WHERE IS POST-DEGRADATION REPAIR LIMITED? EVIDENCE FROM CAMERA-BASED 3D OCCUPANCY

Anonymous authors
Paper under double-blind review

ABSTRACT

In camera-based 3D occupancy, post-deployment repair aims to recover task performance when observation quality degrades, and its effectiveness is commonly evaluated by the amount of performance recovered. Under a fixed model, a small final task gain indicates limited recovery but does not reveal where the repair process is constrained. Motivated by this ambiguity, we separate the analysis into local corrective evidence, privileged task gain, and inference-time task gain. We evaluate their relations using matched interventions across four nuScenes occupancy anchors and a controlled CARLA setting. This evaluation leads to three main observations. First, local corrective evidence can persist through finite-objective improvement while a positive task-metric advantage over the matched control remains unestablished. Second, a positive privileged task gain can correspond to substantially smaller or negative inference-time task gain. Third, inference-time task gain varies with allocation policy and prescribed intervention capacity, and can also change sign across semantic objectives even when the policy is fixed. In summary, these findings motivate examining how corrective evidence translates into finite task gain.

1 INTRODUCTION

Representative camera-based 3D perception systems construct BEV or 3D scene representations from multi-view images (Philion & Fidler, 2020; Huang et al., 2021; Li et al., 2022). Their performance depends on the quality and availability of these observations and can deteriorate when views are corrupted, occluded, missing, or otherwise unavailable after deployment (Xie et al., 2023; Hao et al., 2024). For 3D occupancy perception, such degradation reduces the visual evidence available for scene reconstruction. Existing work addresses degraded perception through several operational routes. Missing-view reconstruction, robust representation learning, and robust sensor fusion improve robustness to degraded observations (Ge et al., 2023; Wang et al., 2024b; Hao et al., 2025). Test-time adaptation uses target observations to adapt deployed predictors, often through parameter or model-state updates (Chen et al., 2025b; Yang et al., 2025; Lin et al., 2024; Wang et al., 2024c; Hu et al., 2025), but its effectiveness can depend on optimization and test conditions, and unstable adaptation may lead to performance degradation, negative transfer, or collapse (Zhao et al., 2023; Niu et al., 2022; 2023; Chen et al., 2025a; Sahoo et al., 2025). Repeated retraining and additional data collection are also used to adapt models to changing deployment conditions (Klinghoffer et al., 2023; Li et al., 2025), but can be costly after deployment. These routes address degradation by modifying the model, training process, or deployed state. Post-deployment correction offers another route by retaining the underlying task model while modifying inputs, auxiliary representations, prompts, or predictions (Gao et al., 2023; Zhang et al., 2025). Such correction can recover degraded performance without updating the task model, while still leaving part of the lost performance unrecovered. Recent correction methods are typically evaluated through the task gains produced by their interventions (Agarwal et al., 2025; Zhang et al., 2025; Maleki & Radha, 2026; Zhang et al., 2024). However, limited final recovery leaves the source of that limitation unclear. We examine what corrective evidence supports the tested intervention and how the choice of allocation affects its task gain.

Answering this question requires a controlled reference for what the tested privileged allocation achieves with additional information that is unavailable during deployment. Privileged information (PI) is traditionally defined as information available during training but unavailable at infer-

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.

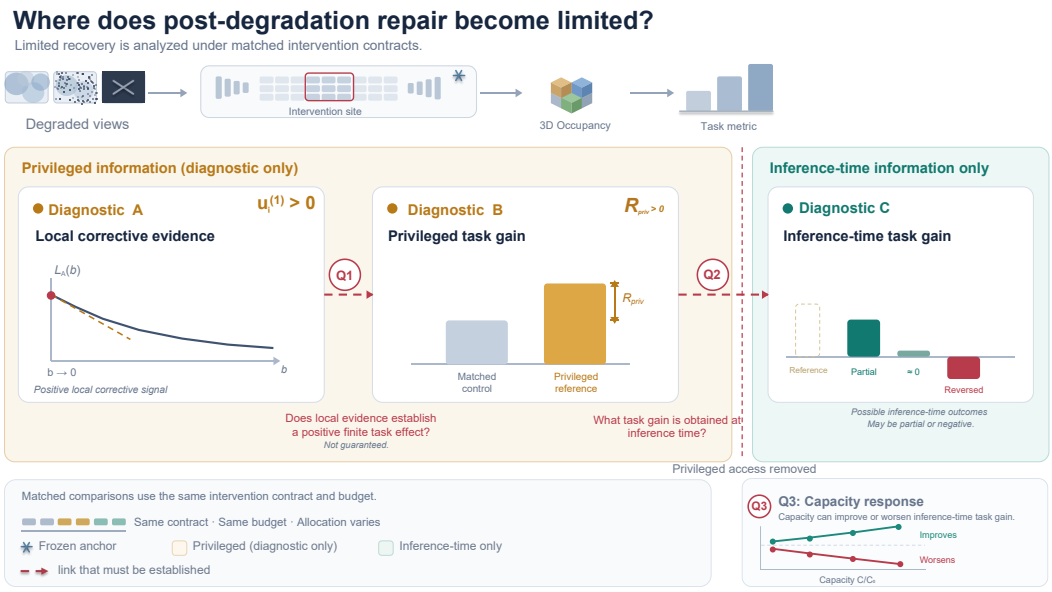


Figure 1: Overview of the diagnostic analysis of limited post-degradation recovery. The three questions share the same diagnostic structure and are evaluated through matched anchor-specific interventions. Q1 tests whether local corrective evidence extends to a positive finite task effect. Q2 compares the task gains of privileged and inference-time allocations against the same matched control. Q3 examines how these task effects change with intervention capacity. Dashed arrows denote relations that are established empirically.

ence (Vapnik & Vashist, 2009). It provides a natural basis for constructing such a reference. However, effective transfer from privileged information is not guaranteed in general (Provodin et al., 2024). PI has been studied in knowledge distillation (Hinton et al., 2015; Lopez-Paz et al., 2016), learning under corrupted supervision (Collier et al., 2022; Ortiz-Jimenez et al., 2023; Wang et al., 2024a), and robust inference (Feldman & Romano, 2025; Feldman et al., 2026). In this study, we use PI to construct a diagnostic reference for corrective intervention. It defines privileged corrective evidence and a reference allocation whose task advantage over the matched control defines the privileged task gain. We then evaluate an inference-time allocation using only information available to the intervention at inference time and measure its task gain against the same control.

For each matched comparison, we keep the intervention site, modification direction, and total intervention strength fixed, and vary only how that strength is allocated across units. This lets us ask three concrete questions: whether the intervention direction is locally corrective, whether privileged allocation improves the finite task metric, and whether an inference-time allocation reproduces that gain. We compare local corrective evidence with the finite task effect, and then compare privileged and inference-time allocations against the same matched control. We examine how these task effects change with intervention capacity. **Q1**: Does local corrective evidence extend to a positive finite task effect? **Q2**: When the privileged allocation produces a positive task gain, what task gain does an inference-time allocation obtain? **Q3**: How do these task effects change with intervention capacity?

Across four camera-based 3D occupancy anchors on nuScenes (Caesar et al., 2020), including FlashOCC (Yu et al., 2023), PanoOcc (Wang et al., 2024d), SparseOcc (Liu et al., 2023), and FB-OCC (Li et al., 2023), along with a controlled CARLA setting Wang et al. (2025), matched interventions reveal distinct relations between corrective evidence and task gain. Local corrective evidence can persist through finite-objective improvement while a positive task advantage over the matched control remains unestablished. Under the same intervention direction and capacity, different inference-time allocations can produce opposite task effects, and this contrast persists as intervention strength increases. In addition, across semantic objectives, privileged task gains remain positive while the estimated effect of a fixed policy can change sign. Our contributions are:

- We use matched exact-capacity interventions to test whether local corrective evidence extends to finite task gain and how privileged and inference-time allocations differ under the same control.
- We show that local corrective evidence can persist through finite-objective improvement while a positive finite task advantage remains unestablished, and that positive privileged task gain can coexist with positive, near-zero, or negative inference-time task gains.
- We characterize how inference-time task gain changes with allocation policy and prescribed intervention capacity, and show that privileged and fixed-policy task effects can vary across semantic occupancy objectives.

2 Controlled Evaluation of Repair Interventions

We study controlled interventions on frozen perception models, which we call *anchors*. Because the anchors expose different internal representations, each matched comparison uses an anchor-specific intervention contract that fixes the intervention site, units, direction construction, capacity definition, and matched control. We evaluate the same questions across these contracts.

2.1 Problem Setting

Each intervention contract exposes a set of units $\mathcal{U}$ in a frozen anchor model. An intervention is specified by coefficients $\mathbf{b} = (b_i)_{i\in\mathcal{U}}$, where b_i scales a fixed additive modification applied to unit i. Within a matched comparison, the intervention site, units, and modification directions are fixed, while only the allocation $\mathbf{b}$ varies. Appendix A specifies the anchor-specific intervention sites and direction constructions.

2.2 Exact-Capacity Intervention Constraint

We define the normalized intervention capacity C as the mean coefficient across intervention units. For a prescribed C, the feasible coefficient vectors (set) satisfy:

$$\Omega(C) = \left\{ \mathbf{b} \in [0, b_{\max}]^{|\mathcal{U}|} \;\middle|\; \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} b_i = C \right\}, \tag{1}$$

where $b_{\max}$ is the per-unit upper bound. Within each matched comparison, the degraded representation, intervention units, realized directions, C, and $b_{\max}$ are fixed, and all allocation policies satisfy the same exact-capacity constraint. Only the distribution of intervention strength across units varies. Because C fixes the mean coefficient exactly, increasing C increases the prescribed intervention strength rather than relaxing an optional budget ceiling.

2.3 Local Corrective Evidence and Finite Effects

Let A index the task objective and let $L_A(\mathbf{b})$ denote its differentiable loss under intervention coefficients $\mathbf{b}$. The first-order corrective utility of unit i is

$$u_i^{(1),A} = -\left.\frac{\partial L_A(\mathbf{b})}{\partial b_i}\right|_{\mathbf{b}=\mathbf{0}}. \tag{2}$$

Positive $u_i^{(1),A}$ indicates locally corrective behavior along the realized direction. Because this quantity describes infinitesimal behavior at zero intervention, we measure finite-objective and finite task effects separately. Here, *finite* denotes execution at a prescribed nonzero intervention strength rather than a computational or resource constraint. For two matched allocations $\mathbf{b}^a, \mathbf{b}^c \in \Omega(C)$, we define

$$G_1^A(\mathbf{b}^a, \mathbf{b}^c) = \left\langle \mathbf{u}^{(1),A}, \mathbf{b}^a - \mathbf{b}^c \right\rangle, \qquad G_L^A(\mathbf{b}^a, \mathbf{b}^c) = L_A(\mathbf{b}^c) - L_A(\mathbf{b}^a). \tag{3}$$

Positive G_1^A indicates that first-order utility favors $\mathbf{b}^a$ over $\mathbf{b}^c$, while positive G_L^A indicates a lower finite objective under $\mathbf{b}^a$. Together with the separately measured task effect, these quantities distinguish local corrective evidence, finite-objective improvement, and finite task-metric change.

Table 1: Intervention contracts used for Q1–Q3. Repeated anchors indicate distinct contracts.

Contract	Direction	Matched comparison
FlashOCC (Q2)	Inference-observable BEV direction	Privileged / inference-time vs. Uniform
FlashOCC (Q3)	Oracle-health-conditioned BEV direction	Privileged / inference-time vs. Uniform
PanoOcc (Q1)	GT-gradient direction	GT-guided vs. Uniform
PanoOcc (Q2/Q3)	Learned voxel direction	Privileged / inference-time vs. Uniform
SparseOcc (Q2)	Fixed-basis token direction	Ranking/profile diagnostic
FB-OCC (Q1)	GT-gradient direction	GT-guided vs. Uniform, with corrected Shuffled diagnostic
CARLA (Q1)	GT-gradient direction	GT-guided vs. Uniform
CARLA (Q2/Q3)	Inference-compatible BEV direction	Privileged / inference-time vs. Uniform

2.4 Privileged and Inference-Time Task Gains

The vector $\mathbf{u}_A^{(1)} = (u_i^{(1),A})_{i\in\mathcal{U}}$ collects the first-order corrective utilities. We construct the privileged allocation $\mathbf{b}_A^{\mathrm{priv}}(C) \in \Omega(C)$ from $[\mathbf{u}_A^{(1)}]_+$. Because the utilities use the task target of the evaluated record, this allocation serves as an in-sample diagnostic reference under the fixed intervention contract. The inference-time allocation $\mathbf{b}^{\mathrm{inf}}(C)$ uses only inference-available information, while the capacity-matched Uniform allocation $\mathbf{b}^{\mathrm{uni}}(C)$ serves as the common control. With $M_A(\mathbf{b})$ denoting the task metric for objective A, we define

$$R^A_{\mathrm{priv}}(C) = M_A\Big(\mathbf{b}_A^{\mathrm{priv}}(C)\Big) - M_A\big(\mathbf{b}^{\mathrm{uni}}(C)\big)\,, \quad R^A_{\mathrm{inf}}(C) = M_A\big(\mathbf{b}^{\mathrm{inf}}(C)\big) - M_A\big(\mathbf{b}^{\mathrm{uni}}(C)\big)\,. \tag{4}$$

R^A_{priv} measures the privileged task gain, while R^A_{inf} measures the inference-time task gain relative to the same capacity-matched control. Because the intervention strength is fixed, these quantities measure the effect of selective allocation rather than absolute recovery from the unintervened degraded anchor. When $R^A_{\mathrm{priv}}(C) > 0$, we also have

$$\mathrm{FRR}^A(C) = \frac{R^A_{\mathrm{inf}}(C)}{R^A_{\mathrm{priv}}(C)}. \tag{5}$$

FRR measures the signed ratio between the inference-time task gain and the privileged task gain under the same intervention setting. We additionally use first-order utility capture (FOUC) to compare the first-order utility captured by the inference-time allocation against the same capacity-matched Uniform and privileged allocations defined above. Under this normalization, the Uniform and privileged allocations correspond to reference values 0 and 1, respectively. FOUC is interpreted separately from finite-objective and finite task effects. When the task objective is fixed, we omit the objective index A from these quantities in subsequent sections.

