GATE3R: TRAINING-FREE PRIOR GATING FOR FEED-FORWARD 3D RECONSTRUCTION

Anonymous authors
Paper under double-blind review

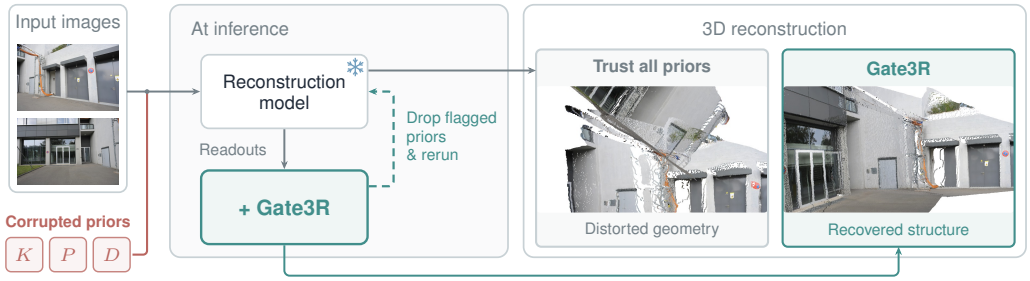

Figure 1: **Training-free prior gating with Gate3R.** Images and priors are first passed through the frozen reconstruction model. Gate3R checks its readouts, drops flagged priors, and reruns the model without training. On ETH3D with corrupted intrinsics K, pose P, and depth D, using all priors degrades the MapAnything reconstruction, while Gate3R recovers the scene structure.

ABSTRACT

Feed-forward reconstruction models such as DUSt3R and its successors recover geometry from unposed images in an end-to-end manner. Some of these models are prior-conditioned, taking optional geometric priors such as camera intrinsics, relative pose, and depth to further improve reconstruction. In practice, however, such priors are often inaccurate or corrupted, and conditioning on them can produce reconstructions worse than those from images alone. We propose Gate3R, a training-free method to discard unreliable priors at test time and fall back to image-only reconstruction. It combines the residual between the prior and the prediction with the frozen model's attention entropy to decide whether each prior is usable. On two prior-conditioned models, Pow3R and MapAnything, Gate3R outperforms both always keeping and always discarding priors in most single-prior settings and in all multi-prior settings across six datasets.

1 INTRODUCTION

Feed-forward models have reshaped 3D reconstruction from unposed images. Since DUSt3R (Wang et al., 2024), a growing family of models (Wang et al., 2025a;b; 2026b;a) can recover dense geometry from unposed images in a single network pass, scaling to many views and streaming input without iterative optimization. Several recent works (Jang et al., 2025; Keetha et al., 2026; Liu et al., 2026) make these models prior-conditioned, accepting optional geometric priors such as camera intrinsics, relative pose, and depth, and conditioning the reconstruction on them. When the provided priors are accurate, conditioning yields substantial gains.

In real-world deployment, however, geometric priors are rarely clean. Camera calibrations go stale as lens parameters drift, and intrinsics estimated from a single image carry residual error (Jin et al., 2023; Veicht et al., 2024). Visual odometry accumulates drift that corrupts relative pose estimates (Mur-Artal et al., 2015; Engel et al., 2018; Teed & Deng, 2021), and monocular depth predictors carry unknown scale and systematic bias (Ranftl et al., 2022; Yang et al., 2024). Corrupted priors are therefore not an edge case but an expected condition in any pipeline that feeds one model's estimate into another. Existing prior-conditioned models nevertheless trust their inputs

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.

unconditionally. Past a certain severity, a corrupted prior silently degrades the reconstruction, often below the quality obtained from images alone, and without any warning. These models provide no mechanism to decide, per image set, whether a given prior is usable or should be discarded.

We show that the answer lies in the model's own computation, in line with recent evidence that frozen geometric foundation models encode rich emergent structure in their intermediate representations (Ziliotto et al., 2026; Bratulić et al., 2026; Han et al., 2026). We first identify an *echo effect*: a frozen model reproduces most of an injected prior error in its output, as if it accepted the wrong prior. Yet the model still registers the error within the same forward pass. At the output, the echo is incomplete, so the residual between the prior and the prediction retains the uncopied error. Inside the network, the attention entropy exposes the conflict between the prior and the image evidence. The two readouts capture the same conflict at different levels, and combining them is more reliable than either alone in most settings.

Building on this finding, we propose Gate3R, a training-free method that discards unreliable priors at test time and falls back to image-only reconstruction. Gate3R reads both readouts from a single conditioned pass of the frozen model and weights them by the heaviness of their upper tails on the deployment batch, without labels or clean reference statistics at test time. The only labeled step is a one-time head selection on a held-out dataset with self-injected corruptions. The combined per-set score decides whether each prior is usable. Across two prior-conditioned models, Pow3R and MapAnything, three prior types, and six datasets under controlled corruption, Gate3R outperforms both always keeping and always discarding priors on eight of the nine single-prior settings, all except Pow3R depth, and on all multi-prior settings. Our contributions are as follows.

- We formulate prior reliability in prior-conditioned feed-forward 3D reconstruction as a test-time decision of whether to keep or discard each geometric prior, and establish a paired corruption protocol over three prior types and six datasets to evaluate it.
- We identify the echo effect: frozen reconstructors reproduce on average up to 93% of an injected prior error in their output, yet the error remains detectable within the same pass, and on the two models we study, the strength of the echo follows how each model injects the prior rather than the prior type.
- We propose Gate3R, a training-free method that fuses the two readouts with weights adapted to each deployment batch, requiring neither labels nor clean reference statistics at test time.

2 Related Work

Prior-conditioned feed-forward reconstruction. Feed-forward models, including DUSt3R (Wang et al., 2024), VGGT (Wang et al., 2025a), π^3 (Wang et al., 2026b), and VGGT-Ω (Wang et al., 2026a), have advanced 3D reconstruction from unposed images. Recent works further condition on available poses, intrinsics, and depth: Pow3R (Jang et al., 2025) injects ray and depth embeddings into its encoder and relative pose into decoder tokens, MapAnything (Keetha et al., 2026) fuses encoded geometric inputs with image features, and WorldMirror (Liu et al., 2026) concatenates camera tokens with image tokens and adds depth embeddings. These designs show the benefit of reliable priors, but a corrupted prior can misguide reconstruction below image-only inference.

Emergent capabilities and training-free adaptation. Large-scale reconstruction training yields internal representations that support capabilities beyond the supervised outputs: probing shows that epipolar geometry emerges in intermediate layers and is linked to attention correspondence patterns (Bratulić et al., 2026), and image diffusion models likewise develop dense correspondence without explicit supervision (Tang et al., 2023). Such representations enable training-free adaptation. Easi3R (Chen et al., 2025) disentangles motion in DUSt3R's attention maps for dynamic scene reconstruction, and RobustVGGT (Han et al., 2026) rejects distractor views using attention or feature similarity, without fine-tuning or outlier supervision.

Uncertainty in feed-forward reconstruction. Feed-forward reconstructors commonly predict a per-pixel confidence alongside geometry (Wang et al., 2024; 2025a; Keetha et al., 2026).

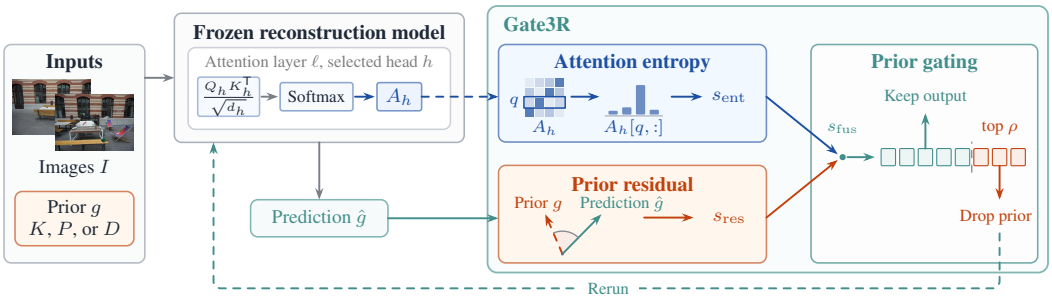

Figure 2: **Overview of Gate3R.** Gate3R reads the attention entropy s_{ent} of a selected head and the residual s_{res} between the prior g and the prediction $\hat{g}$, and fuses them into s_{fus}. The view sets of a batch with the top fraction ρ of s_{fus} drop the prior and rerun on images alone, and the others keep the conditioned output.

Trust3R (Zhu et al., 2026) replaces this heuristic confidence with an evidential head for per-point uncertainty, and conformal prediction yields prediction sets with coverage guarantees for keypoint-based pose estimation (Yang & Pavone, 2023). These methods quantify the uncertainty of the output rather than the reliability of an input prior, and under a high echo the model remains confident about an output that reproduces the prior error. We therefore use the built-in confidence only as a baseline.

Consistency checks and attention entropy. Consistency between redundant predictions is a classical verification signal (Zamir et al., 2020; Manakul et al., 2023); TCO (Zhou et al., 2026) adapts a reconstructor at test time to agree with its priors. Such checks assume the prediction is independent of its reference, which a conditioned model that echoes its prior only partly satisfies, much as language models adopt conflicting context (Xie et al., 2024). Attention maps also expose model failures, from hallucinations (Chuang et al., 2024; Huang et al., 2024; Song et al., 2026) to backdoors (Ge et al., 2026). We bring attention entropy to geometric conditioning and use it alongside the residual as a second readout of whether the model accepts its prior.

