vector encodes object-agnostic grasping representations that transfer across visually diverse scenarios, even though the vector was derived exclusively from in-distribution episodes. These results suggest that the steering vector captures a broadly applicable direction in activation space rather than features specific to any individual task group, and that a single early-layer intervention suffices to propagate this benefit through to end-to-end task success.

Diff-in-Means Steering provides complementary SR improvements. Diff-in-Means Steering at $\ell = 5$, $\alpha = 0.2$ also improves all four categories, with more uniform SR gains ($+0.3$ to $+1.4$ pp) and Language tasks benefiting most ($+1.4$ pp SR, $+1.0$ pp GC); GC gains are more modest, consistent with the noisier failure-derived component of the Diff-in-Means vector. Importantly, the $+1.4$ pp SR gain on Language tasks exceeds that of Mean Steering on the same axis, suggesting that the failure-contrast component encodes task-instruction sensitivity not fully captured by the success mean alone. Together, these two configurations demonstrate that steering at a single early layer suffices to produce consistent improvements across all out-of-distribution categories without any modification to model weights, and that the two methods offer complementary strengths: Mean Steering maximises GC improvements on visually perturbed tasks, while Diff-in-Means provides more balanced SR gains across linguistic reformulations.

4.3 Group-level Tradeoffs

The aggregate results above, however, mask variation across individual task groups: not all configurations improve all three held-out categories simultaneously. Figure A1 (Appendix A) plots, for each evaluated configuration, the mean Δ against the worst-group Δ over the three held-out categories (Vision, Language, Vision+Language).

Balanced improvement is achievable. Several configurations reach the green quadrant where all groups improve simultaneously. Mean Steering at $\ell = 5$, $\alpha = 0.5$ achieves $+0.8$ pp worst-group SR and $+1.2$ pp mean SR; the GC picture is even more favourable, with $+1.0$ pp worst-group GC and $+4.0$ pp mean GC, and three additional Mean Steering configurations also falling in the green quadrant (Table A1, Appendix A). Category-targeted steering vectors derived from group-specific episodes are a natural direction for extending balanced improvements to a wider range of configurations.

5 Conclusion

We have demonstrated that activation steering is a viable training-free intervention for improving out-of-distribution generalization in autoregressive VLA models. Applied exclusively during the pre-grasp phase and derived from a small set of in-distribution successful episodes, steering vectors reliably shift SpatialVLA's internal representations toward directions associated with robust task execution, producing consistent gains in both Success Rate and Grasp Correct across all three out-of-distribution categories: Vision, Language, and Vision+Language. These results suggest that the representations learned by current VLA models already encode generalizable execution knowledge; inference-time steering provides a lightweight mechanism to unlock it without any modification to model parameters, training data, or architecture.

Ethics

This work is concerned with improving the out-of-distribution robustness of robotic manipulation policies in simulation. We do not foresee direct ethical harms from the proposed method itself. However, we note that activation steering, as a lightweight inference-time intervention, could in principle be misused to redirect model behavior in unintended or adversarial directions — a consideration that applies broadly to representation-level control methods. As VLA models are deployed in physical environments, ensuring that steering vectors are derived from verified, high-quality episode sets becomes important for safety. All experiments in this work were conducted in the SIMPLER simulation environment using publicly available benchmarks and model checkpoints, with no involvement of human subjects or sensitive data.

Limitations

Our evaluation is conducted exclusively in the SIMPLER simulation environment; whether the observed gains transfer to real-robot deployment remains to be verified.

Steering is applied only during the pre-grasp phase, leaving post-grasp execution unaffected. This design choice is consistent with our intervention target, but it limits the potential ceiling of SR improvements: errors in placing or releasing the object after a successful grasp lie outside the scope of our method.

Finally, our experiments focus on SpatialVLA as a single representative autoregressive VLA architecture. Generalisation of the proposed steering approach to other autoregressive VLAs, such as OpenVLA or RT-2, is a natural direction for future work.

References

Ravikumar Balakrishnan and Mansi Phute. 2026. VISOR++ : VISUAL INPUT BASED STEERING FOR LARGE VISION LANGUAGE MODELS.

Sombit Dey, Jan-Nico Zaech, Nikolay Nikolov, Luc Gool, and Danda Paudel. 2025. ReVLA: Reverting visual domain limitation of robotic foundation models. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Yuqing Du, Ksenia Konyushkova, Misha Denil, Akhil Raju, Jessica Landon, Felix Hill, Nando de Freitas, and Serkan Cabi. 2023. Vision-language models as success detectors.

Irving Fang, Juexiao Zhang, Shengbang Tong, and Chen Feng. 2025. From intention to execution: Probing the generalization boundaries of vision-language-action models. *arXiv preprint arXiv:2506.09930*.

Senyu Fei, Siyin Wang, Junhao Shi, Zihao Dai, Jikun Cai, Pengfang Qian, Li Ji, Xinzhe He, Shiduo Zhang, Zhaoye Fei, Jinlan Fu, Jingjing Gong, and Xipeng Qiu. 2025. LIBERO-Plus: In-depth robustness analysis of vision-language-action models. *arXiv preprint arXiv:2510.13626*.