2.5 Experimental Protocol

We evaluate four camera-based 3D occupancy anchors on nuScenes: FlashOCC, PanoOcc, SparseOcc, and FB-OCC, using Occ3D (Tian et al., 2023) as the primary occupancy annotation, along with a controlled CARLA setting from UniOcc. All task-model parameters remain frozen during intervention analysis. The primary FlashOCC allocation comparison uses 150 validation scenes, while the primary PanoOcc allocation comparison uses 1,280 corrupted records from 70 development scenes. PanoOcc Q1 uses a separate frozen diagnostic of 128 development keyframes evaluated under five corruption conditions. Table 1 summarizes the intervention contracts used for Q1–Q3. All matched comparisons keep the intervention site, direction construction, capacity definition, and control fixed while changing the allocation under comparison. Full inference observability requires all input-dependent components of the intervention to be available at inference time. We refer to such interventions as fully inference-observable.

The capacity-matched Uniform allocation defined in Sec. 2.4 provides the common reference for finite task effects, while Shuffled serves as a contract-specific auxiliary diagnostic. Capacity sweeps

Table 2: Local-to-finite validation under the Q1 GT-gradient contracts. PanoOcc and CARLA report directional numerical validation and finite task effect. FB-OCC additionally retains corrected Shuffled for its first-order allocation diagnostic. G_1 denotes first-order allocation advantage, G_L finite-objective advantage, FD finite difference, JVP Jacobian–vector product, and Corr. Shuf. corrected Shuffled. Uniform is the common finite-effect reference. Confidence intervals are reported where available.

Anchor	Local evidence	Finite reference	Finite obj.	Task effect
PanoOcc	FD PASS	Uniform	–	$+24.139$ pp $[+22.880, +25.274]$
CARLA	FD/JVP PASS	Uniform	–	$+0.5646$ pp
FB-OCC	$G_1 > 0$ (Corr. Shuf.)	Uniform	$G_L > 0$	-0.0122 pp $[-0.0788, +0.0379]$

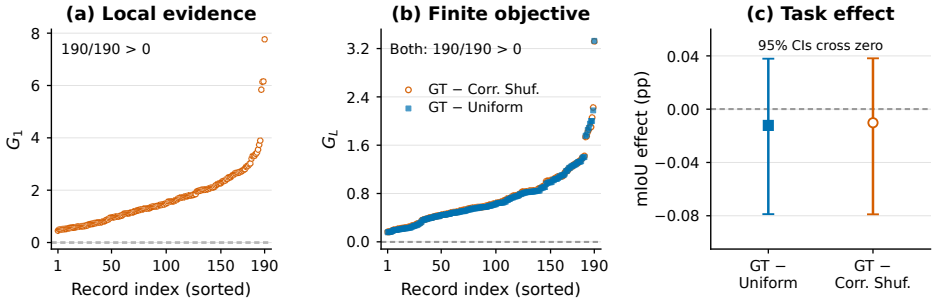


Figure 2: Q1 local-to-finite bridge under the FB-OCC privileged contract. The first-order allocation advantage G_1 is positive on all 190 scene–corruption records, and the finite-objective advantage G_L remains positive against both corrected Shuffled and Uniform. In contrast, the pooled finite task effects against both matched controls have 95% confidence intervals crossing zero. Record-level series in (a) and (b) are sorted independently for visualization.

are normalized by the corresponding C_0 and interpreted within each intervention contract rather than compared directly across anchors. Appendix A provides implementation details, and Appendix B provides evaluation protocols.

3 Results

3.1 Local Corrective Evidence Does Not Establish Positive Finite Task Gain

Q1 asks whether locally validated corrective evidence extends to a positive finite task effect once the intervention is executed at finite strength. FB-OCC evaluates first-order advantage, finite-objective advantage, and finite task effect under the same intervention contract. As shown in Table 2 and visualized in Fig. 2, GT-guided allocation has positive first-order advantage G_1 over corrected Shuffled on all 190 scene–corruption records, with a mean of $+1.6686$. After finite execution, G_L remains positive on all records against corrected Shuffled and Uniform, with means $+0.7274$ and $+0.7171$, respectively. Therefore, the observed separation in FB-OCC is between finite-objective improvement and finite task effect rather than between first-order evidence and finite-objective improvement.

However, the finite-task metric differs despite success on local evidence. Against Uniform, the pooled GT–Uniform mIoU effect is -0.0122 pp with a paired scene-cluster 95% confidence interval of $[-0.0788, +0.0379]$ pp. The GT–corrected-Shuffled effect is -0.0102 pp with confidence interval $[-0.0789, +0.0382]$ pp. Local corrective evidence and finite-objective improvement are positive, while a positive finite task gain is not established. At the prediction level, the GT-guided allocation changes only 1.10% of valid argmax predictions relative to Uniform. Overall voxel correctness increases by 63,923 voxels, while the reported task metric excludes the free class and averages IoU over the 17 evaluated semantic classes. Among these classes, six improve, eight decline, and three remain unchanged.

Table 3: Privileged and inference-time task gains relative to matched Uniform. FlashOCC uses pooled-confusion mIoU differences, while PanoOcc uses mean per-record mIoU differences. Their effects are averaged over three policy seeds, and FRR is the mean seed-wise ratio. SparseOcc reports its fixed-basis diagnostic.

Setting	R_{priv} (pp)	R_{inf} (pp)	FRR
FlashOCC	+0.476	+0.094	0.194
PanoOcc (Learned)	+6.706	-1.265	-0.189
SparseOcc	+1.721	+0.036	0.021
CARLA	+0.065	+0.027	0.410

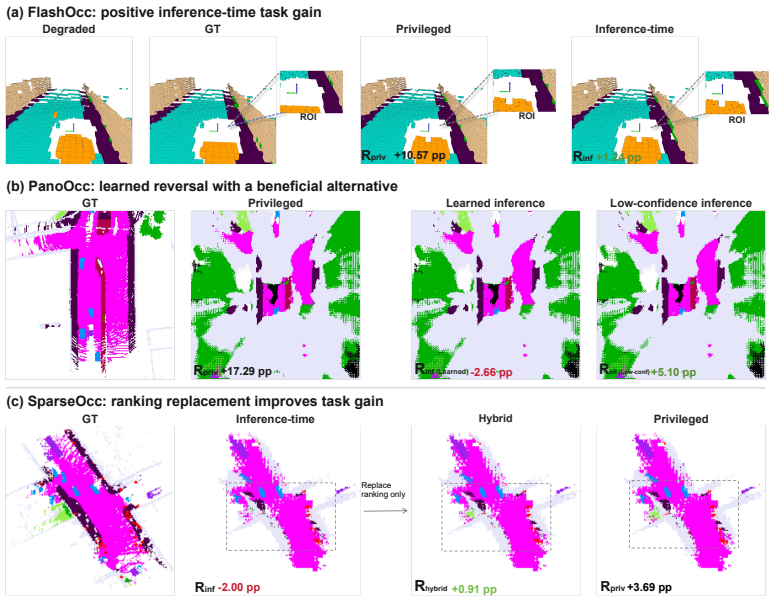


Figure 3: Qualitative diagnostics. Examples are selected from evaluated records for visual clarity. (a) FlashOCC illustrates positive inference-time task gain under its matched intervention. (b) In PanoOcc, the learned allocation yields a negative inference-time task effect, while a low-confidence allocation remains positive under the same intervention setting. (c) In SparseOcc, replacing the inference-time ranking with the privileged ranking improves the task effect while the coefficient profile is fixed. Values are sample-level task-metric differences from matched Uniform.

PanoOcc and CARLA provide contrasting cases. Under their respective privileged GT-gradient contracts, local numerical validation is followed by positive finite task effects of $+24.1392$ pp and $+0.5646$ pp over Uniform. Local corrective evidence can therefore coincide with a positive finite task gain, but does not by itself establish one. In FB-OCC, the separation occurs between finite-objective improvement and finite task effect.

Q1 takeaway. Local corrective evidence and finite-objective improvement can persist without establishing a positive task advantage over the matched finite control.

3.2 Privileged Task Gain Does Not Determine Inference-Time Task Gain

Q2 asks what inference-time task gain is obtained when the privileged allocation produces a positive task gain. Table 3 and Figure 3 summarize the resulting task effects. FlashOCC and CARLA show positive inference-time gains, SparseOcc remains near zero, and PanoOcc reverses sign. Across these contracts, a positive privileged task gain corresponds to different inference-time task gains.

Table 4: PanoOcc spatial allocation and temporal-context diagnostics. Panel (a) reports mIoU under mixed horizontal XY and vertical Z allocations. Panel (b) reports ΔmIoU as History ON minus the corresponding OFF, WRONG, or SHUFFLED condition.

(a) Mixed XY–Z allocation

	Z_{inf}	Z_{priv}
XY_{inf}	54.9463	62.0941
XY_{priv}	57.7218	63.0112

(b) Temporal history

Comparison	ΔmIoU
ON vs. OFF	$+0.608$
ON vs. WRONG	$+3.597$
ON vs. SHUFFLED	$+6.889$

Table 5: SparseOcc ranking/profile diagnostic under the fixed-basis intervention. Priv. and Inf. denote privileged and inference-time components. Loss red. is finite loss reduction. The Inf.-ranking/Priv.-profile configuration is unavailable.

Ranking	Profile	ΔmIoU (pp)	FOUC	Loss red.
Priv.	Priv.	$+1.7205$	1.0000	$+2.63266$
Priv.	Inf.	$+0.5480$	0.2275	$+0.66959$
Inf.	Inf.	$+0.0361$	0.0609	$+0.17971$
Priv. vs. Inf. ranking, fixed Inf. profile		$+0.5119$	$+0.16664$	$+0.48989$

PanoOcc provides the clearest reversal. At C_0, privileged allocation yields $+6.7062$ pp over Uniform, while the learned inference-time allocation yields a three-seed mean effect of -1.2651 pp with a paired scene-cluster 95% confidence interval of $[-1.6698, -0.8181]$ pp. All three seed-level point estimates are negative. Under the same intervention direction and capacity, entropy and low-confidence allocations instead yield positive gains of $+0.1204$ and $+0.1694$ pp. The negative realization therefore depends on the executed allocation policy rather than being determined by the intervention direction alone. Appendix C.2 reports the seed-level and policy-specific results.

To examine the gap between the privileged and learned inference-time task gains in PanoOcc, we next analyze its spatial allocation and temporal context. Table 4 reports two diagnostics. In the spatial diagnostic, starting from the fully inference-time configuration at 54.9463 mIoU, replacing only Z with its privileged allocation raises performance to 62.0941 mIoU and closes 88.6% of the gap to the fully privileged configuration at 63.0112 mIoU. Replacing only XY reaches 57.7218 mIoU. Under the inference-time XY condition, replacing Z produces the larger of the two single-component task effects. These replacements are conditional interventions rather than additive contributions and therefore do not define a unique decomposition of the task-gain gap. Temporal history provides an additional diagnostic of task-relevant context. Under a controlled temporal perturbation setting, correct history outperforms removed, incorrect, and shuffled history by 0.608, 3.597, and 6.889 mIoU, respectively. These results show that temporal context contributes measurably to PanoOcc performance under the evaluated setting.

SparseOcc exposes a different constraint, as shown in Table 5. The fully inference-time ranking/profile combination yields $+0.0361$ pp, compared with $+1.7205$ pp when both components are privileged. Holding the inference-time coefficient profile fixed, replacing only its unit ordering with the privileged ordering raises the finite task gain to $+0.5480$ pp. FOUC increases from 0.0609 to 0.2275, and finite loss reduction from $+0.17971$ to $+0.66959$. Under the fixed inference-time profile, unit ordering contributes to the observed privileged–inference task-gain gap.

Q2 takeaway. A positive privileged task gain can correspond to positive, near-zero, or negative inference-time task gains, depending on the allocation policy and intervention setting.