3 Method

We first formalize the prior-gating task. We then describe how the prior residual and the attention entropy, both obtained from a single conditioned pass, are combined into a gate score that decides whether to keep or discard the prior, as Figure 2 illustrates.

Problem setting. Geometric priors such as camera intrinsics K, relative pose P, and depth D can improve feed-forward 3D reconstruction when they are accurate, but degrade it when they are wrong. Let M be a frozen reconstruction model that takes N images I and optional priors $g \subseteq \{K, P, D\}$ as input and recovers the full 3D geometry. For each image set, our goal is to make a binary decision $d \in \{0, 1\}$: keep the conditioned output $M(I, g)$ when d=0, or discard the prior and re-run as $M(I)$ when d=1. We require d to be determined training-free, from signals internal to the single conditioned pass, without ground-truth geometry.

Prior residual. Given the conditioned pass $M(I, g)$, we compare the injected prior g with the prediction $\hat{g}$ of the same quantity from this pass, and take their discrepancy as the *prior residual*,

$$s_{\text{res}} = d(g, \hat{g}), \tag{1}$$

where d is a distance in the unit of the prior: the per-pixel ray angle averaged over the image for K, the larger of the rotation geodesic and the translation-direction angle for P, and the median log-depth difference over valid pixels for D. Pow3R predicts up-to-scale geometry, so we recover its relative pose by confidence-weighted Umeyama registration of the two point maps and center its log-depth difference by its median. With two views, the set residual is the maximum over views; Section A gives the four-view aggregation. A large residual means the model's own geometry disagrees with the given prior, so the prior is suspect. The residual is a self-consistency check: it uses only the

injected prior and the pass's own output and needs no second forward pass. Its reliability, however, depends on how the model treats a wrong prior. A frozen model partly reproduces an injected error in its output, so prior and prediction partly agree and the residual shrinks. We call this the *echo effect*; on the two models we study, its strength follows how each model injects the prior rather than the prior type.

Attention entropy. Given the same conditioned pass $M(I, g)$, we read the attention map of a selected head h and take its mean Shannon entropy as the *attention entropy*,

$$s_{\text{ent}} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} H\big(\mathbf{A}_h[q, :]\big), \tag{2}$$

where $\mathbf{A}_h[q, :]$ is the attention distribution of query q in head h, $H(\cdot)$ is its Shannon entropy (Shannon, 1948), and $\mathcal{Q}$ is the set of queries. The entropy reads the computation rather than the output. The model fuses the prior with the image evidence in its attention layers (Bratulić et al., 2026), and a conflicting prior disturbs this attention even when the output largely copies the prior. Figure 3 shows that the entropy shifts on the views that receive a corrupted prior. The head is selected once per model and prior type as the one with the highest clean-versus-corrupted AUC on the two-view sets of one held-out dataset, using self-injected corruptions as the only labels. This one-time selection is the only step that uses labels; the head is then frozen for all test datasets, view counts, and seeds, and deployment requires no labels; Table 13 lists the deployed heads.

Gate signal. We fuse the two readouts of every view set in a deployment batch into one *gate signal* with a *tail-weighted fusion*. How informative each readout is depends on the model and the prior type and is unknown at test time, so we weight the two readouts from the unlabeled batch itself. A readout that detects corruption places the corrupted sets above the clean ones, so its distribution over the batch develops a long upper tail, whereas an uninformative readout stays roughly symmetric around its median.

We first standardize each readout over the batch with a robust z-score. For the values $\mathbf{x} = (x_1, \ldots, x_B)$ of one readout on the B view sets,

$$z_b = \frac{x_b - Q_{0.5}(\mathbf{x})}{Q_{0.5}\big(|\mathbf{x} - Q_{0.5}(\mathbf{x})|\big)}, \qquad b = 1, \ldots, B, \tag{3}$$

where Q_q is the q-quantile over the batch. The median and the median absolute deviation depend only on the lower half of the batch, so they reflect the noise of the readout as long as fewer than half of the sets are raised by corruption. We then weight each readout by the asymmetry of its tails and fuse the two,

$$w(\mathbf{x}) = \frac{Q_{0.95}\big(z(\mathbf{x})\big)}{\big|Q_{0.05}\big(z(\mathbf{x})\big)\big|} - 1, \qquad s_{\text{fus}} = w_{\text{res}}\, z(s_{\text{res}}) + w_{\text{ent}}\, z(s_{\text{ent}}), \tag{4}$$

where $z(\mathbf{x}) = (z_1, \ldots, z_B)$, and w_{res} and w_{ent} are w evaluated on the residuals and the entropies of the batch. A symmetric readout has a ratio of one and thus zero weight, while a readout that separates the corrupted sets receives a weight that grows with the separation.

Discard rule. Gate3R ranks the view sets of a batch by s_{fus}, discards the prior of the top fraction ρ, and re-runs these sets on images alone. The budget ρ is not an estimate of how many priors are corrupted, which is unknown at test time. It fixes the compute, $1 + \rho$ forward passes per set, in the same way that selective prediction fixes the coverage and ranks inputs by a confidence score (Geifman & El-Yaniv, 2017). Setting $\rho = 0$ recovers keeping every prior and $\rho = 1$ discarding every prior, so varying ρ traces a curve between the two fixed policies, and discarding at random follows the straight line between them in expectation. We use $\rho = 0.3$ throughout and report the full curve in Figure 5.

4 Experiments

We first describe the data, the deployment protocol, and the metrics, then report gating with a single prior and with several priors, and finally analyze the echo effect and the attention entropy, ablate

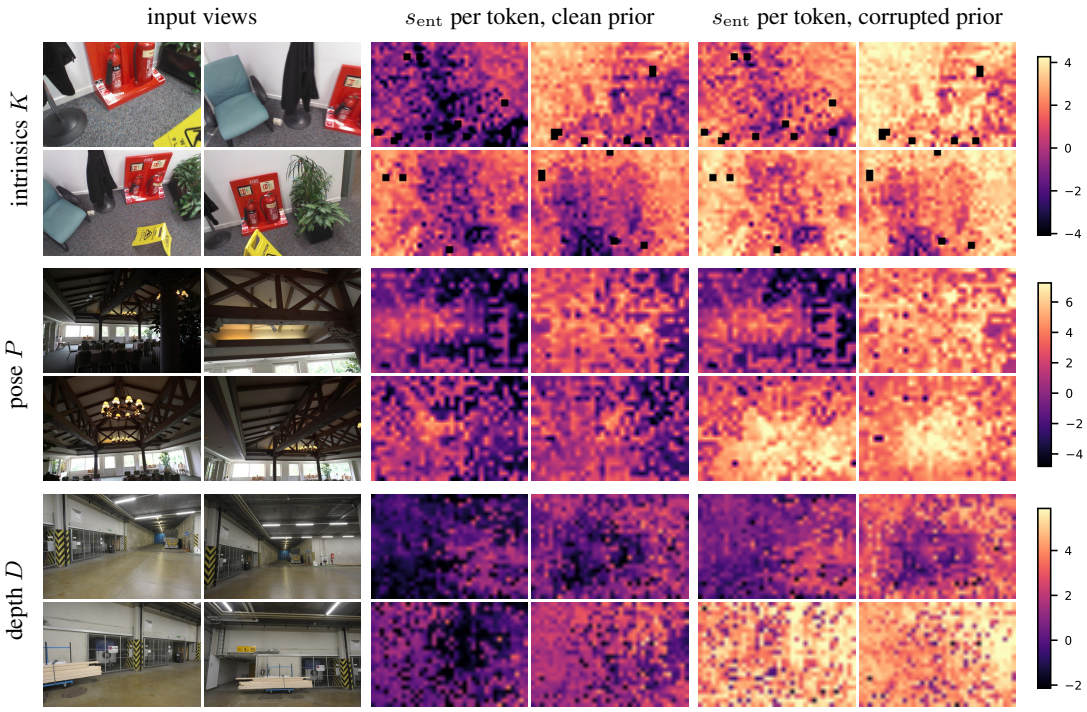

Figure 3: **Attention entropy maps.** Per-token entropy of MapAnything is compared under a clean and a corrupted prior. The entropy shifts on every view that receives a corrupted prior, where brighter indicates higher entropy.

the gate score, study the discard budget, and show qualitative results. Gate3R is training-free: one attention head per model and prior type is selected once on a held-out dataset and frozen for all experiments, and no labels are used at test time.

4.1 Experimental setup

Models and implementation. We study two frozen prior-conditioned models: Pow3R (Jang et al., 2025) with its released ViT-L checkpoint, and MapAnything (Keetha et al., 2026) with the public V1 checkpoint `facebook/map-anything-v1`. Images are resized to a longest side of 512 and 518 pixels, respectively, keeping their aspect ratio. Pow3R injects K-derived rays and mean-normalized depth at its encoder and relative pose as a decoder token, whereas MapAnything fuses K as a ray field with the image features and feeds P and D as input tokens to its transformer. Inference uses bf16 on a single NVIDIA H200 GPU, and the entropy is computed in float32 from the post-softmax attention. Each attention map is discarded right after its entropy is read, so peak memory is unchanged. Reading both readouts costs 1.23–1.39× a plain forward pass, and with the re-runs Gate3R costs 1.47–1.68× on MapAnything and 1.9–3.2× on our unoptimized Pow3R implementation.

Datasets. We evaluate on six datasets: ETH3D, 7-Scenes, 12-Scenes, DTU, LaMAR, and Tanks and Temples. Their scene types and ground-truth sources are listed in Table 6. None of them is part of the training data of either model, so every result is zero-shot. We hold out LaMAR for head selection and test on the other five datasets. Each view set is evaluated with two views and, on MapAnything, also with four views following Keetha et al. (2026).