Zhenglin Hua, Jinghan He, Zijun Yao, Tianxu Han, Haiyun Guo, Yuheng Jia, and Junfeng Fang. 2025. Steering LVLMs via sparse autoencoder for hallucination mitigation. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 10808–10828.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan P Foster, Pannag R Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. 2025. Openvla: An open-source vision-language-action model. In *Proceedings of The 8th Conference on Robot Learning*, volume 270, pages 2679–2713.

Daniil Laptev, Nikita Balagansky, Yaroslav Aksenov, and Daniil Gavrilov. 2025. Analyze feature flow to enhance interpretation and steering in language models. In *Forty-second International Conference on Machine Learning*.

Kenneth Li, Oam Patel, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. 2023. Inference-time intervention: Eliciting truthful answers from a language model. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Xuanlin Li, Kyle Hsu, Jiayuan Gu, Karl Pertsch, Oier Mees, Homer Rich Walke, Chuyuan Fu, Ishikaa Lunawat, Isabel Sieh, Sean Kirmani, Sergey Levine, Jiajun Wu, Chelsea Finn, Hao Su, Quan Vuong, and Ted Xiao. 2024. Evaluating real-world robot manipulation policies in simulation. *arXiv preprint arXiv:2405.05941*.

Zhuo Li, Junjia Liu, Zhipeng Dong, Tao Teng, Quentin Rouxel, Darwin Caldwell, and Fei Chen. 2025. Towards deploying VLA without fine-tuning: Plug-and-play inference-time VLA policy steering via embodied evolutionary diffusion. *arXiv preprint arXiv:2511.14178*.

Shuo Liu, Ishneet Singh, Yiqing Xu, Jiafei Duan, and Ranjay Krishna. 2026. VLS: Steering pretrained robot policies via vision–language models. *arXiv preprint arXiv:2602.03973*.

Xing Han Lù, Amirhossein Kazemnejad, Nicholas Meade, Arkil Patel, Dongchan Shin, Alejandra Zambrano, Karolina Stanczak, Peter Shaw, Christopher Pal, and Siva Reddy. 2025. Agentrewardbench: Evaluating automatic evaluations of web agent trajectories. In *Second Conference on Language Modeling*.

Jayneel Parekh, Pegah Khayatan, Mustafa Shukor, Arnaud Dapogny, Alasdair Newson, and Matthieu Cord. 2025. Learning to steer: Input-dependent steering for multimodal llms. *Advances in Neural Information Processing Systems*.

Delin Qu, Haoming Song, Qizhi Chen, Yuanqi Yao, Xinyi Ye, Yan Ding, Zhigang Wang, JiaYuan Gu, Bin Zhao, and Dong Wang. 2025. SpatialVLA: Exploring spatial representations for visual-language-action model. In *Robotics: Science and Systems (RSS)*.

Nina Rimsky, Nick Gabrieli, Julian Schulz, Meg Tong, Evan Hubinger, and Alexander Turner. 2024. Steering Llama 2 via contrastive activation addition. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15504–15522.

Anushka Sivakumar, Andrew Zhang, Zaber Ibn Abdul Hakim, and Chris Thomas. 2025. SteerVLM: Robust model control through lightweight activation steering for vision language models. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 23640–23665, Suzhou, China. Association for Computational Linguistics.

Yunzhou Song, Cleave Le, Jongwoo Park, Ruijie Wang, Haochen Shi, Minghan Liu, Jiayi Gu, Eric Eaton, Dinesh Jayaraman, and Kostas Daniilidis. 2026. OmniGuide: Universal guidance fields for enhancing generalist robot policies. *arXiv preprint arXiv:2603.10052*.

Alexander Matt Turner, Lisa Thiergart, David Udell, Gavin Leech, Ulisse Mini, and Monte Pelrine. 2024. Steering language models with activation engineering. *arXiv preprint arXiv:2308.10248*.

Yilin Wu, Ran Tian, Gokul Swamy, and Andrea Bajcsy. 2025. From foresight to forethought: Vlm-in-the-loop policy steering via latent alignment. In *Robotics: Science and Systems (RSS)*.

Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, and Ayzaan Wahid. 2023. RT-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning (CoRL)*, volume 229 of *Proceedings of Machine Learning Research*, pages 2165–2183.

Andy Zou, Long Phan, Sarah Chen, James Campbell, Phillip Guo, Richard Ren, Alexander Pan, Xuwang Yin, Mantas Mazeika, and Ann-Kathrin Dombrowski. 2023. Representation engineering: A top-down approach to AI transparency. *arXiv preprint arXiv:2310.01405*.

A Balanced Tradeoff Analysis

For each evaluated configuration we plot the mean Δ against the worst-group Δ across the three held-out categories (Vision, Language, Vision+Language). Figure A1 shows this for SR and GC; Table A1 lists all configurations whose worst-group Δ is non-negative on at least one metric.