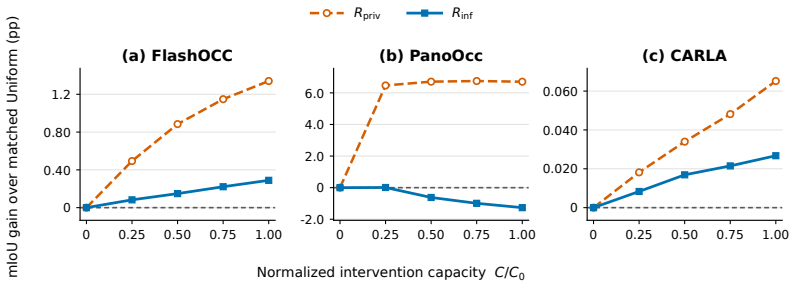


Figure 4: Capacity response under the corresponding frozen Q3 intervention contracts. FlashOCC uses the oracle-health-conditioned direction, PanoOcc the frozen voxel direction, and CARLA the inference-compatible direction. FlashOCC and CARLA show increasing positive inference-time task gains as intervention capacity grows. In PanoOcc, privileged task gain remains positive and approximately stable while the learned inference-time task gain decreases and becomes increasingly negative. R_{priv} and R_{inf} are measured relative to capacity-matched Uniform. Capacities are comparable only within each contract.

3.3 How Does Intervention Capacity Affect Inference-Time Task Gain?

Q3 examines how inference-time task gain changes as the prescribed intervention strength increases. We compare these trajectories within each intervention contract and, where multiple inference-time allocations are available, examine whether their capacity responses remain consistent.

Figure 4 shows that additional capacity has no uniform effect on the inference-time task gain. Under the FlashOCC oracle-health-conditioned contract, R_{priv} increases from $+0.4934$ to $+1.3404$ pp as capacity grows from $0.25C_0$ to C_0, while R_{inf} increases from $+0.0826$ to $+0.2891$ pp. The corresponding FRR rises from 0.1675 to 0.2157. CARLA also shows an increasing positive inference-time task gain, with R_{inf} rising from $+0.0083$ to $+0.0267$ pp. Its FRR, however, is non-monotonic and changes from 0.4586 at $0.25C_0$ to 0.4101 at C_0. CARLA separates first-order utility capture from inference-time task gain: FOUC decreases from 0.7236 to 0.4337 while R_{inf} increases.

PanoOcc exhibits a different capacity response. The privileged task gain remains approximately stable across the sweep, with R_{priv} values of $+6.4692$, $+6.7100$, $+6.7492$, and $+6.7062$ pp from $0.25C_0$ to C_0. Over the same capacities, the learned R_{inf}, averaged across three seeds, changes from $+0.0100$ to -0.6206, -0.9910, and -1.2651 pp, with the same decreasing trend in every seed. Thus, the privileged task gain remains similar while the learned inference-time task gain progressively deteriorates under this intervention setting.

Entropy-based and low-confidence allocations remain positive and outperform the learned allocation at all four tested capacities (Table 22). Their gains increase from $+0.0285$ and $+0.0534$ pp at $0.25C_0$ to $+0.1204$ and $+0.1694$ pp at C_0, respectively. These policies share the intervention direction, eligible units, and exact-capacity constraint. The same direction therefore supports positive task gains at capacities where the learned allocation underperforms Uniform, showing that the capacity response depends on how intervention strength is allocated. A separate single-seed development diagnostic examines the signed first-order utility captured by a matched utility-supervised and Task-only policy pair (Table 27). The utility-supervised policy captures larger positive and negative contributions, while its net captured utility remains positive. This diagnostic characterizes the first-order structure of the allocation without attributing the primary capacity response to a single mechanism.

Q3 takeaway. Increasing prescribed intervention strength can produce different task-gain trajectories under different allocation policies, even within the same intervention contract.

3.4 Sensitivity to Semantic Occupancy Objectives

Previous experiments use Occ3D as the primary semantic occupancy objective. We examine annotation sensitivity through two complementary evaluations: constructing privileged corrective utility

Table 6: Annotation-conditioned task gains (pp) relative to matched Uniform. R_{priv} rows report source-specific privileged diagnostics with annotation-specific task-objective and gradient construction, whereas R_{inf} rows report frozen-policy evaluations without re-inducing privileged gradients.

Anchor	Effect	Evaluation scope	Occ3D	OpenOccupancy	SurroundOcc
FlashOCC	R_{priv}	150 validation scenes	$+0.476$	$+0.157$	$+0.104$
	R_{inf}	150 validation scenes	$+0.094$	$+0.063$	$+0.045$
PanoOcc	R_{priv}	15 frames from 15 scenes	$+5.417$	$+2.152$	$+0.520$
	R_{inf}	6,019 frames from 150 scenes	-0.8577	$+0.2953$	$+0.0371$

FlashOCC uses native 17-class pooled mIoU for Occ3D and common 10-class mean record mIoU for the alternative annotations. PanoOcc fixed-policy evaluation uses pooled common-space mIoU.

separately from each occupancy objective, and evaluating a frozen Occ3D-derived policy across annotation objectives. The former recomputes the annotation-specific task objective and feature gradient, whereas the fixed-policy evaluation changes the annotation/evaluation target without re-inducing privileged gradients or allocations. Table 6 reports their task gains and evaluation scopes. The privileged diagnostics yield positive gains under all three objectives. In the PanoOcc fixed-policy evaluation, the task-gain estimate changes from negative under Occ3D to positive under OpenOccupancy and SurroundOcc. FlashOCC also yields positive effects under each reported evaluation protocol. Together, these experiments establish positive privileged task gains across independently induced objectives and show that the effectiveness of a fixed inference-time policy can change across occupancy annotations.

4 Discussion and Limitations

Interpretation. In summary, the results point to more than one way in which recovery can become limited. Local corrective evidence can persist into finite-objective improvement. The task-metric advantage can still remain unresolved. A positive privileged gain can also coexist with weak or negative inference-time realization. The policy comparisons further show that the same intervention direction can yield opposite task effects under different inference-time allocations. Component replacements narrow this difference to specific aspects of the allocation structure in the evaluated cases. These results help identify which stage of the repair process deserves closer examination.

Implications for future repair methods. These findings suggest evaluating intervention design and inference-time allocation as two linked parts of repair. The privileged comparison provides a task-gain reference for the tested intervention. Comparisons among inference-time policies assess how each policy uses the available signals. Capacity sweeps show how task gain changes as the prescribed intervention becomes stronger and can help guide the choice of strength for each policy.

Limitations. Our evaluation focuses on camera-based 3D occupancy on nuScenes, with CARLA providing a controlled synthetic setting. The reported effects are specific to the evaluated intervention contracts, inference-time policies, and annotation-supported evaluation scopes. Capacity effects are interpreted within each contract because intervention geometry, unit definition, and coefficient semantics differ across anchors. Broader cross-dataset generalization and finer causal attribution are beyond the scope of the present study.

5 Conclusion

This work studies limited post-degradation recovery as a diagnostic problem under a fixed task model. The results show that corrective evidence, privileged task gain, and inference-time task gain capture distinct aspects of repair under matched interventions. For future repair methods, two questions are particularly important: how to preserve positive privileged task gain when moving to inference-time allocation, and how to increase intervention strength without degrading the resulting task effect. More broadly, our results motivate evaluating not only final recovered performance, but also how corrective evidence is translated into finite task gain at inference time.

AI USE STATEMENT

Generative AI tools were used to assist with code editing, manuscript editing, literature search, feedback on research methodology and partial experimental design, interpretation of experimental results, and partial mathematical formulation and derivation. The authors reviewed and verified all AI-assisted code changes, mathematical arguments, experimental interpretations, and manuscript content. The authors take responsibility for the final content of this work, including all reported results, claims, and artifacts.

REPRODUCIBILITY STATEMENT

The main text defines the intervention contracts, local corrective utility, exact-capacity allocation, and evaluation quantities used throughout the study. Appendix A documents the contract-specific task objectives, intervention sites and directions, exact-capacity allocation, inference-time observables, policy training, and capacity construction. Appendix B specifies the evaluation scopes, corruption protocols, statistical procedures, and numerical validity checks. Appendices C and D provide the extended quantitative results and controlled task-effect diagnostics underlying the reported findings.

REFERENCES

Aishwarya Agarwal, Srikrishna Karanam, and Vineet Gandhi. Tide: Training locally interpretable domain generalization models enables test-time correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30210–30220, June 2025.

Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020.

Mingcai Chen, Baoming Zhang, Zongbo Han, Wenyu Jiang, Yanmeng Wang, Shuai Feng, Yuntao Du., and Bingkun Bao. Test-time selective adaptation for uni-modal distribution shift in multi-modal data. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 9711–9727. PMLR, 13–19 Jul 2025a. URL https://proceedings.mlr.press/v267/chen25ch.html.

Zhuoxiao Chen, Junjie Meng, Mahsa Baktashmotlagh, Yonggang Zhang, Zi Huang, and Yadan Luo. Mos: Model synergy for test-time adaptation on lidar-based 3d object detection. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 92784–92808, 2025b. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/e7c3ac813288e4aff606c24e59a8410a-Paper-Conference.pdf.

Mark Collier, Rodolphe Jenatton, Efi Kokiopoulou, and Jesse Berent. Transfer and marginalize: Explaining away label noise with privileged information, 2022. URL https://arxiv.org/abs/2202.09244.

Shai Feldman and Yaniv Romano. Robust conformal prediction using privileged information, 2025. URL https://arxiv.org/abs/2406.05405.

Shai Feldman, Stephen Bates, and Yaniv Romano. Conformal prediction with corrupted labels: Uncertain imputation and robust re-weighting, 2026. URL https://arxiv.org/abs/2505.04733.

Jin Gao, Jialing Zhang, Xihui Liu, Trevor Darrell, Evan Shelhamer, and Dequan Wang. Back to the source: Diffusion-driven adaptation to test-time corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11786–11796, June 2023.

Chongjian Ge, Junsong Chen, Enze Xie, Zhongdao Wang, Lanqing Hong, Huchuan Lu, Zhenguo Li, and Ping Luo. Metabev: Solving sensor failures for bev detection and map segmentation, 2023. URL https://arxiv.org/abs/2304.09801.

Xiaoshuai Hao, Mengchuan Wei, Yifan Yang, Haimei Zhao, Hui Zhang, Yi Zhou, Qiang Wang, Weiming Li, Lingdong Kong, and Jing Zhang. Is your hd map constructor reliable under sensor corruptions? *Advances in Neural Information Processing Systems*, 37:22441–22482, 2024.

Xiaoshuai Hao, Lingdong Kong, Rong Yin, Pengwei Wang, Jing Zhang, Yunfeng Diao, and Shu Zhao. Safemap: Robust hd map construction from incomplete observations. *arXiv preprint arXiv:2507.00861*, 2025.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015. URL https://arxiv.org/abs/1503.02531.

Zixuan Hu, Dongxiao Li, Xinzhu Ma, Shixiang Tang, Xiaotong Li, Wenhan Yang, and Ling-Yu Duan. Adaptive dual uncertainty optimization: Boosting monocular 3d object detection under test-time shifts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7273–7283, October 2025.

Junjie Huang, Guan Huang, Zheng Zhu, Ye Yun, and Dalong Du. Bevdet: High-performance multi-camera 3d object detection in bird-eye-view. *arXiv preprint arXiv:2112.11790*, 2021.

Tzofi Klinghoffer, Jonah Philion, Wenzheng Chen, Or Litany, Zan Gojcic, Jungseock Joo, Ramesh Raskar, Sanja Fidler, and Jose M. Alvarez. Towards viewpoint robustness in bird's eye view segmentation, 2023. URL https://arxiv.org/abs/2309.05192.

Ye Li, Wenzhao Zheng, Xiaonan Huang, and Kurt Keutzer. Unidrive: Towards universal driving perception across camera configurations. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 26250–26264, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/41badd36e935f8a80175e95d8bc6192e-Paper-Conference.pdf.

Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Yu Qiao, and Jifeng Dai. Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. *arXiv preprint arXiv:2203.17270*, 2022.

Zhiqi Li, Zhiding Yu, David Austin, Mingsheng Fang, Shiyi Lan, Jan Kautz, and Jose M Alvarez. FB-OCC: 3D occupancy prediction based on forward-backward view transformation. *arXiv:2307.01492*, 2023.

Hongbin Lin, Yifan Zhang, Shuaicheng Niu, Shuguang Cui, and Zhen Li. MonoTTA: Fully Test-Time Adaptation for Monocular 3D Object Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. doi: 10.1007/978-3-031-72784-9_6. URL https://mlanthology.org/eccv/2024/lin2024eccv-monotta/.

Haisong Liu, Haiguang Wang, Yang Chen, Zetong Yang, Jia Zeng, Li Chen, and Limin Wang. Fully sparse 3d panoptic occupancy prediction. *arXiv preprint arXiv:2312.17118*, 2023.

David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying Distillation and Privileged Information. In *International Conference on Learning Representations*, 2016. URL https://mlanthology.org/iclr/2016/lopezpaz2016iclr-unifying/.

Armin Maleki and Hayder Radha. Heteroprompt: A real-time and privacy-preserving heterogeneous collaborative perception framework, 2026. URL https://arxiv.org/abs/2607.26283.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16888–16905. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/niu22a.html.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world, 2023. URL https://arxiv.org/abs/2302.12400.