Corruptions. We corrupt each prior at five severities a, with each corruption type modeled on a physical failure and summarized in Table 1. For intrinsics, the focal length of each view is scaled by $e^{\pm a}$ with an independent sign, as from a stale calibration or a wrong EXIF entry. For pose, to mimic odometry drift, we rotate both the relative rotation and the translation direction of every

Table 1: **Corruption protocol.** A corrupted prior is perturbed at one of five severities a. Here ε is a random-sign magnitude equal to a, and s is the median valid depth of the view.

Prior	Source	Corruption	Levels a	Across views
K	stale calibration, wrong EXIF	$f \leftarrow f\, e^{\varepsilon}$	$\{0.02, 0.05, 0.1, 0.2, 0.4\}$	independent
P	VO / SLAM drift	$R \leftarrow \Delta R(a)\, R$, $\hat{t} \leftarrow$ rotated by a	$\{1, 5, 15, 45, 90\}^\circ$	non-reference views
D	mis-scaled depth estimator	$d \leftarrow d\, e^{\varepsilon_1} + \varepsilon_2\, s$	$\{0.01, 0.02, 0.05, 0.1, 0.2\}$	independent

non-reference view by exactly a in random directions, keeping $\|t\|$. For depth, each view receives a scale drift and an additive bias of random signs, as from a mis-scaled depth estimator. Since signs are independent across views, about half of the sets are consistently wrong, and the rest carry a cross-view contradiction of up to $2a$. Seeds are fixed per set and shared across models, levels, view counts, and strategies, so all comparisons are paired.

Deployment protocol. Every number is measured on a deployment batch that mixes clean and corrupted priors without labels. The batch of a test dataset contains all of its view sets. With a single prior, one sixth of the sets keep the clean prior and each of the five corruption levels takes another sixth. With several priors, each fed prior is drawn independently and uniformly from its clean value and its five levels, which covers all 36 configurations of two priors and all 216 of $K{+}P{+}D$. All policies share the same draws, and we repeat them over 10 seeds. Gate3R discards at the budget $\rho = 0.3$. With several priors, each fed prior is ranked by its own score under its own budget.

Baselines. We compare with two fixed policies: All keeps every prior, and None discards every prior. Two further baselines discard at the same budget ρ as Gate3R: Random selects the sets at random, and Confidence selects the sets with the lowest mean predicted confidence. As a reference, the Oracle uses the ground truth to pick the better of All and None for each set, or the best subset of the fed priors with several priors. This per-set selection with the ground truth is optimistic.

Evaluation metrics. Following Keetha et al. (2026), we score each prior by the output quantity it constrains: the ray-direction error in degrees for K, the relative-pose AUC@5° for P, and the z-depth inlier ratio $\delta{<}1.03$ for D, computed after normalizing both geometries by their mean point distance. The depth metric excludes Tanks and Temples, whose ground-truth depth is sparse.

4.2 Results

Main results. Gate3R outperforms both fixed policies on eight of the nine single-prior axes in Table 2 and on all 27 multi-prior cells in Table 3, where each fed prior is gated separately with the same frozen head. The only exception is Pow3R depth, where All performs best. Pow3R's depth normalization removes most of the corruption before it reaches the network, so most corrupted depth priors still help, and discarding a fixed 30% of each batch removes more useful priors than harmful ones. At the same budget, Gate3R also outperforms Random and Confidence in all settings except two-view depth, where it is on par with Confidence. Its gains therefore come from choosing which priors to discard rather than from the budget itself. The two-view heads transfer unchanged to four views, where Gate3R remains the best on all three axes, even on depth, although keeping the depth prior hurts on average. Table 8 gives per-dataset results.

Echo effect. A frozen model partly reproduces an injected prior error in its output, so the residual sees only the part it does not reproduce. We quantify this echo by perturbing a clean prior $g^\star$ into g and measuring how much of the perturbation reappears in the output,

$$e = \frac{d(\hat{g},\, g^\star) - d(\hat{g}^\star,\, g^\star)}{d(g,\, g^\star)}, \tag{5}$$

where $\hat{g}^\star$ is the prediction under the clean prior. The numerator is the increase in output error caused by the perturbation, in the unit of the injected error. We read the echo at the severest level of each axis, where the injected error dominates the clean error; it serves only for analysis and is never computed at deployment. Averaged over datasets, every axis echoes 66 to 93% of the error that reaches the network, as Table 4 shows, although single datasets fall lower, such as 46% for

Table 2: **Gating results.** Means over the five test datasets and ten seeds; 2v and 4v denote two and four views. Depth excludes Tanks and Temples, whose ground-truth depth is sparse. Random and Confidence discard at the same budget $\rho = 0.3$ as Gate3R. Oracle picks the better of None and All for each view set. Bold marks the best result per model, excluding Oracle.

Model	Priors used	K: ray err (deg) ↓	P: AUC@5° ↑	D: inlier δ<1.03 ↑
Pow3R (2v)	None	1.80	30.0	0.400
	All	2.07	24.3	**0.590**
	Random	1.99	25.9	0.526
	Confidence	1.87	30.9	0.558
	Gate3R (ours)	**1.54**	**32.1**	0.558
	Oracle	0.88	41.9	0.645
MapAnything (2v)	None	2.59	29.0	0.375
	All	2.20	28.4	0.393
	Random	2.34	27.7	0.386
	Confidence	2.18	34.2	**0.438**
	Gate3R (ours)	**1.79**	**36.1**	0.435
	Oracle	1.14	46.6	0.499
MapAnything (4v)	None	2.56	29.4	0.324
	All	2.21	24.9	0.301
	Random	2.33	26.1	0.309
	Confidence	2.12	32.0	0.339
	Gate3R (ours)	**1.76**	**33.1**	**0.355**
	Oracle	1.17	43.4	0.415

Table 3: **Gating with multiple priors.** Each column lists the priors fed together, such as KP for intrinsics and pose, and each fed prior is independently clean or corrupted. Oracle picks the best subset of the fed priors for each view set. Means over the five test datasets and ten seeds. Bold marks the best result per model, excluding Oracle.

		K: ray err (deg) ↓			P: AUC@5° ↑			D: inlier δ<1.03 ↑		
Model	Priors used	KP	KD	KPD	KP	PD	KPD	KD	PD	KPD
Pow3R (2v)	None	1.65	1.65	1.65	32.2	32.2	32.2	0.409	0.409	0.409
	All	2.10	2.19	2.15	20.7	23.1	20.1	0.501	0.426	0.362
	Random	2.01	1.94	1.98	24.1	25.5	23.3	0.493	0.433	0.403
	Confidence	1.81	1.71	1.81	28.0	30.9	27.9	0.539	0.489	0.443
	Gate3R (ours)	**1.53**	**1.33**	**1.40**	**32.5**	**33.0**	**33.4**	**0.562**	**0.515**	**0.506**
	Oracle	0.79	0.68	0.61	52.6	53.4	63.2	0.740	0.703	0.790
MapAnything (2v)	None	2.19	2.19	2.19	34.1	34.1	34.1	0.394	0.394	0.394
	All	2.25	2.27	2.27	27.1	29.4	30.5	0.451	0.339	0.353
	Random	2.22	2.27	2.29	28.4	29.6	29.3	0.417	0.342	0.355
	Confidence	2.25	2.02	2.26	27.1	32.5	32.7	0.496	0.388	0.417
	Gate3R (ours)	**1.70**	**1.68**	**1.69**	**37.3**	**37.5**	**38.6**	**0.517**	**0.414**	**0.462**
	Oracle	0.98	0.98	0.92	55.8	56.6	64.2	0.654	0.572	0.699
MapAnything (4v)	None	2.16	2.16	2.16	34.0	34.0	34.0	0.340	0.340	0.340
	All	2.20	2.21	2.31	24.3	27.7	27.4	0.358	0.287	0.309
	Random	2.20	2.23	2.32	26.0	27.6	26.5	0.344	0.285	0.303
	Confidence	2.20	2.32	2.22	24.3	30.9	30.9	0.368	0.339	0.363
	Gate3R (ours)	**1.64**	**1.64**	**1.71**	**34.3**	**35.4**	**35.3**	**0.436**	**0.367**	**0.415**
	Oracle	0.97	0.98	0.93	50.5	52.4	55.9	0.561	0.496	0.616

MapAnything depth on DTU. On the two models we study, the echo follows the injection mechanism rather than the prior type. MapAnything reproduces 92% of a pose error but 79% of an intrinsics error, whereas Pow3R reverses the order, 66% against 85%, and the echo barely changes from two to four views. Pow3R depth further shows that a mechanism can hide part of an error from the network. Its per-view mean normalization cancels a global depth rescaling, so only a small part of

Table 4: **Echo per dataset.** Echo in percent at the severest level of each prior, read on the rotation error for pose. TNT denotes Tanks and Temples, whose sparse ground-truth depth gives no depth entries.