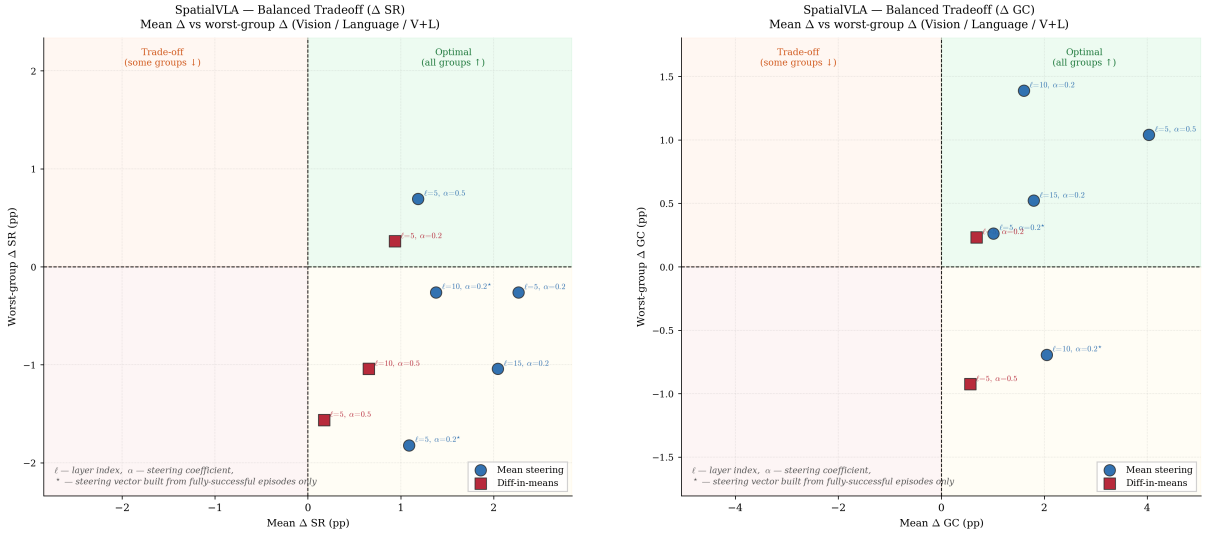

Figure A1: Balanced tradeoff analysis for SpatialVLA: mean Δ vs. worst-group Δ across Vision, Language, and Vision+Language categories, for SR (left) and GC (right). Each point represents one steering configuration (layer, α, source set). The green quadrant (upper-right) marks configurations where *all* task groups improve simultaneously; the orange quadrant (upper-left) marks configurations with a positive mean gain but at least one group regressing. Mean Steering at $\ell = 5$, $\alpha = 0.5$ reaches the green quadrant on both metrics, demonstrating that uniform cross-group improvement is achievable.

Table A1: SpatialVLA activation steering: balanced configurations. SR = Success Rate (%), GC = Grasp Correct (%). Deltas relative to baseline (+ improvement, − degradation). * steering vector built from fully-successful episodes only.

			Vision		*Language*		*Vision+Language*		**Overall**	
Method	ℓ	α	SR	GC	SR	GC	SR	GC	SR	GC
Baseline	—	—	16.4	33.9	28.1	46.2	16.4	38.4	21.3	40.5
Mean steer.	5	0.2	16.1 (−0.3)	33.1 (−0.8)	31.2 (+3.1)	47.6 (+1.4)	20.3 (+3.9)	40.7 (+2.3)	23.1 (+1.8)	41.7 (+1.2)
	5	0.5	17.2 (+0.8)	39.6 (+5.7)	28.8 (+0.7)	47.2 (+1.0)	18.5 (+2.1)	43.7 (+5.3)	22.2 (+0.9)	44.3 (+3.8)
	15	0.2	15.4 (−1.0)	34.4 (+0.5)	30.2 (+2.1)	47.6 (+1.4)	21.5 (+5.1)	41.9 (+3.5)	23.6 (+2.3)	42.8 (+2.3)
	10	0.2*	16.1 (−0.3)	37.0 (+3.1)	30.9 (+2.8)	45.5 (−0.7)	18.0 (+1.6)	42.1 (+3.7)	22.5 (+1.2)	42.8 (+2.3)
	5	0.2*	14.6 (−1.8)	34.2 (+0.3)	31.6 (+3.5)	46.9 (+0.7)	18.0 (+1.6)	40.5 (+2.1)	22.4 (+1.1)	41.8 (+1.3)
Diff-in-means	5	0.2	16.7 (+0.3)	34.7 (+0.8)	29.5 (+1.4)	47.2 (+1.0)	17.6 (+1.2)	38.6 (+0.2)	22.0 (+0.7)	41.0 (+0.5)
	10	0.5	15.4 (−1.0)	34.2 (+0.3)	28.8 (+0.7)	44.1 (−2.1)	18.7 (+2.3)	37.9 (−0.5)	22.2 (+0.9)	40.2 (−0.3)
	5	0.5	14.8 (−1.6)	35.5 (+1.6)	30.2 (+2.1)	47.2 (+1.0)	16.4	37.5 (−0.9)	20.9 (−0.4)	41.2 (+0.7)