Guillermo Ortiz-Jimenez, Mark Collier, Anant Nawalgaria, Alexander Nicholas D'Amour, Jesse Berent, Rodolphe Jenatton, and Efi Kokiopoulou. When does privileged information explain away label noise? In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 26646–26669. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/ortiz-jimenez23a.html.

Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In *Proceedings of the European Conference on Computer Vision*, 2020.

Danil Provodin, Bram van den Akker, Christina Katsimerou, Maurits Kaptein, and Mykola Pechenizkiy. Rethinking knowledge transfer in learning using privileged information, 2024. URL https://arxiv.org/abs/2408.14319.

Sabyasachi Sahoo, Mostafa ElAraby, Jonas Ngnawe, Yann Pequignot, Frederic Precioso, and Christian Gagne. A layer selection approach to test time adaptation, 2025. URL https://arxiv.org/abs/2404.03784.

Xiaoyu Tian, Tao Jiang, Longfei Yun, Yue Wang, Yilun Wang, and Hang Zhao. Occ3d: A large-scale 3d occupancy prediction benchmark for autonomous driving. *arXiv preprint arXiv:2304.14365*, 2023.

Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5):544–557, 2009. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2009.06.042. URL https://www.sciencedirect.com/science/article/pii/S0893608009001130. Advances in Neural Networks Research: IJCNN2009.

Ke Wang, Guillermo Ortiz-Jimenez, Rodolphe Jenatton, Mark Collier, Efi Kokiopoulou, and Pascal Frossard. Pi-DUAL: Using privileged information to distinguish clean from noisy labels. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 51214–51236. PMLR, 21–27 Jul 2024a. URL https://proceedings.mlr.press/v235/wang24bb.html.

Shiming Wang, Holger Caesar, Liangliang Nan, and Julian F. P. Kooij. Unibev: Multi-modal 3d object detection with uniform bev encoders for robustness against missing sensor modalities, 2024b. URL https://arxiv.org/abs/2309.14516.

Yanshuo Wang, Ali Cheraghian, Zeeshan Hayder, Jie Hong, Sameera Ramasinghe, Shafin Rahman, David Ahmedt-Aristizabal, Xuesong Li, Lars Petersson, and Mehrtash Harandi. Backpropagation-free network for 3d test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23231–23241, June 2024c.

Yuping Wang, Xiangyu Huang, Xiaokang Sun, Mingxuan Yan, Shuo Xing, Zhengzhong Tu, and Jiachen Li. Uniocc: A unified benchmark for occupancy forecasting and prediction in autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2025.

Yuqi Wang, Yuntao Chen, Xingyu Liao, Lue Fan, and Zhaoxiang Zhang. Panoocc: Unified occupancy representation for camera-based 3d panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17158–17168, 2024d.

Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. RoboBEV: Towards robust bird's eye view perception under corruptions. *arXiv preprint arXiv:2304.06719*, 2023.

Huitong Yang, Zhuoxiao Chen, Fengyi Zhang, Zi Huang, and Yadan Luo. Codemerge: Codebook-guided model merging for robust test-time adaptation in autonomous driving, 2025.

Zichen Yu, Changyong Shu, Jiajun Deng, Kangjie Lu, Zongdai Liu, Jiangyong Yu, Dawei Yang, Hui Li, and Yan Chen. Flashocc: Fast and memory-efficient occupancy prediction via channel-to-height plugin. 2023.

Hanxue Zhang, Zetong Yang, Yanan Sun, Li Chen, Fei Xia, Fatma Güney, and Hongyang Li. Test-time correction: An online 3d detection system via visual prompting, 2025. URL https://arxiv.org/abs/2412.07768.

Yunbei Zhang, Akshay Mehra, and Jihun Hamm. Ot-vp: Optimal transport-guided visual prompting for test-time adaptation, 2024. URL https://arxiv.org/abs/2407.09498.

Hao Zhao, Yuejiang Liu, Alexandre Alahi, and Tao Lin. On pitfalls of test-time adaptation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 42058–42080. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/zhao23d.html.

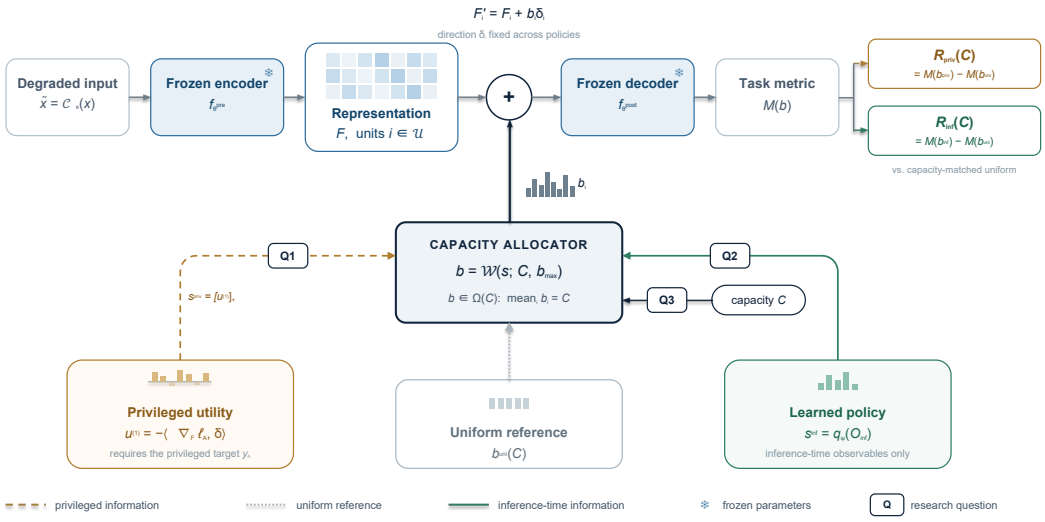


Figure 5: Controlled finite-intervention evaluation framework. A degraded input is mapped by the frozen encoder to intervention units F_i. Within each comparison, the realized direction $\boldsymbol{\delta}_i$ is fixed, while the capacity allocator maps privileged scores $s_{\text{priv}} = [u_i^{(1),A}]_+$ or learned inference-time scores $s_{\text{inf}} = q_\psi(\mathcal{O}_{\text{inf}})$ to intervention coefficients $\mathbf{b}$ at prescribed capacity C. The resulting intervention $F_i' = F_i + b_i\boldsymbol{\delta}_i$ is evaluated through the frozen decoder and task metric $M(\mathbf{b})$. R_{priv} measures privileged task gain relative to capacity-matched Uniform, while R_{inf} measures inference-time task gain relative to the same reference. Q1 tests whether local corrective evidence extends to a positive finite task effect. Q2 compares the privileged and inference-time task gains under the same matched control. Q3 examines how these task effects change with intervention capacity.

A Finite-Intervention Implementation Details

This appendix provides implementation details for the controlled finite-intervention evaluation in Sec. 2. It specifies the anchor-specific intervention sites and directions, contract-specific capacity definitions, exact-capacity allocation, and the construction of privileged, Uniform, and inference-time allocations. The inference-time policy implementations and observability of the corresponding intervention components are also documented. Throughout this appendix, A indexes the task objective, while a indexes an intervention contract.

A.1 Intervention Overview and Local-to-Finite Relation

Throughout this paper, *finite intervention* refers to executing a nonzero intervention at prescribed strength, in contrast to first-order analysis at zero intervention. Here, *finite* denotes non-infinitesimal execution rather than computational or resource constraints.

Figure 5 summarizes the controlled finite-intervention pipeline. The degraded input is mapped by the frozen encoder to an anchor-specific representation with intervention units F_i. Each unit is modified as $F_i' = F_i + b_i\boldsymbol{\delta}_i$, where the realized direction $\boldsymbol{\delta}_i$ is fixed within each allocation comparison and b_i specifies the intervention strength assigned to that unit. The capacity allocator $\mathcal{W}$ maps privileged or inference-time allocation scores to coefficients under the prescribed capacity C, with Uniform providing the capacity-matched equal-allocation reference. The modified representation is evaluated by task metric $M(\mathbf{b})$. Privileged task gain and inference-time task gain are defined by comparing the corresponding allocations with this common Uniform reference.

The distinction between local corrective evidence and finite intervention used in Q1 can be made explicit through the local behavior of the task objective. For task objective A, let

$$\mathcal{L}_A(\mathbf{b}) = \ell_A(\hat{y}(\mathbf{b}), y_A)\,, \qquad u_i^{(1),A} = -\left.\frac{\partial \mathcal{L}_A(\mathbf{b})}{\partial b_i}\right|_{\mathbf{b}=\mathbf{0}}, \tag{6}$$

and let $\mathbf{u}_A^{(1)} = (u_i^{(1),A})_{i\in\mathcal{U}}$ collect the unit-wise first-order corrective utilities. The sign convention makes $u_i^{(1),A} > 0$ correspond to a locally loss-reducing increase in the intervention coefficient b_i.

Around zero intervention, the objective at a finite coefficient vector $\mathbf{b}$ can be written as

$$\mathcal{L}_A(\mathbf{b}) = \mathcal{L}_A(\mathbf{0}) - \left\langle \mathbf{u}_A^{(1)}, \mathbf{b} \right\rangle + r_A(\mathbf{b}), \tag{7}$$

where the remainder is

$$r_A(\mathbf{b}) = \mathcal{L}_A(\mathbf{b}) - \mathcal{L}_A(\mathbf{0}) + \left\langle \mathbf{u}_A^{(1)}, \mathbf{b} \right\rangle. \tag{8}$$

The inner-product term is the first-order prediction of the objective change obtained from the local utilities at $\mathbf{b} = \mathbf{0}$, while $r_A(\mathbf{b})$ collects the discrepancy that appears once the intervention is executed at finite strength. If $\mathcal{L}_A$ is twice differentiable near $\mathbf{0}$ with locally bounded curvature, then $r_A(\mathbf{b}) = O(\|\mathbf{b}\|^2)$ as $\|\mathbf{b}\| \to 0$. Thus, first-order utility characterizes the infinitesimal preference over intervention directions, whereas the executed finite intervention can additionally reflect higher-order interactions among the modified units. This distinction motivates evaluating first-order allocation advantage, finite-objective advantage, and finite task effect separately in Q1.

The same first-order utility can be interpreted directly in feature space. For unit i, the coefficient b_i moves the representation along the realized direction δ_i. Thus, we have

$$\begin{aligned} \mathbf{F}_i'(b_i) &= \mathbf{F}_i + b_i \boldsymbol{\delta}_i, \\ \left.\frac{\partial \mathcal{L}_A}{\partial b_i}\right|_{\mathbf{b}=\mathbf{0}} &= \langle \nabla_{\mathbf{F}_i} \ell_A(\hat{\mathbf{y}}_0, y_A), \boldsymbol{\delta}_i \rangle , \\ u_i^{(1),A} &= -\langle \nabla_{\mathbf{F}_i} \ell_A(\hat{\mathbf{y}}_0, y_A), \boldsymbol{\delta}_i \rangle . \end{aligned} \tag{9}$$

The second line follows from the chain rule, since b_i changes $\mathbf{F}_i$ only along $\boldsymbol{\delta}_i$. For a sufficiently small positive intervention ϵ applied only to unit i,

$$\mathcal{L}_A(\epsilon \mathbf{e}_i) - \mathcal{L}_A(\mathbf{0}) = -\epsilon u_i^{(1),A} + O(\epsilon^2). \tag{10}$$

Thus, $u_i^{(1),A} > 0$ means that a small positive intervention along $\boldsymbol{\delta}_i$ is locally predicted to reduce the task objective, while its magnitude gives the corresponding first-order loss-reduction rate. The numerical validity of this gradient–direction form is examined in Appendix B.5.

Contract-specific task objectives. The differentiable task objective $\mathcal{L}_A$ is instantiated by the frozen task head of each intervention contract rather than by the allocation-policy training loss. Table 7 summarizes the objectives used to construct first-order corrective utility and, where retained, finite-objective effects.

Finite-objective effects are reported only for contracts that retain post-intervention task loss. PanoOcc Q1, FB-OCC Q1, and CARLA evaluate the same task objective used to construct their local corrective utility.

The differentiable objectives and reported task metrics use different aggregations. PanoOcc Q1 includes the free class in cross-entropy but excludes it from non-free mIoU. FB-OCC uses class- and component-weighted differentiable losses. CARLA uses weighted cross-entropy while reporting non-free IoU. These distinctions motivate evaluating finite-objective and task-metric effects separately.

Common annotation space. The source-specific annotation diagnostic uses a locked common 11-class probability space. Native occupancy probabilities are projected differentiably according to

$$\begin{array}{lllll} 1 \leftarrow \{3,4,5,9,10\}, & 2 \leftarrow \{2\}, & 3 \leftarrow \{6\}, & 4 \leftarrow \{7\}, & 5 \leftarrow \{8\}, \\ 6 \leftarrow \{16\}, & 7 \leftarrow \{11\}, & 8 \leftarrow \{12,13,14\}, & 9 \leftarrow \{1,15\}, & 10 \leftarrow \{17\}, \end{array}$$

with common class $0 \leftarrow \{0\}$ ignored. The valid support is the intersection of the three source-valid masks and requires each source label to map to common classes 1–10. The common-space NLL and its feature gradient are recomputed separately for each annotation source.