Model	12-Scenes			7-Scenes			DTU			ETH3D			LaMAR			TNT		Mean		
	K	P	D	K	P	D	K	P	D	K	P	D	K	P	D	K	P	K	P	D
Pow3R (2v)	90	73	93	88	68	93	92	75	89	80	60	90	84	65	95	77	55	85	66	92
MapAnything (2v)	85	93	85	84	97	82	57	92	46	83	91	80	80	92	92	82	90	79	92	77
MapAnything (4v)	84	94	80	83	97	81	55	92	57	83	89	77	78	92	86	81	91	77	93	76

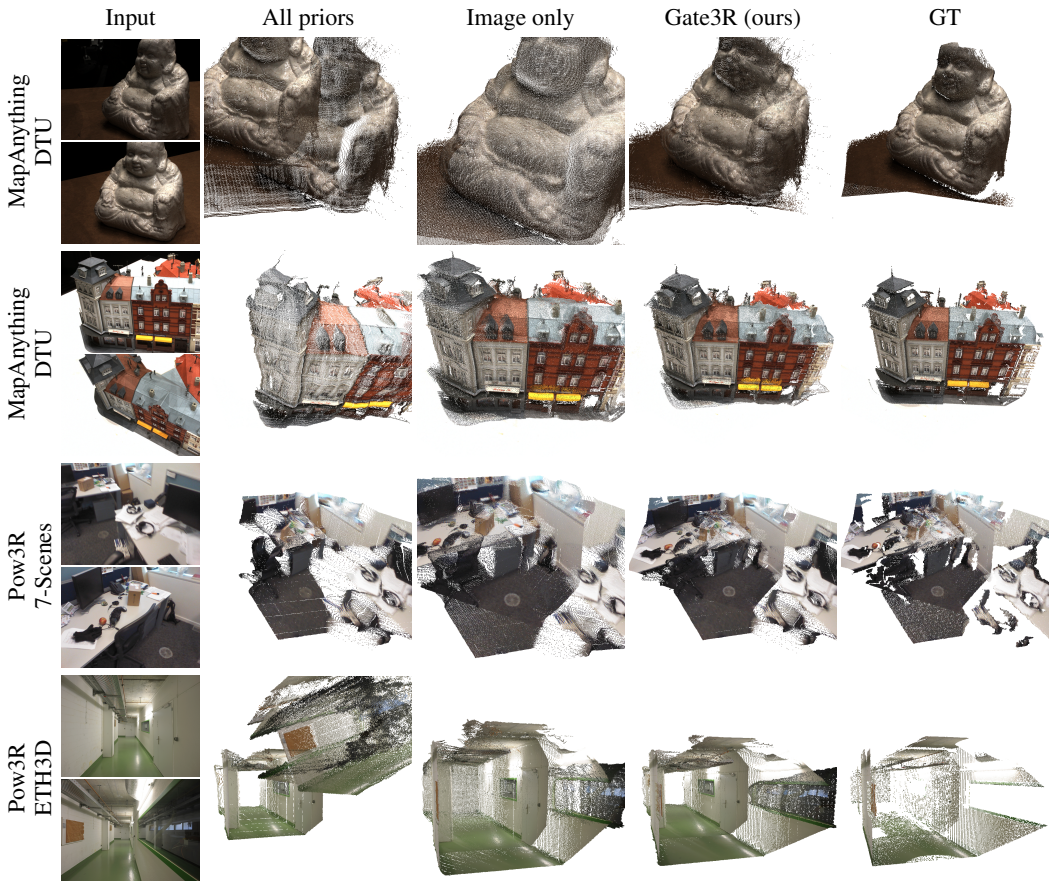

Figure 4: **Qualitative results.** Two-view reconstructions with all priors, with images alone, and with Gate3R. GT denotes the ground truth.

the corruption reaches the network, which is consistent with the Pow3R depth exception in the main results. A high echo thus weakens the residual, and the entropy is needed as a second readout.

Attention entropy. The entropy registers a corrupted prior inside the network even when the output copies it. In Figure 3, the entropy map of the selected MapAnything head changes on every view that receives a corrupted prior. At the severest level, the entropy alone separates clean from corrupted priors with an AUC of 0.64 to 0.98 on seven of the nine axes and is close to chance only on MapAnything intrinsics, as Table 7 reports. It is weaker than the residual, yet fusing the two raises the AUC over the residual alone on seven of the nine axes, by up to 0.026 on Pow3R intrinsics.

Ablation study. Fusing the two readouts gives the most consistent gate score. Table 5 replaces the fused score with the residual alone, the entropy alone, and the larger of the two readouts after standardizing each by clean statistics. Our fusion is best or tied on seven of the nine axes, and it

Table 5: **Gate score ablation.** Each row replaces the fused score of Gate3R and keeps all other settings fixed. Single-prior means over the five test datasets and ten seeds. Bold marks the best score per column, excluding the fixed policies.

		K: ray err (deg) ↓			P: AUC@5° ↑			D: inlier δ<1.03 ↑		
	Variant	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v
Fixed	None	1.80	2.59	2.56	30.0	29.0	29.4	0.400	0.375	0.324
	All	2.07	2.20	2.21	24.3	28.4	24.9	0.590	0.393	0.301
Gate score	Residual only	1.63	1.80	1.77	**32.3**	35.9	**33.1**	**0.561**	0.433	0.350
	Entropy only	1.62	2.30	2.29	31.9	32.8	31.3	0.553	0.417	0.337
	Max, clean z	1.61	1.85	1.79	**32.3**	35.6	**33.1**	0.554	0.433	0.350
Gate3R (ours)		**1.54**	**1.79**	**1.76**	32.1	**36.1**	**33.1**	0.558	**0.435**	**0.355**

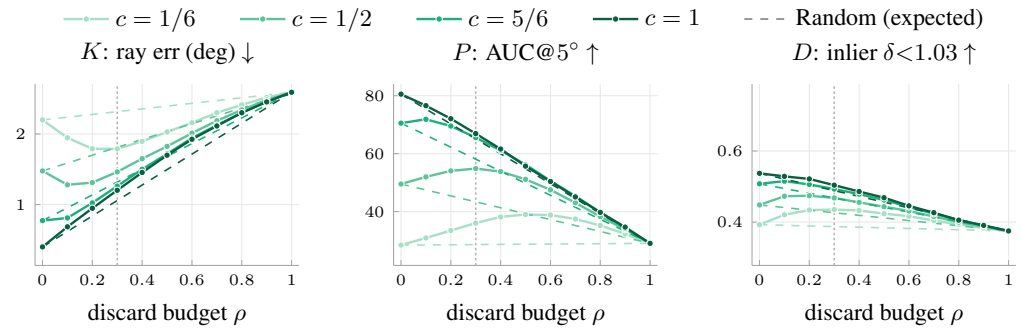

Figure 5: **Discard budget.** Single-prior results of Gate3R on MapAnything with two views as the discard budget ρ grows from 0 (All) to 1 (None), for clean fractions c of the batch. Dashed lines join All and None and give the expected result of discarding at random. The dotted line marks $\rho = 0.3$. Depth excludes Tanks and Temples.

gains most on Pow3R intrinsics, $1.54°$ against $1.63°$ for the residual alone, where both readouts are informative. The clean-statistics maximum needs clean data per model, prior, and view count, yet exceeds our fusion only on Pow3R pose, by 0.2 AUC. Section D further shows that a single head suffices, and that the choice of held-out dataset has little effect.

Discard budget. Figure 5 traces Gate3R from $\rho = 0$ to $\rho = 1$ for clean fractions c from $1/6$ to 1. Whenever some priors are corrupted, the curve lies on the better side of the random line at every budget on pose and depth, and the best budget shrinks as clean priors become common, from 0.5 at $c = 1/6$ to 0.1 at $c = 5/6$ on pose. When every prior is clean, the intrinsics curve falls below the line, since the gate can then only discard clean priors. With multiple priors, every budget from 0.3 to 0.5 beats both fixed policies on all 27 cells. As c is unknown at deployment, we keep $\rho = 0.3$ fixed.

Qualitative results. Figure 4 compares the reconstructions with all priors, with images alone, and with Gate3R on four two-view scenes, two for each model.

5 Conclusion

We studied when a prior-conditioned feed-forward reconstruction model should trust the geometric priors it receives. Although a frozen model echoes most of a wrong prior into its output, it still registers the error within the same pass, both in the prior residual and in the attention entropy. Gate3R fuses these two readouts to discard unreliable priors at test time without training, and it outperforms both fixed policies in most single-prior settings and in all multi-prior settings. Gate3R assumes that the model also supports image-only inference, and our study covers two such models. A clear gap to the Oracle remains, and per-view decisions and further prior-conditioned models are natural extensions.

AI USE STATEMENT

Large language models were used as a writing and engineering assistant under author direction: drafting and revising manuscript text, editing analysis and plotting scripts, and preparing experiment launch scripts. All experimental design decisions, all reported numbers, and all claims were produced by the authors' experiment pipeline and verified by the authors against the underlying run logs; no experimental result was generated by an LLM.

REFERENCES

Jelena Bratulić, Sudhanshu Mittal, Thomas Brox, and Christian Rupprecht. On geometric understanding and learned priors in feed-forward 3D reconstruction models. In *European Conference on Computer Vision (ECCV)*, 2026.

Xingyu Chen, Yue Chen, Yuliang Xiu, Andreas Geiger, and Anpei Chen. Easi3R: Estimating disentangled motion from DUSt3R without training. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9158–9168, 2025.

Yung-Sung Chuang, Linlu Qiu, Cheng-Yu Hsieh, Ranjay Krishna, Yoon Kim, and James R. Glass. Lookback lens: Detecting and mitigating contextual hallucinations in large language models using only attention maps. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1419–1436, 2024.

Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):611–625, 2018.

Xuanyu Ge, Zhongqi Wang, Jie Zhang, Shiguang Shan, and Xilin Chen. EntropyScan: Towards model-level backdoor detection in LVLMs via visual attention entropy, 2026. arXiv preprint arXiv:2605.15711.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Jisang Han, Sunghwan Hong, Jaewoo Jung, Wooseok Jang, Honggyu An, Qianqian Wang, Seungryong Kim, and Chen Feng. Emergent outlier view rejection in visual geometry grounded transformers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 427–437, 2026.

Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, and Nenghai Yu. OPERA: Alleviating hallucination in multi-modal large language models via over-trust penalty and retrospection-allocation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Wonbong Jang, Philippe Weinzaepfel, Vincent Leroy, Lourdes Agapito, and Jérôme Revaud. Pow3R: Empowering unconstrained 3D reconstruction with camera and scene priors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1071–1081, 2025.

Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 406–413, 2014.

Linyi Jin, Jianming Zhang, Yannick Hold-Geoffroy, Oliver Wang, Kevin Blackburn-Matzen, Matthew Sticha, and David F. Fouhey. Perspective fields for single image camera calibration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17307–17316, 2023.

Nikhil Keetha, Norman Müller, Johannes Schönberger, Lorenzo Porzi, Yuchen Zhang, Tobias Fischer, Arno Knapitsch, Duncan Zauss, Ethan Weber, Nelson Antunes, Jonathon Luiten, Manuel Lopez-Antequera, Samuel Rota Bulò, Christian Richardt, Deva Ramanan, Sebastian Scherer, and Peter Kontschieder. MapAnything: Universal feed-forward metric 3D reconstruction. In *International Conference on 3D Vision (3DV)*, 2026.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017.

Yifan Liu, Zhiyuan Min, Zhenwei Wang, Junta Wu, Tengfei Wang, Yixuan Yuan, Yawei Luo, and Chunchao Guo. WorldMirror: Universal 3D world reconstruction with any-prior prompting. In *International Conference on Machine Learning (ICML)*, 2026.

Potsawee Manakul, Adian Liusie, and Mark Gales. SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 9004–9017, 2023.

Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.

René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1623–1637, 2022.

Paul-Edouard Sarlin, Mihai Dusmanu, Johannes L. Schönberger, Pablo Speciale, Lukas Gruber, Viktor Larsson, Ondrej Miksik, and Marc Pollefeys. LaMAR: Benchmarking localization and mapping for augmented reality. In *European Conference on Computer Vision (ECCV)*, 2022.

Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew Fitzgibbon. Scene coordinate regression forests for camera relocalization in RGB-D images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2930–2937, 2013.

Jiale Song, Jiaxin Luo, Xue-song Tang, Kuangrong Hao, and Mingbo Zhao. Segmentation-based attention entropy: Detecting and mitigating object hallucinations in large vision-language models. In *ACM International Conference on Multimedia (MM)*, 2026. Accepted; arXiv:2603.16558.

Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Zachary Teed and Jia Deng. DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 16558–16569, 2021.

Julien Valentin, Angela Dai, Matthias Nießner, Pushmeet Kohli, Philip Torr, Shahram Izadi, and Cem Keskin. Learning to navigate the energy landscape. In *International Conference on 3D Vision (3DV)*, pp. 323–332, 2016.

Alexander Veicht, Paul-Edouard Sarlin, Philipp Lindenberger, and Marc Pollefeys. GeoCalib: Learning single-image calibration with geometric optimization. In *European Conference on Computer Vision (ECCV)*, 2024.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual geometry grounded transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306, 2025a.

Jianyuan Wang, Minghao Chen, Shangzhan Zhang, Nikita Karaev, Johannes Schönberger, Patrick Labatut, Piotr Bojanowski, David Novotny, Andrea Vedaldi, and Christian Rupprecht. VGGT-Ω. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026a.

Qianqian Wang, Yifei Zhang, Aleksander Holynski, Alexei A. Efros, and Angjoo Kanazawa. Continuous 3D perception model with persistent state. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10510–10522, 2025b.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jérôme Revaud. DUSt3R: Geometric 3D vision made easy. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709, 2024.

Yifan Wang, Jianjun Zhou, Haoyi Zhu, Wenzheng Chang, Yang Zhou, Zizun Li, Junyi Chen, Jiangmiao Pang, Chunhua Shen, and Tong He. π^3: Permutation-equivariant visual geometry learning. In *International Conference on Learning Representations (ICLR)*, 2026b.

Jian Xie, Kai Zhang, Jiangjie Chen, Renze Lou, and Yu Su. Adaptive chameleon or stubborn sloth: Revealing the behavior of large language models in knowledge conflicts. In *International Conference on Learning Representations (ICLR)*, 2024.

Heng Yang and Marco Pavone. Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8947–8958, 2023.

Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10371–10381, 2024.

Amir R. Zamir, Alexander Sax, Nikhil Cheerla, Rohan Suri, Zhangjie Cao, Jitendra Malik, and Leonidas J. Guibas. Robust learning through cross-task consistency. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11197–11206, 2020.

Lei Zhou, Haoyu Wu, Akshat Dave, and Dimitris Samaras. Learning 3D reconstruction with priors in test time. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 36550–36560, 2026.

Zihao Zhu, Wenyuan Zhao, Nuo Chen, Chao Tian, and Zhiwen Fan. Trust it or not: Evidential uncertainty for feed-forward 3D reconstruction with Trust3R. In *International Conference on Machine Learning (ICML)*, 2026.

Filippo Ziliotto, Luciano Serafini, Lamberto Ballan, and Tommaso Campari. What VGGT knows about overlap: Probing geometric foundation models for co-visibility. *arXiv preprint arXiv:2607.09503*, 2026.

A Experimental Details

This appendix details the protocol shared by all tables and then reports the echo and detection results, the per-dataset gating results, further ablations, and the deployed attention heads.

Protocol. Only the entropy head is selected offline, on the 90 two-view LaMAR sets of each model and prior; the robust z-scores, the tail weights, and the ranking that selects the sets to re-run are all computed on the deployment batch. The five test datasets contribute 640 view sets, of which 430 enter the depth metric. We call the largest corruption level severe and abbreviate the datasets as 12Sc for 12-Scenes, 7Sc for 7-Scenes, ETH for ETH3D, LaM for LaMAR, and TnT for Tanks and Temples. Their scene types and ground-truth sources are listed in Table 6.

Four views. The four-view results use the same sets with four views each, on MapAnything. The intrinsics and depth corruptions are drawn independently per view. The pose corruption is applied to every non-reference view with the same angle but independent axes and directions, while the reference view fixes the gauge and stays clean. The residual is averaged over views for K and D and taken as the maximum over view pairs for P, and discarding a prior removes it from all views. The heads are the two-view ones of Table 13, reused unchanged, and the fusion statistics come from the four-view deployment batch itself.

Dataset caveats. Two datasets need care. The depth prior of Tanks and Temples comes from sparse COLMAP tracks and is 99.9% holes, so its depth axis carries little signal for any model, and we exclude it from the depth metric. On DTU, the narrow field of view maps the severe intrinsics level $a = 0.4$ to only 4.2° of injected ray error, against 7 to 9° elsewhere, while MapAnything's image-only intrinsics are poor at 5.2°. Its K prior on DTU therefore helps at every level.

Table 6: **Evaluation datasets.** Scene type and ground-truth source of ETH3D (Schöps et al., 2017), DTU (Jensen et al., 2014), Tanks and Temples (Knapitsch et al., 2017), LaMAR (Sarlin et al., 2022), 7-Scenes (Shotton et al., 2013), and 12-Scenes (Valentin et al., 2016). 7-Scenes and 12-Scenes rely on pseudo ground truth from RGB-D SLAM.

Dataset	Scene type	GT pose	GT depth
ETH3D	indoor + outdoor	SfM + laser scan	laser scan
DTU	object-centric, lab	robot-arm calibration	structured-light scan
Tanks and Temples	indoor + outdoor, large-scale	SfM + laser scan	sparse SfM (not used)
LaMAR	indoor + outdoor, AR	lidar-registered pipeline	lidar mesh
7-Scenes	indoor, small rooms	RGB-D SLAM	Kinect sensor
12-Scenes	indoor, rooms	RGB-D SLAM	Structure sensor

B Echo and Detection

B.1 Echo

For a given model and prior, the echo is stable across datasets, including the held-out LaMAR. MapAnything reproduces 90 to 97% of a severe pose error on every dataset, and Pow3R 55 to 75%. DTU is the outlier for MapAnything intrinsics and depth, at 57% and 46%, coinciding with its narrow field of view and small injected error. Pow3R depth echoes 89 to 95% on every dataset in its scale-invariant depth metric.

B.2 Detection

Table 7 reports how well each signal separates clean from corrupted priors within a test dataset. At the severe level, the residual alone reaches an AUC of 0.86 to 0.99 on intrinsics and pose and 0.70 to 0.87 on depth. The fusion also raises Pow3R intrinsics from 0.859 to 0.885, and on the two MapAnything intrinsics axes, where the entropy head carries little information, it stays within 0.01 of the residual. The clean-reference maximum instead falls below the residual on three axes, since an uninformative readout can raise the score through its own noise. Averaged over all five levels,

every AUC is lower, because the mild levels barely perturb the prior and are hard to separate from clean.

Table 7: **Detection AUC.** Separation of clean and corrupted priors, averaged over the five test datasets, at the severest level and over all five levels. Max takes the larger of s_{res} and s_{ent} after standardizing each by its LaMAR clean statistics. Bold marks the best signal per column and block. Depth excludes Tanks and Temples.