Table 7: Differentiable task objectives used for first-order corrective utility and finite-objective analysis. Main FlashOCC and three-seed PanoOcc allocation evaluations retain task metrics rather than finite-loss statistics, so G_L is reported only where it was directly evaluated.

Contract	Differentiable objective $\mathcal{L}_A$	Valid support / class treatment	Finite-objective status
FlashOCC (Q2)	Unweighted 18-class occupancy cross-entropy (`loss_occ`)	Camera-valid voxels, free class 17 included, label 255 ignored, normalized by valid camera-mask count	Not retained in the primary allocation evaluation
PanoOcc (Q1)	18-class occupancy cross-entropy (`loss_occ`) with coefficient 12	Camera-valid voxels, free class 17 included, normalized by valid camera-mask count	Measured with the same `loss_occ` after finite intervention
PanoOcc (Q2/Q3)	18-class occupancy cross-entropy (`loss_occ`) with coefficient 12	Same native occupancy target and camera-valid support	Finite loss is available in separate direction diagnostics, the primary three-seed allocation evaluation retains task metrics only
FB-OCC (Q1)	Sum of weighted focal occupancy loss, semantic scaling, geometric scaling, and Lovasz loss	Raw 19-class target, label 255 ignored, class- and component-specific weighting	Measured with the same four frozen head components
SparseOcc (fixed basis)	Sum of frozen head loss-key means under fixed Hungarian assignment	Sparse semantic and instance/mask branches, native free class 17, semantic label 255 ignored	Measured under the frozen fixed-basis replay
CARLA	11-class weighted camera-masked cross-entropy	Camera-valid support, free class 10 has positive weight, unsupported classes have zero weight	Measured with the same weighted cross-entropy
Annotation-induced privileged diagnostic	Mean NLL after differentiable projection from the native 18-class prediction to the common annotation space	Intersection of source-valid masks and labels mapped to common classes 1–10, common free class 10 included, common class 0 ignored	Measured with the same source-specific common-space NLL

A.2 Anchor-Specific Intervention Sites and Units

The intervention is applied to an internal representation of each frozen anchor rather than directly to the input images or final semantic predictions. Because the evaluated anchors expose different internal representations, their intervention sites and unit definitions are architecture-specific. Table 8 summarizes the representations used by the evaluated contracts.

For FlashOCC, intervention coefficients are assigned to BEV locations of the degraded internal feature, while the feature-channel direction at each location is provided by the corresponding direction mechanism. PanoOcc exposes a dense 3D voxel representation, so intervention is defined over in-

Table 8: Anchor-specific intervention representations and units. SparseOcc uses a fixed-basis contract for the ranking/profile diagnostic and a separate query-space contract for exploratory diagnostics.

Anchor / contract	Intervention representation	Unit
FlashOCC	Dense BEV feature $F \in \mathbb{R}^{H \times W \times D}$	BEV location (x, y)
PanoOcc	Dense 3D feature $[B, 32, 16, 200, 200]$	Voxel
SparseOcc (fixed basis)	Frozen fixed-basis sparse representation	Fixed-basis sparse unit
SparseOcc (query space)	Original sparse-query representation	Query
FB-OCC	Frozen encoder input $[B, C, Z, H, W]$	BEV column
CARLA	Frozen dense 2D BEV latent $[B, 64, 200, 200]$	BEV cell (x, y)

dividual voxels of the full spatial grid. FB-OCC operates on BEV columns of the frozen encoder input. The two CARLA direction contracts share the same BEV intervention site and unit definition, while their direction constructions differ as described in Sec. A.3.

SparseOcc requires separate treatment because its representation is organized as sparse queries rather than a dense spatial grid. We therefore use two intervention contracts. The fixed-basis contract provides a common intervention geometry for the main realization analysis and enables the ranking/profile diagnostic by separating the ordering of intervention units from the coefficient profile assigned over that ordering. The query-space contract operates directly on the original sparse queries and is used only for exploratory diagnostics. The corresponding ranking/profile results are reported in Table 5. The fixed-basis diagnostic evaluates unit ordering while holding the inference-time coefficient profile fixed.

A.3 Realized Intervention Directions

The first-order utility in Eq. 9 is defined with respect to the realized intervention direction $\boldsymbol{\delta}_i$. This section specifies how that direction is instantiated for each intervention contract. Within each matched allocation comparison, the realized direction is fixed and only the allocation coefficients vary. For contracts with a frozen learned direction, the privileged task gradient evaluates the local utility of that realized direction without redefining it. In the privileged-direction diagnostics of PanoOcc, FB-OCC, and CARLA, the direction itself is constructed from the normalized negative task gradient and then held fixed across the corresponding matched allocation controls.

We use d_i for the unit-normalized direction produced by the direction mechanism and $\boldsymbol{\delta}_i$ for the realized intervention vector used by the finite executor. For the frozen learned-direction contracts,

$$\boldsymbol{\delta}_i = \|F_i\|_2 d_i, \qquad F_i' = F_i + b_i \boldsymbol{\delta}_i. \tag{11}$$

The feature norm therefore sets the local perturbation scale, while d_i determines its direction and b_i controls the allocated intervention strength.

Table 9 summarizes the direction construction used by each contract. FlashOCC uses separate directions for Q2 and Q3. The Q2 contract uses a fully inference-observable BEV direction, whereas the Q3 capacity analysis uses an oracle-health-conditioned BEV direction. The oracle-health signal conditions the frozen Q3 direction construction and is not an input to the inference-time allocation policy. The Q3 experiment therefore evaluates allocation and capacity response under this privileged direction contract rather than full inference observability of the complete intervention. These are treated as separate intervention contracts. In both cases, direction construction is distinct from the coefficient-prediction policy described in Appendix A.6.

For the frozen learned-direction contracts, PanoOcc maps the detached 32-channel voxel feature and normalized (x, y, z) coordinates through two Conv3D–GroupNorm–SiLU blocks and a 32-channel output head, followed by per-voxel channel normalization. The separate PanoOcc Q1 contract constructs its privileged direction from the negative task gradient at each voxel and normalizes it over feature channels using $\epsilon = 10^{-12}$. The finite executor applies the corresponding feature-norm scale to the normalized direction. The learned voxel adapter used in Q2/Q3 instead uses $\epsilon = 10^{-8}$ and sets invalid directions to zero.

Table 9: Realized intervention directions under the anchor-specific contracts. Distinct FlashOCC rows separate the fully inference-observable Q2 contract from the oracle-health-conditioned Q3 capacity contract. SparseOcc query space is used only for exploratory diagnostics.

Contract	Direction construction	Role
FlashOCC (Q2)	Frozen fully inference-observable BEV direction	Inference-time realization
FlashOCC (Q3)	Frozen oracle-health-conditioned BEV direction	Capacity response
PanoOcc (Q1)	Privileged GT-gradient direction	Local-to-finite diagnostic
PanoOcc (Q2/Q3)	Frozen feature–coordinate voxel adapter	Matched allocation and capacity response
SparseOcc (fixed basis)	Frozen feature–coordinate token adapter	Ranking/profile realization diagnostic
SparseOcc (query space)	Fixed-basis direction restricted to query support	Exploratory diagnostic
FB-OCC	Privileged GT-gradient direction over BEV columns	Local-to-finite diagnostic
CARLA (privileged)	Privileged GT-gradient direction	Direction diagnostic
CARLA (inference-compatible)	Frozen inference-observable BEV adapter	Allocation realization

Table 10: Reference-capacity parameters for the anchor-specific intervention contracts. Shared numerical values of C_0 and $b_{\max}$ do not define a common intervention scale across contracts.

Contract	$C_0^{(a)}$	$b_{\max}^{(a)}$	Capacity evaluation
FlashOCC (Q2, inference-observable)	0.004786	0.10	Reference capacity
FlashOCC (Q3, oracle-health-conditioned)	0.004786	0.10	Four-capacity sweep
PanoOcc	0.004786	0.10	Four-capacity sweep
SparseOcc (fixed basis)	0.004786	0.10	Reference-capacity diagnostic
SparseOcc (query space)	0.004786	–	Separate norm-weighted capacity
FB-OCC	0.004786	0.10	Reference-capacity diagnostic
CARLA	0.004786	0.10	Four-capacity sweep

SparseOcc fixed basis applies a token MLP with dimensions $259 \to 128 \to 128 \to 256$ to the detached token feature and normalized coordinates, followed by per-token ℓ_2 normalization. The feature norm in $\boldsymbol{\delta}_i = \|F_i\|_2 d_i$ is taken over the channel dimension of each intervention unit, with b_i broadcast over the corresponding feature channels. The SparseOcc query-space contract restricts the fixed-basis direction to the original query support and is used only for exploratory diagnostics.

FB-OCC constructs d_i directly from the normalized negative task gradient over each BEV-column unit and uses this privileged direction for the local-to-finite diagnostic. CARLA uses a privileged GT-gradient direction for its direction diagnostic and a separate frozen inference-compatible BEV direction for allocation realization.

A.4 Intervention Strength and Contract-Specific Capacity

Intervention strength is defined within each contract rather than through a universal perturbation scale. We denote the reference capacity and coefficient bound for contract a by $C_0^{(a)}$ and $b_{\max}^{(a)}$, respectively. In the evaluated contracts, $C_0^{(a)}$ takes the common numerical value 0.004786, inherited from the dimensionless residual-to-anchor-norm scaling used in the FlashOCC intervention design. Because the intervention representation, unit definition, and realized direction are contract-specific, the same numerical capacity does not correspond to the same physical feature perturbation across anchors. Table 10 summarizes these quantities.

For the capacity-response experiments, intervention strength is expressed relative to the corresponding contract-specific reference $C_0^{(a)}$. The FlashOCC contract, PanoOcc contract, and CARLA contract are evaluated at 0.25, 0.50, 0.75, and $1.00\, C_0^{(a)}$. The fully inference-observable FlashOCC Q2 contract, SparseOcc fixed-basis contract, and FB-OCC contract are evaluated only at their reference

Table 11: Contract-specific implementation of the coefficient-mean exact-capacity allocator, including score preprocessing, saturation, residual completion, and tie handling.

Component	FlashOCC	PanoOcc	SparseOcc	CARLA
Score preprocessing	$s = \mathrm{softplus}(z/0.5)$, $w = s \odot m$. If no positive eligible weight remains, the allocator falls back to m, and runtime weights are floored at $\epsilon = 10^{-8}$.	$s = \mathrm{softplus}(z)$, $w = s \odot m$ in FP64. Zero-score eligible units remain available for residual completion.	$s = \mathrm{softplus}(z)$, $w = s \odot m$. If no positive eligible weight remains, the allocator falls back to m, followed by a valid-token floor of $\epsilon = 10^{-12}$.	$s = \mathrm{softplus}(z/0.5)$, $w = s \odot m$ in FP64. If no positive eligible weight remains, the allocator falls back to m, followed by a floor of $\epsilon = 10^{-12}$.
Bisection and saturation	128 iterations followed by analytic free-mass scaling.	64 iterations followed by analytic free-mass scaling.	48 iterations followed by analytic free-mass scaling.	64 FP64 iterations followed by analytic free-mass scaling.
Residual completion	The score-weighted solution is final.	If $C_{\mathrm{f}} - C_{\mathrm{stage1}} > 10^{-6}$, remaining eligible zero-score capacity is filled uniformly, followed by analytic level recomputation and post-cast correction.	The score-weighted solution is final.	Runtime-dtype undershoot is filled uniformly over remaining eligible caps, while overshoot is corrected by rescaling the complete allocation.
Tie rule	Elementwise water filling.	Elementwise water filling.	Elementwise water filling. The ranking–profile replay additionally uses stable descending token order.	Elementwise water filling.

capacities. SparseOcc query space uses a separate norm-weighted capacity definition without a per-query coefficient upper bound and is evaluated independently from the coefficient-mean capacity sweeps.

A.5 Exact-Capacity Allocation Implementation

For intervention contract a, let B_i denote the local coefficient cap of intervention unit i, with $B_i = 0$ for ineligible units and $B_i = b_{\max}^{(a)}$ otherwise. Given a nonnegative score vector $\mathbf{s}$, the exact-capacity allocator produces

$$\mathbf{b} = \mathcal{W}_a\left(\mathbf{s}; C, b_{\max}^{(a)}\right), \qquad 0 \le b_i \le B_i, \qquad \frac{1}{N}\sum_{i=1}^{N} b_i = C. \tag{12}$$

All matched allocations within a contract use the same eligible support, prescribed capacity, coefficient bounds, and solver implementation.

The privileged and Uniform references differ only in the score vectors supplied to the allocator:

$$\mathbf{b}^{\mathrm{priv}}(C) = \mathcal{W}_a\left([\mathbf{u}^{(1)}]_+; C, b_{\max}^{(a)}\right), \qquad \mathbf{b}^{\mathrm{uni}}(C) = \mathcal{W}_a\left(\mathbf{1}_{\mathcal{E}_a}; C, b_{\max}^{(a)}\right), \tag{13}$$

where $\mathcal{E}_a$ denotes the eligible intervention units of contract a. Inference-time and diagnostic allocations use the same allocator with their corresponding score vectors.

For a contract with N intervention units, the solver flattens the intervention-unit axis, applies the eligibility mask, and converts the supplied scores into nonnegative weights w_i. Given the local

coefficient caps B_i, the allocation follows the water-filling form

$$b_i = B_i \min\{1, \alpha w_i\}, \tag{14}$$

where the scale α is chosen so that the realized mean coefficient matches the prescribed capacity. For numerical feasibility, the implementation defines

$$C_{\mathrm{f}} = \min\left(C, \frac{1}{N}\sum_i B_i\right). \tag{15}$$

When the prescribed capacity is feasible under the contract-specific coefficient caps, $C_{\mathrm{f}} = C$. The clipping therefore acts only as a solver-level feasibility safeguard. The scale α is first estimated by fixed-iteration bisection. Given the resulting saturated set

$$\mathcal{S} = \{i : \alpha_{\mathrm{hi}} w_i \geq 1\}, \tag{16}$$

the remaining free mass is assigned using

$$\alpha = \max\left(0, \frac{NC_{\mathrm{f}} - \sum_{i\in\mathcal{S}} B_i}{\max\left(\sum_{i\notin\mathcal{S}} B_i w_i, \epsilon\right)}\right). \tag{17}$$

When the free-mass denominator vanishes, the bisection upper bound is retained. Equal weights and equal caps follow the same elementwise water-filling rule without top-k selection or a spatial-index tie breaker. SparseOcc uses stable descending token order only in the ranking–profile replay.

Table 11 summarizes the contract-specific numerical implementations. They differ in score transformation, numerical floor, iteration count, arithmetic precision, and residual completion while implementing the same coefficient-mean capacity constraint. PanoOcc, SparseOcc, and CARLA determine the minimum safe weight over eligible units when constructing the initial bisection bracket, while FlashOCC first floors the flattened masked weight tensor. FB-OCC uses the PanoOcc-style FP64 allocator for its local-to-finite diagnostic and is therefore omitted as a separate table column.

Shuffled uses the same contract-specific allocation procedure and prescribed capacity as its matched reference while reassigning the allocation signal across intervention units. FlashOCC and PanoOcc permute structured scores before exact-capacity allocation. FB-OCC permutes privileged utility before recomputing the allocation. The SparseOcc ranking–profile diagnostic reassigns the coefficient profile under a changed token ordering. CARLA permutes the signed teacher utility before allocation. In each case, the contract-specific allocation structure and prescribed capacity are preserved while unit-wise correspondence is changed.

A.6 Inference-Time Allocation Instantiation

The inference-time allocator separates score prediction from capacity enforcement. Each anchor uses a frozen scoring policy q_ψ operating on inference-time observables $\mathcal{O}_{\mathrm{inf}}$. The predicted unit-wise scores are passed through the same contract-specific exact-capacity allocator used by the matched privileged and Uniform references:

$$\mathbf{s}^{\mathrm{inf}} = q_\psi(\mathcal{O}_{\mathrm{inf}}), \qquad \mathbf{b}^{\mathrm{inf}}(C) = \mathcal{W}_a\left(\mathbf{s}^{\mathrm{inf}}; C, b^{(a)}_{\max}\right). \tag{18}$$

Evaluation-time observables are detached quantities computed from the degraded inference stream and are summarized for each contract in Table 12. Privileged supervision may be used during policy training, while evaluation is restricted to $\mathcal{O}_{\mathrm{inf}}$.

The semantic statistics are derived from the corrupted logits. Let $p = \mathrm{softmax}(\ell^{\mathrm{corr}})$ along the class dimension. We compute

$$\begin{aligned} H(p) &= -\sum_c p_c \log \max(p_c, 10^{-8}), \quad P_{\max}(p) = \max_c p_c, \\ P_{\mathrm{fg}}(p) &= 1 - p_{\mathrm{free}}, \qquad\qquad M(p) = p_{(1)} - p_{(2)}. \end{aligned} \tag{19}$$

where $H(p)$ is predictive entropy, $P_{\max}(p)$ is the maximum class probability, $P_{\mathrm{fg}}(p)$ is the non-free probability used as the foreground summary, and $M(p)$ is the top-two probability margin, with $p_{(1)}$ and $p_{(2)}$ denoting the largest and second-largest class probabilities. FlashOCC and CARLA

Table 12: Inference-time observables and frozen coefficient-prediction policies used by the anchor-specific intervention contracts.

Contract	Inference-time observables	Policy construction
FlashOCC	$\mathcal{O}_{\text{inf}}^{\text{Flash}} = \{f_{256\times H\times W}^{\text{corr}}, \ell^{\text{corr}}, (x, y)\}$, with $(x, y) \in [-1, 1]^2$.	The policy receives the corrupted BEV feature, semantic $(\text{entropy}, \text{confidence}, \text{foreground})$ summaries, and normalized (x, y). Its five-channel health branch is fixed to structural zeros under the `semantic_only` configuration.
PanoOcc	$\mathcal{O}_{\text{inf}}^{\text{Pano}} = \{f_{32\times Z\times H\times W}^{\text{corr}}, \ell^{\text{corr}}, (x, y, z)\}$, with $(x, y, z) \in [-1, 1]^3$.	The policy receives the corrupted voxel feature, per-voxel semantic $(\text{entropy}, \text{confidence}, \text{foreground})$ summaries, and normalized (x, y, z).
SparseOcc fixed basis	$\mathcal{O}_{\text{inf}}^{\text{Sparse}} = \{\ell_{N\times K}^{\text{corr}}, (x, y, z), \mathcal{N}_6\}$, where $\mathcal{N}_6$ denotes the fixed six-neighbor topology.	The policy receives four semantic statistics, normalized (x, y, z), and deterministic six-neighbor mean context. The fixed-basis diagnostic replays the corresponding scores and coefficient profiles.
CARLA	$\mathcal{O}_{\text{inf}}^{\text{CARLA}} = \{f_{64\times H\times W}^{\text{corr}}, \ell^{\text{corr}}\}$, with $\ell^{\text{corr}} \in \mathbb{R}^{C\times X\times Y\times Z}$.	The policy receives the corrupted BEV latent and latent-depth mean semantic summaries. Its five-channel observability branch is fixed to structural zeros.

Table 13: Coefficient-prediction policies and recorded training iterations. Parameter counts exclude perception anchors and direction modules. The FlashOCC row refers to the primary inference-observable allocation policies.

Contract	Parameters	Training iterations	Objective
FlashOCC	88,161	75,954 per seed	KL + 0.25 ranking + 4 regret
PanoOcc	58,753	6,144	KL + 0.25 ranking + 4 regret
SparseOcc	57,529	–	Forward allocation KL
CARLA	75,873	2,400	KL + 0.25 ranking + 4 regret

use latent-depth means of $\big(H(p), P_{\max}(p), P_{\text{fg}}(p)\big)$, with free classes 17 and 10, respectively. PanoOcc uses the same three statistics directly at each voxel with free class 17. SparseOcc uses $\big(H(p), P_{\max}(p), P_{\text{fg}}(p), M(p)\big)$, also with free class 17. For FlashOCC, PanoOcc, and CARLA, the coefficient-prediction policy is trained with

$$\mathcal{L}_{\text{policy}} = D_{\text{KL}}(\pi^T \| \pi^S) + 0.25\,\mathcal{L}_{\text{rank}} + 4\,\mathcal{L}_{\text{regret}}, \tag{20}$$

where π^T and π^S are the teacher and student allocation distributions normalized over the common eligible support. The teacher allocation and privileged signed utility used for supervision are detached. The ranking term compares the mean Student score over the top quartile of positive-utility units with the mean score over non-positive units using a softplus margin. The regret term measures the shortfall in captured privileged utility, defined as one minus the utility captured by the Student allocation relative to the detached teacher utility, with numerical ϵ protection. The three terms therefore supervise allocation-distribution matching, utility-informed score ordering, and captured corrective utility. SparseOcc uses forward allocation KL only under its fixed-basis contract.

Table 13 summarizes the frozen coefficient-prediction policies and their retained training settings. These policies are lightweight anchor-specific predictors and are separate from the direction modules in Table 9. FlashOCC combines BEV-feature, semantic, coordinate, and spatial-context branches, PanoOcc uses voxel-feature, semantic, coordinate, and local 3D-context branches, SparseOcc uses a semantic-spatial predictor with deterministic six-neighbor context, and CARLA combines BEV-feature, semantic, and spatial-context branches. Each policy produces one scalar allocation score for each eligible intervention unit.

Privileged labels, task gradients, clean representations, teacher allocations, and corruption identity or severity are excluded from the evaluation interface. Capacity is applied only through the exact-capacity allocator in Eq. 18, allowing the same frozen scoring policy to be replayed across the corresponding capacity sweep. PanoOcc uses $2{,}048$ training records over three epochs, yielding $6{,}144$ recorded optimization steps, with the epoch-3 checkpoint fixed before evaluation. The primary allocation analysis evaluates three learned-policy seeds on a frozen development package. The auxiliary utility-supervised and Task-only policies use an earlier training and development package and are reported separately. CARLA uses 8 training scenes and 2 development scenes, with $2{,}400$ optimization steps and checkpoint selection on the development split before evaluation on 3 locked scenes. FlashOCC uses three independently trained policies for the primary inference-observable comparison. Each policy contains $88{,}161$ parameters and is trained for $75{,}954$ iterations over three epochs. The corresponding direction module contains a separate $33{,}856$ parameters and is excluded from the allocation-policy count.

First-order utility capture. First-order utility capture (FOUC) measures how strongly the inference-time allocation captures privileged first-order corrective structure under the same prescribed-capacity contract. For capacity C,

$$\mathrm{FOUC}(C) = \frac{\left\langle \mathbf{u}^{(1)}, \mathbf{b}^{\mathrm{inf}}(C) \right\rangle - \left\langle \mathbf{u}^{(1)}, \mathbf{b}^{\mathrm{uni}}(C) \right\rangle}{\left\langle \mathbf{u}^{(1)}, \mathbf{b}^{\mathrm{priv}}(C) \right\rangle - \left\langle \mathbf{u}^{(1)}, \mathbf{b}^{\mathrm{uni}}(C) \right\rangle}. \tag{21}$$

FOUC uses the same capacity-matched Uniform and privileged allocations as the finite allocation comparisons. Under this normalization, Uniform and privileged allocation correspond to reference values 0 and 1. FOUC is reported when the privileged-reference denominator is nonzero and is interpreted as a signed normalized measure of first-order utility capture. It characterizes local allocation structure separately from finite-objective improvement, finite loss reduction, and finite task benefit. When multiple corrective objectives are compared, we restore the objective index A.

B Experimental and Numerical Protocols

This appendix provides implementation-level details for the experimental protocol in Sec. 2.5, including the evaluation scopes of the frozen intervention contracts, corruption protocols, matched-comparison rules, statistical procedures, and numerical validity checks.

B.1 Frozen Anchors and Evaluation Scope

Table 14 summarizes the analysis-specific evaluation scopes. Repeated anchors denote distinct intervention contracts or evaluation subsets. Annotation-induced privileged diagnostics, fixed-policy annotation evaluations, and larger-sample confirmations are listed separately.

B.2 Corruption Protocols

Table 15 summarizes the corruption protocols associated with the frozen experimental contracts. FlashOCC Q2 uses five image-space corruption families under the fully inference-observable realization contract, while the Q3 capacity analysis uses a separate oracle-health-conditioned replay contract. PanoOcc, SparseOcc fixed basis, FB-OCC, and CARLA use the same five corruption families but apply them under contract-specific camera and temporal rules. PanoOcc fixes one physical camera for each scene–condition pair, SparseOcc preserves the same corrupted camera across eight temporal frames, FB-OCC applies corruption to one fixed camera in the current keyframe, and CARLA preserves one of four corrupted cameras within each locked scene–condition pair.

Within each five-family corruption contract, *severity 3* denotes the frozen image-space corruption setting used for evaluation. The label identifies a contract-specific parameterization rather than a common numerical corruption magnitude across anchors. The corresponding settings are summarized in Table 16.

Table 14: Frozen anchors and analysis-specific evaluation scopes. Multiple rows for the same anchor denote distinct intervention contracts or evaluation subsets.