		Pow3R (2v)			MapAny. (2v)			MapAny. (4v)		
	Signal	K	P	D	K	P	D	K	P	D
Severe	s_{res}	0.859	0.993	0.700	**0.903**	0.956	0.775	**0.931**	0.980	0.866
	s_{ent}	0.831	0.984	0.642	0.586	0.807	0.754	0.623	0.901	0.863
	max	0.839	0.993	0.656	0.901	0.959	0.781	**0.931**	0.980	0.867
	s_{fus}	**0.885**	**0.996**	**0.724**	0.901	**0.961**	**0.790**	0.921	**0.981**	**0.881**
All levels	s_{res}	0.681	**0.840**	0.567	**0.683**	0.769	0.626	**0.701**	0.800	0.656
	s_{ent}	0.616	0.796	0.541	0.536	0.668	0.595	0.547	0.740	0.640
	max	0.622	0.837	0.548	0.660	0.757	0.624	0.679	0.794	0.654
	s_{fus}	**0.693**	0.833	**0.574**	0.674	**0.773**	**0.632**	0.683	**0.806**	**0.666**

C GATING

Per dataset. We break the single-prior results down by dataset in Table 8. Gate3R is never below both fixed policies on any of the 42 dataset cells, and it is the best of the three on 24 of them and tied on one more. Where it is not the best, one fixed policy is far ahead for a reason specific to the dataset. On ETH3D and Tanks and Temples, the image-only predictions are often accurate enough that discarding every prior wins, *e.g.* 53.8 against 47.2 for MapAnything pose on ETH3D. On DTU, MapAnything's image-only intrinsics are poor, so keeping every intrinsics prior wins, 1.60° against 3.12° for Gate3R. Pow3R depth is the one axis where keeping every prior is best on three of the four datasets, and this axis is the only single-prior loss of Gate3R.

Per noise level. We split the single-prior results by noise level in Figure 6. On clean and mildly corrupted sets, where the prior helps, Gate3R follows keeping every prior; at the severest level, where the prior hurts, it approaches discarding every prior, *e.g.* 26.9 against 29.1 for None and 0.2 for All on Pow3R pose at 90°. The loss concentrates at intermediate levels. At 15° pose noise, keeping the prior already drives AUC@5° to near zero, yet only 15 to 28% of these sets are discarded, so Gate3R reaches 3.8 to 7.6 against about 29 for None. The fixed budget also discards some clean priors, and this costs most on Pow3R depth, 0.683 against 0.756 for All.

Table 8: **Gating results per test dataset.** Bold marks the best of All, None, and Gate3R on each dataset. Depth excludes TnT, whose ground-truth depth is sparse.

		K: ray err (deg) ↓						P: AUC@5° ↑						D: inlier δ<1.03 ↑				
Model	Policy	12Sc	7Sc	DTU	ETH	TnT	Mean	12Sc	7Sc	DTU	ETH	TnT	Mean	12Sc	7Sc	DTU	ETH	Mean
Pow3R (2v)	All	2.21	2.04	**1.29**	2.51	2.29	2.07	21.8	25.9	19.2	24.0	30.4	24.3	0.645	**0.596**	**0.624**	**0.496**	0.590
	None	**0.78**	1.75	3.75	1.84	**0.91**	1.80	25.2	31.1	14.4	**35.7**	**43.7**	30.0	0.497	0.322	0.377	0.401	0.400
	Gate3R (ours)	1.11	**1.46**	2.41	**1.63**	1.11	1.54	**28.4**	**34.0**	**21.6**	32.8	43.6	32.1	**0.649**	0.534	0.555	**0.496**	0.558
	Oracle	0.55	1.01	1.05	1.07	0.72	0.88	38.3	43.9	28.9	43.5	55.1	41.9	0.720	0.623	0.653	0.583	0.645
MapAny. (2v)	All	2.27	2.12	**1.60**	2.62	2.38	2.20	28.1	30.6	19.5	32.3	31.6	28.4	0.481	**0.342**	0.279	0.468	0.393
	None	2.21	3.59	5.17	**1.25**	**0.74**	2.59	13.8	15.4	9.8	**53.8**	**52.3**	29.0	0.416	0.248	0.338	0.498	0.375
	Gate3R (ours)	**1.58**	**1.95**	3.12	1.27	1.02	1.79	**30.0**	**32.9**	**22.9**	47.2	47.2	36.1	**0.520**	0.333	**0.367**	**0.521**	0.435
	Oracle	1.24	1.54	1.53	0.84	0.55	1.14	36.2	39.7	26.3	66.2	64.6	46.6	0.570	0.383	0.425	0.617	0.499
MapAny. (4v)	All	2.37	2.26	**1.60**	2.61	2.23	2.21	23.1	24.9	15.4	29.9	31.3	24.9	0.320	0.266	0.222	0.397	0.301
	None	2.17	3.51	5.11	1.34	**0.66**	2.56	16.0	18.0	12.9	**48.3**	**52.0**	29.4	0.358	0.217	0.279	0.441	0.324
	Gate3R (ours)	**1.58**	**1.92**	3.16	**1.25**	0.88	1.76	**27.2**	**29.2**	**17.6**	44.5	47.2	33.1	**0.385**	**0.274**	**0.291**	**0.468**	0.355
	Oracle	1.26	1.70	1.54	0.88	0.50	1.17	33.1	36.1	23.7	60.2	63.9	43.4	0.450	0.323	0.349	0.539	0.415

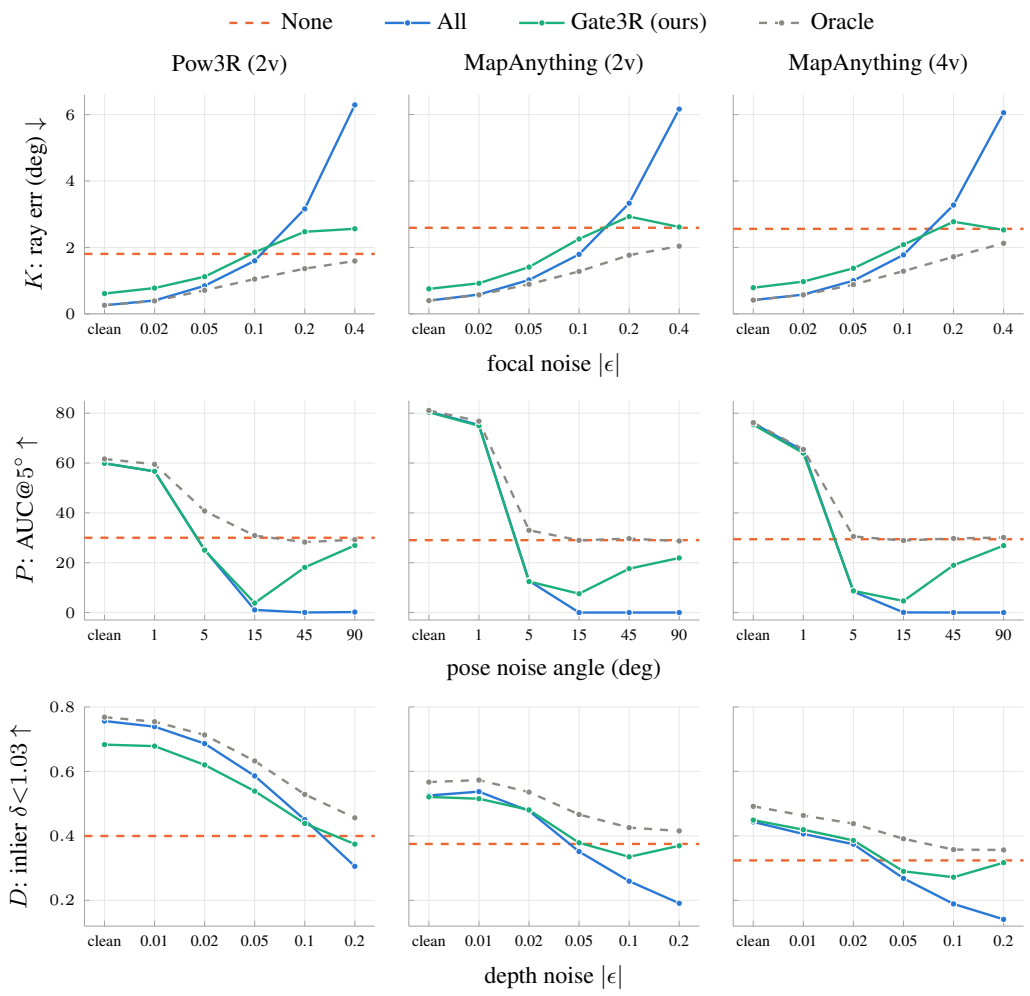

Figure 6: **Single-prior gating by noise level.** Means over the five test datasets and ten seeds. None ignores the prior and is drawn at its mean. Pose noise turns both the rotation and the translation direction by the given angle.

D Ablations

This section sweeps the budget ρ and the number of entropy heads, and varies the dataset on which the heads are selected. Single-prior results are given in Figure 7 and Table 10 and the 27 multi-prior cells in Table 11. All other settings follow the main experiments.

Discard budget. Figure 7 traces Gate3R as the budget grows from $\rho = 0$, which keeps every prior, to $\rho = 1$, which discards every prior, for clean fractions c from $1/6$ to 1. Discarding at random follows the dashed line between the two ends in expectation. Whenever some priors are corrupted, the curve lies on the better side of this line at every budget on pose and depth, and at all but one budget up to $c = 1/2$ on intrinsics. The best budget shrinks as clean priors become common, *e.g.* on MapAnything pose from 0.5 at $c = 1/6$ to 0.3 at $c = 1/2$ and 0.1 at $c = 5/6$. When every prior is clean, the ranking carries no corruption signal. The pose and depth curves then stay close to or above the line, whereas the intrinsics curves fall below it: a clean intrinsics prior has the highest residual on the view sets where the image-only intrinsics are worst, so the gate discards the clean priors that help most. With multiple priors the trends are the same. Every budget from 0.3 to 0.5 beats both fixed policies on all 27 cells, whereas $\rho = 0.2$ falls below discarding every prior on 7 pose cells and $\rho = 0.1$ fails on 14 cells, since a small budget keeps too many badly wrong priors.