Anchor / analysis	Domain	Evaluation scope
FlashOCC (Q2)	nuScenes	150 validation scenes $\times$ 5 corruptions, 3 policy seeds
FlashOCC (Q3)	nuScenes	70-scene capacity replay, 4 capacities
PanoOcc (Q1)	nuScenes	Development split: 70 scenes, 128 keyframes $\times$ 5 corruptions, 640 records
PanoOcc (Q2/Q3)	nuScenes	70 development scenes, 1,280 corrupted records, 3 policy seeds, 4 capacities
SparseOcc (Q2)	nuScenes	Fixed-basis ranking/profile diagnostic
FB-OCC (Q1)	nuScenes	38 scenes $\times$ 5 corruptions, 190 records
CARLA (direction diagnostic)	UniOcc CARLA	3 locked evaluation scenes at reference capacity C_0
CARLA (Q2/Q3)	UniOcc CARLA	3 locked evaluation scenes, 360 frames, 4 capacities
FlashOCC / PanoOcc (annotation-induced privileged diagnostic)	nuScenes	15 scenes, 15 diagnostic frames, 3 occupancy objectives
FlashOCC (annotation-specific evaluation)	nuScenes	150 validation scenes, 5 corruption families, 3 primary allocation-policy seeds
PanoOcc (fixed-policy annotation evaluation)	nuScenes	150 validation scenes, 6,019 frames under 5 corruption conditions
PanoOcc (privileged confirmation)	nuScenes	70 scenes, 350 pairs from the previously evaluated validation pool

B.3 Evaluation Metrics and Statistical Protocol

Each intervention contract is evaluated using its corresponding frozen task metric and matched control. Finite task effects are reported as differences relative to the matched reference, with mIoU differences expressed in percentage points where applicable. Uniform provides the common reference for R_{priv} and R_{inf}, while corrected Shuffled is additionally used in the FB-OCC local-to-finite diagnostic. SparseOcc retains RayIoU for baseline reproduction, and CARLA uses pooled non-free mIoU on the locked evaluation split. For the annotation-specific FlashOCC results, Occ3D uses native 17-class pooled mIoU, while OpenOccupancy and SurroundOcc use common 10-class mean record mIoU. PanoOcc fixed-policy evaluation uses pooled common-space mIoU. Privileged annotation diagnostics are summarized separately within their respective evaluation sets.

The primary FlashOCC task gains are computed from pooled confusion matrices, while the primary PanoOcc gains are means of per-record mIoU differences. Confidence intervals use paired scene-cluster resampling, with the corresponding point-estimate statistic recomputed in each draw.

Table 15: Frozen corruption protocols used by the anchor-specific experimental branches. Distinct FlashOCC rows correspond to separate intervention contracts and evaluation scopes.

Anchor / contract	Corruption families	Protocol scope
FlashOCC (Q2)	Box blur, Gaussian noise, exposure shift, camera dropout, patch occlusion	Sev. 3, 150 scenes, 3 seeds
FlashOCC (Q3)	Frozen capacity-replay corruption set	Oracle-health-conditioned capacity protocol
PanoOcc	Box blur, Gaussian noise, exposure shift, camera dropout, patch occlusion	Sev. 3, 1 fixed camera per scene–condition
SparseOcc (fixed basis)	Box blur, Gaussian noise, exposure shift, camera dropout, patch occlusion	Sev. 3, same camera across 8 frames, 64-record diagnostic
FB-OCC	Box blur, Gaussian noise, exposure shift, camera dropout, patch occlusion	Sev. 3, 1 fixed camera per scene–condition, current frame only
CARLA	Box blur, Gaussian noise, exposure shift, camera dropout, patch occlusion	Sev. 3, 1 of 4 cameras fixed per scene–condition, 360 base frames $\times$ 5 corruption families

Table 16: Frozen severity-3 image-space parameters for the parameterized corruption families. Blur is parameterized by kernel size k, except for FB-OCC, which uses PIL radius r.

Contract	Blur	Noise σ	Exposure	Patch
FlashOCC	$k = 21$	30	$\times 1.60$	area 0.40
PanoOcc	$k = 9$	12	$\times 1.45$	side 0.28
SparseOcc	$k = 7$	24	$\times 0.40$	side 0.34
FB-OCC	$r = 5$	30	$\times 1.60$	area 0.40
CARLA	$k = 9$	12	$\times 1.45$	side 0.28

FlashOCC resamples 150 validation scenes, PanoOcc 70 development scenes, and FB-OCC 38 evaluation scenes. CARLA uses temporal-block resampling. Multi-seed summaries average the seed-level effects, while FRR is averaged over the seed-wise ratios.

The principal finite task-effect quantities are R_{priv} and R_{inf}, both evaluated relative to capacity-matched Uniform. FRR summarizes the signed ratio of inference-time task gain to privileged task gain when the corresponding privileged reference is positive. FOUC is reported separately as a normalized first-order allocation diagnostic and is interpreted independently of finite-objective, finite-loss, and task-metric effects.

B.4 Absolute Task-Metric Context

The main task-gain quantities are defined relative to capacity-matched Uniform. For context, Table 17 reports the available absolute task metrics for the primary matched comparisons. These values are interpreted within each intervention contract and are not compared across anchors.

B.5 Numerical and FP64 Validity Checks

The local directional-utility computation is checked against directional finite differences and, where available, independent Jacobian–vector product (JVP) calculations under the corresponding frozen intervention contract. Table 18 summarizes the retained numerical checks.

PanoOcc passes the retained directional finite-difference check. The SparseOcc query-space diagnostic passes its support-aware FP64 derivative check. FB-OCC shows complete directional sign agreement, near-perfect finite-difference rank agreement, and a maximum JVP/gradient-dot discrepancy of 2.21×10^{-17}. CARLA likewise shows complete sign agreement together with near-perfect finite-difference and JVP rank agreement.

Table 17: Absolute task metrics for the primary matched comparisons. FlashOCC reports seed-mean pooled-confusion mIoU on validation records, PanoOcc reports seed-mean record mIoU on development records, and FB-OCC reports pooled-confusion mIoU for the corrected local-to-finite comparison. Task gains remain relative to Uniform.

Setting	Degraded	Uniform	Reference allocation	Inference-time
FlashOCC Q2	21.8359	21.8730	22.3486	21.9665
PanoOcc Q2	57.1245	57.2092	63.9155	55.9441
FB-OCC Q1	3.1633	3.2039	3.1917	–

Table 18: Numerical validity checks for the local directional-utility computation. Sign denotes directional sign agreement, FD denotes directional finite difference, and JVP denotes Jacobian–vector product. Rank agreement is measured by Spearman correlation where available. Max discrepancy reports the retained JVP/gradient-dot discrepancy.

Contract	Check	Sign	FD rank	JVP rank	Max discrepancy
PanoOcc	Directional FD	PASS	–	–	–
SparseOcc (query space)	Support-aware FP64	PASS	–	–	–
FB-OCC	FD + JVP	1.000	$\rho = 0.999891$	–	2.21×10^{-17}
CARLA	FD + JVP	1.000	0.999884	0.999998	–

The finite-execution implementation is checked separately from the local derivative calculation. For CARLA, the maximum recorded realized-perturbation error is 2.66×10^{-12}, and the maximum capacity error is 1.59×10^{-11}. These checks verify the numerical consistency of the local directional-utility and finite-execution computations under their corresponding frozen contracts. Finite-objective and task-metric effects are evaluated separately in the corresponding experimental analyses. A separate 15-frame annotation-induced diagnostic also passes directional finite-difference checks for FlashOCC and PanoOcc under Occ3D, OpenOccupancy, and SurroundOcc. The corresponding privileged task gains are reported in Table 25.

C Extended Quantitative Results

This appendix provides complementary quantitative evidence for Sec. 3. We report privileged and inference-time task gains, capacity-dependent behavior, and objective sensitivity under the corresponding intervention and evaluation contracts.

C.1 Privileged Task Gains

Table 19 collects the privileged references used by the corresponding analyses. FlashOCC, PanoOcc, and CARLA provide positive references for their matched inference-time comparisons, while FB-OCC remains part of the Q1 local-to-finite analysis because its finite task endpoint does not establish positive task gain.

C.2 Inference-Time Task Gains

Table 20 summarizes inference-time task gains for contracts with a positive privileged task-gain reference and a matched inference-time allocation. FlashOCC, PanoOcc, and CARLA are evaluated at their contract-specific $C_0^{(a)}$, while SparseOcc uses the fixed-basis ranking/profile replay at its retained reference capacity. FB-OCC remains part of the Q1 local-to-finite analysis because its finite GT–Uniform task effect does not establish a positive privileged task gain at the reference capacity.

Table 21 reports the seed-level estimates underlying the aggregate results. Table 22 provides the matched PanoOcc policy comparisons and capacity trajectories, while Table 23 reports the fixed-profile SparseOcc ranking diagnostic.

Table 19: Finite privileged reference effects under the corresponding frozen intervention contracts. R_{priv} is measured relative to capacity-matched Uniform under the same intervention contract.

Anchor / analysis	Reference comparison	R_{priv} (pp)	95% CI	Role
FlashOCC (Q2)	Privileged vs. Uniform at $C_0^{(a)}$ under the fully inference-observable direction	$+0.4757$	–	Inference-time task-gain reference
PanoOcc (Q2/Q3)	Privileged vs. Uniform at $C_0^{(a)}$ under the frozen voxel direction	$+6.7062$	–	Policy and capacity reference
FB-OCC (Q1)	GT-guided vs. Uniform at $C_0^{(a)}$	-0.0122	$[-0.0788, +0.0379]$	Corresponding finite task endpoint
CARLA	Privileged vs. Uniform at $C_0^{(a)}$ under the inference-compatible direction	$+0.0652$	$[+0.0589, +0.0715]$	Inference-time allocation reference

Table 20: Privileged and inference-time task gains under the corresponding intervention contracts. R_{priv} and R_{inf} are measured relative to capacity-matched Uniform. For FlashOCC and PanoOcc, the reported R_{priv} and R_{inf} values are three-seed mean effects, while FRR is the mean of the corresponding seed-wise ratios.

Anchor / contract	R_{priv} (pp)	R_{inf} (pp)	FRR	Robustness evidence
FlashOCC (fully inference-observable)	$+0.4757$	$+0.0935$	0.1937	3/3 seed CIs positive, 14/15 seed–corruption effects positive
PanoOcc (learned allocation)	$+6.7062$	-1.2651	-0.1886	3/3 seed CIs negative, mean-effect 95% CI $[-1.6698, -0.8181]$
SparseOcc (ranking/profile replay)	$+1.7205$	$+0.0361$	0.0210	Fixed-basis diagnostic
CARLA (inference-compatible direction)	$+0.0652$	$+0.0267$	0.4101	R_{inf} 95% CI $[+0.0095, +0.0429]$

Table 21: Seed-level privileged and inference-time task gains at $C_0^{(a)}$. FlashOCC uses pooled-confusion mIoU differences, while PanoOcc uses mean per-record mIoU differences. Confidence intervals use 10,000 paired scene-cluster bootstrap draws with the corresponding estimator. The PanoOcc privileged allocation is independent of the learned-policy seed.

Contract	Seed	R_{priv} (pp)	95% CI	R_{inf} (pp)	95% CI
FlashOCC	0	$+0.479197$	$[+0.439513, +0.521658]$	$+0.109358$	$[+0.096913, +0.122441]$
FlashOCC	1	$+0.511255$	$[+0.464313, +0.559054]$	$+0.118130$	$[+0.103944, +0.133057]$
FlashOCC	2	$+0.436506$	$[+0.401568, +0.472099]$	$+0.053148$	$[+0.044059, +0.063078]$
PanoOcc	0	$+6.706212$	$[+6.024972, +6.871645]$	-1.320571	$[-1.785094, -1.063730]$
PanoOcc	1	$+6.706212$	$[+6.035469, +6.854599]$	-1.245245	$[-1.724424, -0.999573]$
PanoOcc	2	$+6.706212$	$[+6.027251, +6.867693]$	-1.229491	$[-1.699052, -0.985909]$

C.3 Capacity-Dependent Realization

Table 24 reports the complete four-capacity trajectories underlying Sec. 3.3. Figure 6 further shows the seed-level PanoOcc learned response and the contrasting trajectories of the alternative allocation policies.

Table 22: PanoOcc allocation effects under the frozen voxel intervention contract. Panel A compares inference-time policies and the shuffled-teacher diagnostic at the reference capacity $C_0^{(a)}$. Panel B reports the capacity trajectories of the inference-time policies. Values are three-seed mean task-metric differences relative to capacity-matched Uniform.

Panel A: Reference-capacity allocation effects

Allocation	Role	Effect vs. Uniform (pp)
Learned	Inference-time	-1.2651
Entropy	Inference-time	$+0.1204$
Low confidence	Inference-time	$+0.1694$
Foreground uncertainty	Inference-time	-0.0018
Shuffled teacher	Diagnostic control	-0.2395

Panel B: Inference-time policy capacity response

Allocation policy	$0.25C_0^{(a)}$	$0.50C_0^{(a)}$	$0.75C_0^{(a)}$	$1.00C_0^{(a)}$
Learned	$+0.0100$	-0.6206	-0.9910	-1.2651
Entropy	$+0.0285$	$+0.0580$	$+0.1056$	$+0.1204$
Low confidence	$+0.0534$	$+0.0867$	$+0.1535$	$+0.1694$
Foreground uncertainty	-0.0037	-0.0089	$+0.0019$	-0.0018

Table 23: SparseOcc ranking/profile diagnostic under the fixed-basis intervention. Priv. and Inf. denote privileged and inference-time components. Loss red. denotes finite loss reduction.