Since c is unknown at deployment, we treat ρ as a compute budget rather than tune it to c, and fix it to 0.3.

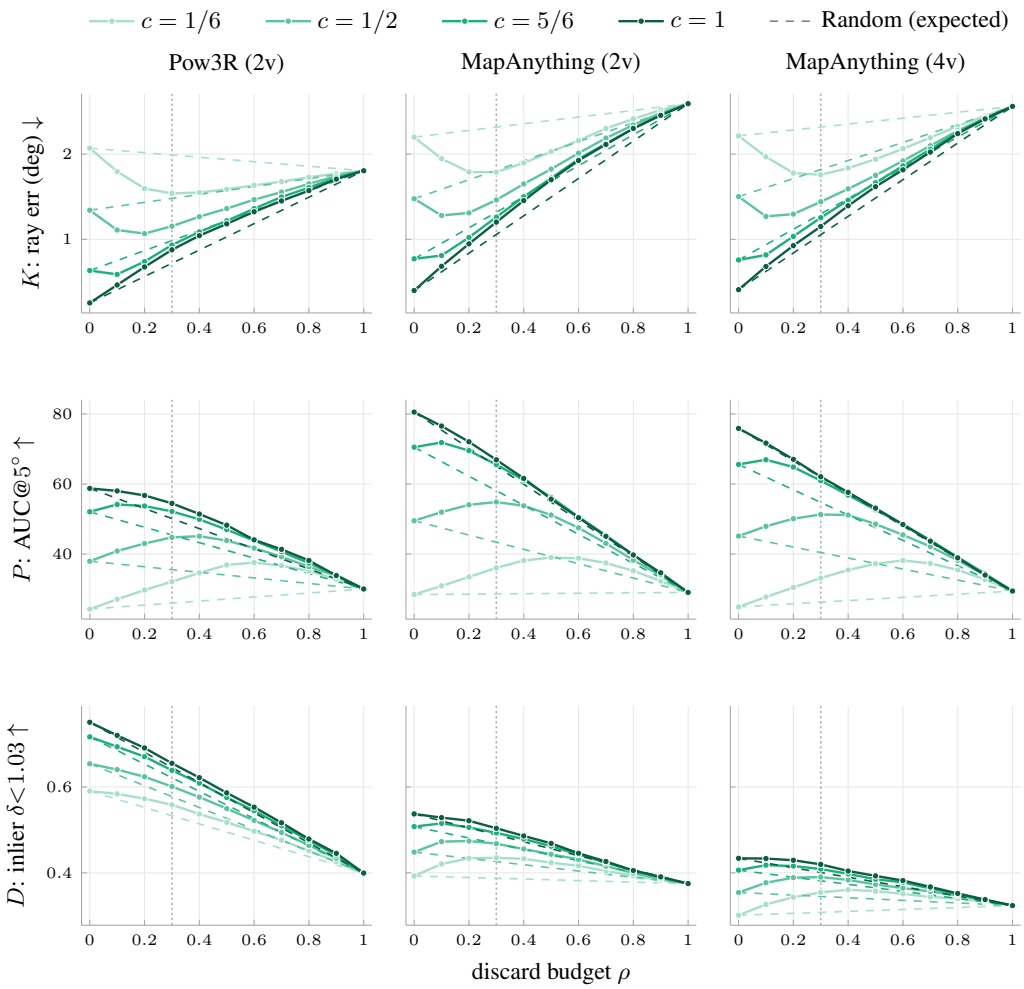

Figure 7: **Discard budget.** Single-prior results of Gate3R as the discard budget ρ grows from 0 (All) to 1 (None), for clean fractions c of the batch. Dashed lines join All and None and give the expected result of discarding at random. The dotted line marks $\rho = 0.3$. Depth excludes Tanks and Temples.

Corruption rate. All results above use a deployment batch in which $1/6$ of the view sets carry a clean prior. In Table 9, we raise this clean fraction c to $1/3$ and $1/2$ and spread the remaining sets uniformly over the five levels, reusing the same per-set outputs, heads, and test sets. At every c, Gate3R with $\rho = 0.3$ beats both fixed policies on all intrinsics and pose axes and on both MapAnything depth axes, *e.g.* 46.5 against 39.4 AUC on MapAnything pose at $c = 1/3$, and Pow3R depth stays the one loss. As clean priors become more common, keeping them all improves, and on MapAnything intrinsics the margin narrows from 0.41° at $c = 1/6$ to 0.01° at $c = 1/2$. The best budget shrinks accordingly: on pose it moves from 0.5 at $c = 1/6$ and $1/3$ to 0.3 at $c = 1/2$, and on intrinsics from 0.3 to 0.1 at $c = 1/2$.

Number of heads. Averaging the k top-ranked heads strengthens the entropy alone, *e.g.* its severe AUC on MapAnything pose rises from 0.81 with one head to 0.88 with four and 0.93 with eight, and its gating result from 32.8 to 35.4 AUC. Four heads, however, move the fused gate by at most 0.04°, 0.1 AUC, and 0.006 on any single-prior axis, since the residual already detects most of what the extra heads add. Pow3R intrinsics is the one axis where four heads help the fused gate,

$1.51°$ against $1.54°$, and with multiple priors Pow3R also gains up to 0.7 pose AUC, whereas on MapAnything one head is best or tied on all 18 multi-prior cells. We therefore keep a single head, which also keeps the gate to one entropy readout per prior.

Head-selection dataset. In Table 12, the heads are selected on each other dataset instead of LaMAR. Many sources select the same head, *e.g.* Pow3R picks the same pose head on four of the five sources, and the gate moves little. The one sensitive axis is Pow3R intrinsics, where the DTU head gives $1.43°$ against $1.32°$ for the LaMAR head. Over all sources, 37 of the 42 single-prior cells and 121 of the 135 multi-prior cells beat both fixed policies, exactly as many as with the LaMAR head on the same datasets.

Table 9: **Corruption rate.** Single-prior results when a fraction c of each batch carries the clean prior and the rest is spread uniformly over the five corruption levels; the main results use $c = 1/6$. Depth excludes Tanks and Temples. Bold marks the best non-oracle policy per column within each block.

	K: ray err (deg) ↓			P: AUC@5° ↑			D: inlier δ<1.03 ↑		
Policy	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v
Clean fraction $c = 1/6$									
None	1.80	2.59	2.56	30.0	29.0	29.4	0.400	0.375	0.324
All	2.07	2.20	2.21	24.3	28.4	24.9	**0.590**	0.393	0.301
$\rho = 0.1$	1.80	1.95	1.97	27.1	30.9	27.7	0.584	0.421	0.327
$\rho = 0.3$	**1.54**	**1.79**	**1.76**	32.1	36.1	33.1	0.558	**0.435**	0.355
$\rho = 0.5$	1.59	2.03	1.94	**36.9**	**39.0**	**37.2**	0.518	0.424	**0.357**
Oracle	0.88	1.14	1.17	41.9	46.6	43.4	0.645	0.499	0.415
Clean fraction $c = 1/3$									
None	1.80	2.59	2.56	30.0	29.0	29.4	0.400	0.375	0.324
All	1.70	1.83	1.84	31.4	39.4	35.2	**0.623**	0.420	0.329
$\rho = 0.1$	1.44	**1.60**	1.59	34.2	41.9	38.1	0.614	0.447	0.353
$\rho = 0.3$	**1.34**	**1.60**	**1.58**	39.1	46.5	42.8	0.581	**0.453**	**0.373**
$\rho = 0.5$	1.46	1.91	1.84	**41.5**	**46.8**	**44.3**	0.533	0.436	0.366
Oracle	0.75	0.99	1.02	45.7	54.0	50.1	0.669	0.513	0.430
Clean fraction $c = 1/2$									
None	1.80	2.59	2.56	30.0	29.0	29.4	0.400	0.375	0.324
All	1.34	1.48	1.50	38.0	49.5	45.1	**0.654**	0.448	0.354
$\rho = 0.1$	**1.11**	**1.28**	**1.27**	40.9	52.0	47.9	0.640	**0.473**	0.377
$\rho = 0.3$	1.15	1.46	1.44	**44.8**	**54.8**	**51.3**	0.601	0.468	**0.390**
$\rho = 0.5$	1.36	1.82	1.75	43.8	51.1	48.6	0.550	0.443	0.373
Oracle	0.62	0.84	0.88	49.3	60.8	56.4	0.693	0.529	0.444

Table 10: **Number of entropy heads.** The entropy readout averages the k top-ranked LaMAR heads, each standardized by its LaMAR clean statistics. Top: the entropy alone. Bottom: our fusion. Bold marks the best k per column within each block.

	K: ray err (deg) ↓			P: AUC@5° ↑			D: inlier δ<1.03 ↑		
Heads	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v
Entropy only									
$k = 1$	1.62	2.30	**2.29**	31.9	32.8	31.3	**0.553**	**0.417**	**0.337**
$k = 2$	1.61	**2.29**	2.33	32.4	33.6	32.1	0.548	**0.417**	0.328
$k = 4$	**1.52**	**2.29**	2.31	31.9	34.9	32.3	0.550	0.415	0.326
$k = 8$	1.55	**2.29**	2.31	**32.6**	**35.4**	**33.3**	0.549	**0.417**	0.324
Gate3R									
$k = 1$	1.54	**1.79**	**1.76**	32.1	36.1	**33.1**	**0.558**	0.435	**0.355**
$k = 2$	1.55	**1.79**	1.77	32.2	36.0	**33.1**	**0.558**	**0.436**	0.352
$k = 4$	**1.51**	**1.79**	**1.76**	32.2	36.0	**33.1**	0.552	0.434	0.349
$k = 8$	1.52	**1.79**	**1.76**	**32.3**	**36.2**	**33.1**	0.557	0.434	0.347

Table 11: **Budget and heads with multiple priors.** Each row is one metric of one prior combination. Left: the budget ρ with one head. Right: the number of entropy heads k at $\rho = 0.3$. Bold marks the best value per row within each block.