Ranking	Profile	ΔmIoU (pp)	FOUC	Loss red.
Priv.	Priv.	$+1.7205$	1.0000	$+2.63266$
Priv.	Inf.	$+0.5480$	0.2275	$+0.66959$
Inf.	Inf.	$+0.0361$	0.0609	$+0.17971$

Table 24: Capacity-dependent realization under the corresponding frozen Q3 contracts. FlashOCC, PanoOcc, and CARLA report matched R_{priv} and R_{inf} trajectories across all four capacities. PanoOcc values are three-seed means. Dashes indicate quantities not retained as formal aggregate estimates for the corresponding evaluation.

Contract	$C/C_0^{(a)}$	R_{priv} (pp)	R_{inf} (pp)	FRR	FOUC
FlashOCC (oracle-health-conditioned)	0.25	$+0.4934$	$+0.0826$	0.1675	0.1217
	0.50	$+0.8844$	$+0.1486$	0.1680	0.1357
	0.75	$+1.1482$	$+0.2214$	0.1929	0.1506
	1.00	$+1.3404$	$+0.2891$	0.2157	0.1668
PanoOcc (learned)	0.25	$+6.4692$	$+0.0100$	–	–
	0.50	$+6.7100$	-0.6206	–	–
	0.75	$+6.7492$	-0.9910	–	–
	1.00	$+6.7062$	-1.2651	-0.1886	–
CARLA (inference-compatible)	0.25	$+0.0182$	$+0.0083$	0.4586	0.7236
	0.50	$+0.0340$	$+0.0169$	0.4972	0.5576
	0.75	$+0.0481$	$+0.0215$	0.4463	0.4818
	1.00	$+0.0652$	$+0.0267$	0.4101	0.4337

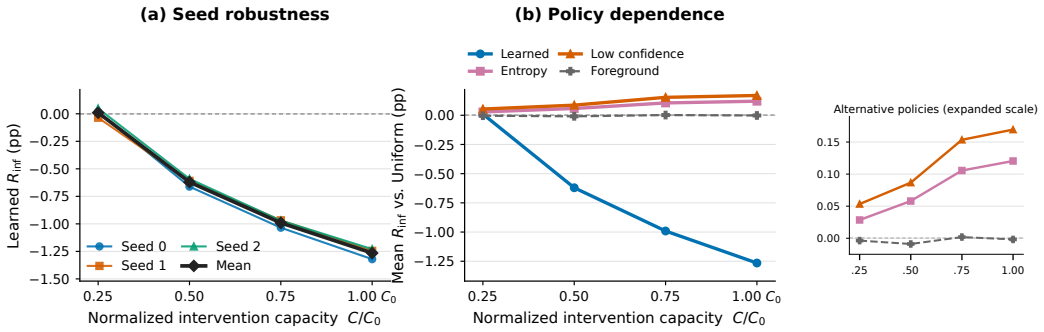


Figure 6: PanoOcc capacity-response diagnostics under the frozen voxel intervention contract. (a) Seed-level learned R_{inf} trajectories for the three independently trained policies and their mean. (b) Three-seed mean R_{inf} trajectories for the learned and alternative inference-time allocation policies. The auxiliary view on the right expands the near-zero range of the alternative policies. All effects are measured relative to capacity-matched Uniform at the corresponding prescribed capacity.

Table 25: Privileged task gains (pp) in the 15-frame, 15-scene annotation-induced diagnostic. Each objective separately defines the task loss and corrective utility. Directional finite-difference checks pass under all three objectives for both anchors.

Anchor	Occ3D	OpenOccupancy	SurroundOcc
FlashOCC	+0.638	+0.076	+0.051
PanoOcc	+5.417	+2.152	+0.520

C.4 Additional Objective-Sensitivity Evidence

Fixed-policy cross-objective evaluation. The fixed-policy PanoOcc experiment in Table 6 evaluates the same Occ3D-derived policy across Occ3D, OpenOccupancy, and SurroundOcc on 150 validation scenes and 6,019 frames under five corruption conditions. The annotation/evaluation target changes while the policy remains fixed. The task-effect point estimate changes from negative under Occ3D to positive under the alternative annotations. FlashOCC yields positive effects under each reported annotation-specific evaluation, with uncertainty estimates reported in Table 26.

Annotation-induced privileged utility. For each annotation objective A, the privileged diagnostic separately constructs the common-space objective $\mathcal{L}_A$, computes $\nabla_F \mathcal{L}_A$, and derives $\mathbf{u}_A^{(1)}$ before exact-capacity allocation. The same annotation-specific common-space NLL is used for the corresponding finite-objective evaluation. Directional finite-difference checks pass for FlashOCC and PanoOcc under all three objectives in the 15-frame, 15-scene diagnostic. Table 25 reports the corresponding finite task gains. Both anchors obtain positive privileged gains under each independently induced objective.

Larger-sample privileged confirmation. PanoOcc additionally evaluates privileged allocation on 70 scenes and 350 pairs. These scenes are disjoint from the 15-scene diagnostic and belong to the previously evaluated validation pool. As shown in Table 26, the privileged effects under OpenOccupancy and SurroundOcc are positive relative to Uniform, with confidence intervals excluding zero, and are also positive relative to Shuffled. This provides larger-sample support for the privileged gains observed in the annotation-induced diagnostic.

Table 26: Additional statistical evidence for the annotation experiments. FlashOCC reports common-class mean record effects with paired scene-cluster bootstrap intervals. PanoOcc reports the separate 70-scene privileged confirmation. Confidence intervals correspond to the Uniform comparison. Shuffled is an auxiliary control.

Analysis	Annotation	Quantity	vs. Uniform (pp)	95% CI	vs. Shuffled (pp)
FlashOCC	OpenOccupancy	R_{priv}	$+0.1570$	$[+0.1465, +0.1675]$	–
	OpenOccupancy	R_{inf}	$+0.0628$	$[+0.0580, +0.0679]$	–
	SurroundOcc	R_{priv}	$+0.1043$	$[+0.0968, +0.1120]$	–
	SurroundOcc	R_{inf}	$+0.0448$	$[+0.0411, +0.0486]$	–
PanoOcc privileged confirmation	OpenOccupancy	R_{priv}	$+2.1264$	$[+1.8465, +2.3853]$	$+2.1377$
	SurroundOcc	R_{priv}	$+0.5638$	$[+0.4862, +0.6350]$	$+0.5629$

D TASK-GAIN DIAGNOSTICS

This section uses controlled diagnostics to localize structure associated with the observed privileged–inference task-gain gaps while keeping the corresponding intervention contracts and evaluation metrics fixed.

D.1 PANOOCC AUXILIARY POLICY DIAGNOSTIC

Table 27: Signed first-order utility at C_0 for a single-seed PanoOcc policy pair on a separate 1,280-record development package. G_+ and G_- denote contributions from positive- and negative-utility units, with $G = G_+ + G_-$.

Allocation policy	G_+	G_-	G
Utility-supervised	$+0.230934$	-0.128021	$+0.102913$
Task-only	$+0.132888$	-0.046946	$+0.085942$

Table 27 compares a matched single-seed utility-supervised and Task-only policy pair trained with the same architecture, initialization, records, duration, and fixed epoch-3 selection. The utility-supervised policy captures larger positive and negative first-order contributions, while the net captured utility remains positive for both.

On the same auxiliary development package, the frozen Task-only policy increases mean record mIoU from 56.1973 under Uniform to 56.6403, yielding $+0.4430$ pp. The primary three-seed PanoOcc analysis uses a separate frozen record package. Both analyses average per-record mIoU.

D.2 PANOOCC SPATIAL ALLOCATION AND TEMPORAL-CONTEXT DIAGNOSTICS

Figure 7 reports the four conditional XY–Z replacements. Replacing Z under the inference-time XY condition produces the larger single-component effect and closes 88.6% of the diagnostic gap. Because these are conditional replacements, they are not interpreted as an additive decomposition.

The four conditions combine privileged or inference-time XY and Z allocation components within the same frozen mixed-allocation contract. The intervention direction, prescribed capacity, and evaluation records are fixed, while the source of the XY or Z allocation component is exchanged.

Temporal-history diagnostic. The temporal diagnostic varies history while preserving the current-frame image, pose, and calibration. Correct history improves task performance relative to OFF, WRONG, and SHUFFLED histories. The native temporal representation differs structurally from the intervention grid, so this comparison is used as evidence of task-relevant temporal context rather than as a repair-grid attribution.

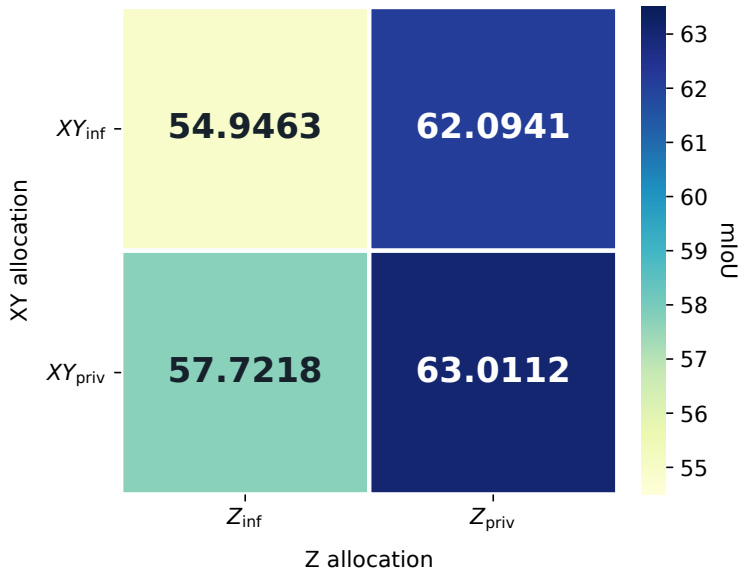


Figure 7: PanoOcc XY–Z allocation diagnostic under the frozen mixed-allocation contract. Each cell reports mIoU under the corresponding combination of inference-time or privileged XY and Z allocations. Starting from the fully inference-time configuration, replacing only the Z allocation with its privileged counterpart increases mIoU from 54.9463 to 62.0941, closing 88.6% of the diagnostic gap to the fully privileged configuration.

Together, these diagnostics identify sensitivity to conditional vertical allocation structure and to temporal context under their respective evaluation settings. The spatial replacement directly probes the privileged–inference task-gain gap, while the temporal comparison is used only as supporting evidence that historical context affects task performance.

D.3 CARLA Direction and Allocation Diagnostic

Table 28 separates variation across intervention directions from allocation realization under a fixed inference-compatible direction. Under the inference-compatible direction, replacing privileged allocation with inference-time allocation yields an FRR of 0.4101.

Table 28: CARLA direction and allocation diagnostic at $C_0^{(a)}$. The privileged-allocation rows compare finite task effects across two direction contracts. Under the inference-compatible direction, the final two rows compare privileged and inference-time allocation relative to their matched Uniform reference.

Direction	Allocation	Δ mIoU (pp)
Privileged GT-gradient	Privileged	+0.56456
Inference-compatible	Privileged	+0.06518
Inference-compatible	Inference-time	+0.02673

D.4 CARLA Finite-Loss Bridge

CARLA additionally provides matched post-intervention loss values, allowing the same intervention contract to be examined in differentiable-objective space and task-metric space. We define the finite loss-reduction effects relative to the capacity-matched Uniform allocation as

$$R^{\text{loss}}_{\text{priv}}(C) = \mathcal{L}\big(\mathbf{b}^{\text{uni}}(C)\big) - \mathcal{L}\big(\mathbf{b}^{\text{priv}}(C)\big)\,, \qquad R^{\text{loss}}_{\text{inf}}(C) = \mathcal{L}\big(\mathbf{b}^{\text{uni}}(C)\big) - \mathcal{L}\big(\mathbf{b}^{\text{inf}}(C)\big)\,, \quad (22)$$

so positive values indicate lower post-intervention loss than the matched Uniform reference. Table 29 reports these quantities across the frozen capacity sweep together with FOUC and the corresponding inference-time mIoU effect.

Table 29: CARLA finite-loss bridge across the frozen capacity sweep. $R^{\mathrm{loss}}_{\mathrm{priv}}$ and $R^{\mathrm{loss}}_{\mathrm{inf}}$ denote finite loss reductions relative to capacity-matched Uniform, with $\mathrm{FRR}_{\mathrm{loss}} = R^{\mathrm{loss}}_{\mathrm{inf}} / R^{\mathrm{loss}}_{\mathrm{priv}}$. The final column reports the corresponding inference-time mIoU effect.

$C/C_0^{(a)}$	FOUC	$R^{\mathrm{loss}}_{\mathrm{priv}}$	$R^{\mathrm{loss}}_{\mathrm{inf}}$	$\mathrm{FRR}_{\mathrm{loss}}$	$R^{\mathrm{mIoU}}_{\mathrm{inf}}$ (pp)
0.25	0.7236	+0.008014	+0.006099	0.7611	+0.00833
0.50	0.5576	+0.014001	+0.008079	0.5770	+0.01689
0.75	0.4818	+0.019277	+0.009517	0.4937	+0.02148
1.00	0.4337	+0.024190	+0.010679	0.4415	+0.02673

Across all four capacities, first-order utility capture, finite loss reduction, and inference-time mIoU gain remain positive. Their magnitudes follow different trajectories, showing that normalized first-order capture and finite task gain measure different aspects of the intervention response.