					Budget ρ					Heads k			
Model	Priors	Metric	None	All	0.1	0.2	0.3	0.4	0.5	1	2	4	8
Pow3R	K+P	K	1.65	2.10	1.83	1.64	1.52	**1.48**	**1.48**	1.52	1.54	**1.48**	1.50
	K+P	P	32.2	20.7	25.6	29.6	32.5	35.2	**37.2**	32.5	33.0	**33.1**	**33.1**
	K+D	K	1.65	2.19	1.80	1.50	1.33	**1.31**	1.34	1.33	1.35	**1.32**	1.34
	K+D	D	0.409	0.501	0.539	0.556	**0.562**	0.551	0.536	**0.562**	0.561	**0.562**	0.561
	P+D	P	32.2	23.1	26.6	30.0	33.0	35.9	**38.8**	33.0	32.9	**33.1**	**33.1**
	P+D	D	0.409	0.426	0.469	0.500	0.515	0.526	**0.531**	0.515	**0.518**	0.516	0.516
	K+P+D	K	1.65	2.15	1.80	1.55	1.40	**1.33**	1.35	1.40	1.40	**1.37**	1.41
	K+P+D	P	32.2	20.1	25.4	29.6	33.4	36.3	**38.4**	33.4	33.5	33.8	**33.9**
	K+P+D	D	0.409	0.362	0.433	0.477	0.506	0.526	**0.533**	0.506	0.509	**0.513**	0.512
MapAny	K+P	K	2.19	2.25	1.94	1.75	**1.70**	1.73	1.81	**1.70**	**1.70**	1.71	1.71
	K+P	P	34.1	27.1	30.4	33.9	37.3	39.8	**41.5**	**37.3**	37.2	**37.3**	**37.3**
	K+D	K	2.19	2.27	1.96	1.75	**1.68**	1.74	1.82	**1.68**	1.69	1.69	**1.68**
	K+D	D	0.394	0.451	0.488	0.510	**0.517**	0.508	0.492	**0.517**	0.516	0.516	**0.517**
	P+D	P	34.1	29.4	32.2	35.0	37.5	39.2	**40.8**	**37.5**	**37.5**	37.3	37.3
	P+D	D	0.394	0.339	0.379	0.398	0.414	**0.419**	0.418	**0.414**	0.413	0.413	0.411
	K+P+D	K	2.19	2.27	1.92	1.72	**1.69**	1.76	1.83	**1.69**	1.71	1.72	1.72
	K+P+D	P	34.1	30.5	33.4	36.5	38.6	40.3	**41.2**	**38.6**	38.5	38.5	38.5
	K+P+D	D	0.394	0.353	0.411	0.445	**0.462**	0.460	0.456	**0.462**	0.458	0.458	0.456
MapAny 4v	K+P	K	2.16	2.20	1.89	1.69	**1.64**	1.69	1.76	**1.64**	1.65	1.65	1.65
	K+P	P	34.0	24.3	27.3	30.7	34.3	37.5	**40.0**	**34.3**	34.2	34.2	34.2
	K+D	K	2.16	2.21	1.90	1.68	**1.64**	1.70	1.77	**1.64**	1.67	1.67	1.65
	K+D	D	0.340	0.358	0.394	0.423	**0.436**	**0.436**	0.428	**0.436**	0.431	0.431	0.432
	P+D	P	34.0	27.7	30.2	32.7	35.4	38.0	**39.8**	**35.4**	**35.4**	35.3	35.3
	P+D	D	0.340	0.287	0.321	0.346	0.367	0.371	**0.375**	**0.367**	0.363	0.358	0.357
	K+P+D	K	2.16	2.31	1.98	1.75	**1.72**	1.75	1.81	**1.72**	1.74	1.74	1.74
	K+P+D	P	34.0	27.4	29.7	32.5	35.3	37.8	**39.8**	**35.3**	**35.3**	**35.3**	**35.3**
	K+P+D	D	0.340	0.309	0.358	0.394	0.415	**0.421**	0.419	**0.415**	0.411	0.409	0.406

Table 12: **Head-selection dataset.** Single-prior results with the heads selected on each listed dataset, shown as source head / LaMAR head. Both are evaluated on the remaining test datasets, and depth also excludes Tanks and Temples. The four-view rows reuse the two-view head.

	K: ray err (deg) ↓			P: AUC@5° ↑			D: inlier δ<1.03 ↑		
Source	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v	Pow3R	MapAny	MapAny 4v
ETH3D	1.52 / 1.52	1.94 / 1.92	1.91 / 1.89	32.2 / 31.9	33.3 / 33.3	30.3 / 30.3	0.584 / 0.579	0.406 / 0.406	0.317 / 0.317
7-Scenes	1.59 / 1.56	1.76 / 1.75	1.72 / 1.72	31.6 / 31.6	36.9 / 36.9	34.2 / 34.1	0.567 / 0.566	0.469 / 0.469	0.378 / 0.381
12-Scenes	1.69 / 1.65	1.83 / 1.84	1.79 / 1.80	33.0 / 33.0	37.4 / 37.6	34.5 / 34.6	0.529 / 0.528	0.407 / 0.407	0.344 / 0.344
DTU	1.43 / 1.32	1.46 / 1.46	1.42 / 1.41	34.7 / 34.7	39.2 / 39.3	37.0 / 37.0	0.563 / 0.560	0.458 / 0.458	0.376 / 0.376
Tanks and Temples	1.65 / 1.65	1.98 / 1.98	1.98 / 1.98	29.2 / 29.2	33.0 / 33.3	29.6 / 29.6	–	–	–

E DEPLOYED HEADS

The deployed heads and their behavior on the test datasets are listed in Table 13. Each is the top-ranked of the 144 later-half heads of MapAnything or the 528 of Pow3R, and beats the median head by 0.05 to 0.24 LaMAR AUC.

Location. Pose and depth heads concentrate late: 8 to 10 of the 10 best heads by LaMAR AUC lie in the later half of the network, and on Pow3R pose all ten lie in its last quarter, seven of them cross-view. On MapAnything the pose and depth heads even share one global-attention block, block 16. MapAnything intrinsics is the exception, since its best heads sit in the first quarter, *e.g.* frame block 1 with an AUC of 0.70. Restricting the search to the later half therefore leaves a weak intrinsics head, 0.55, which the fusion compensates with the residual.

Response to a corrupted prior. In eight of the nine cells a corrupted prior lowers the head's entropy, by 1.5% to 24% of its clean value at the severest level, *i.e.* the head attends more sharply. The shift is largest where the head detects best, on Pow3R pose, and its direction matches LaMAR on every test dataset except DTU for Pow3R intrinsics. Pow3R depth is the one head whose entropy rises. Detection grows with severity: on the test datasets the AUC rises from about 0.5 at the mildest level to between 0.59 and 0.98 at the severest, so the level-averaged AUC understates the head on the harmful levels.

Transfer. The heads are selected on LaMAR alone, yet their severe-level AUC on the test datasets matches or exceeds the LaMAR value on pose and depth. Pow3R intrinsics is the one drop, from 1.00 to 0.83, driven by DTU, which is the weakest target in almost every cell. The two-view heads also transfer to four views without re-selection: their entropy shifts in the same direction on every dataset, rank 1, 2, and 4 among the four-view heads of MapAnything, and detect better than with two views, *e.g.* 0.90 against 0.81 severe AUC on pose.

Complementarity with the residual. The entropy is nearly uncorrelated with the residual on clean sets, with a Spearman correlation between 0.02 and 0.23, and moderately correlated on corrupted sets, between 0.20 and 0.68, since both respond to the same error. The entropy matters most on Pow3R intrinsics and depth, where 28% and 31% of the corrupted sets are discarded only because of it, against at most 6% on MapAnything.

Table 13: **Deployed attention heads.** Each head is selected on two-view LaMAR and reused on every test dataset and on four views. Frame and global denote self and cross attention, L the layer, and H the head. LaMAR and Test give the detection AUC, the latter averaged over levels and at the severest level. ΔH is the severe-level entropy shift, Agree the fraction of test datasets on which the shift has the same direction as on LaMAR, r_s the Spearman correlation with the residual, and Ent. only the fraction of corrupted sets discarded by the entropy alone.

Model	Prior	Head	LaMAR	Test mean / sev.	ΔH (%)	Agree	r_s	Ent. only
MapAnything	K	frame L21 H3	0.55	0.54 / 0.59	−1.5	1.00	0.20	0.04
	P	global L16 H5	0.73	0.67 / 0.81	−7.9	1.00	0.31	0.02
	D	global L16 H6	0.58	0.59 / 0.75	−2.8	1.00	0.33	0.06
MapAnything 4v	K	same	0.54	0.55 / 0.62	−2.1	1.00	0.22	0.02
	P	same	0.81	0.74 / 0.90	−12.6	1.00	0.49	0.00
	D	same	0.59	0.64 / 0.86	−3.7	1.00	0.47	0.05
Pow3R	K	frame L20 H8	0.72	0.62 / 0.83	−12.0	0.80	0.43	0.28
	P	global L21 H2	0.74	0.80 / 0.98	−23.6	1.00	0.68	0.00
	D	frame L34 H4	0.54	0.54 / 0.64	+8.5	1.00	0.21	0.31