Supplementary Information

Limits of continual test-time adaptation for camera–LiDAR 3D detection under simulated view loss

Peiyuan Qu and Mengxing Huang

Supplementary Note: Evaluation and audit scope

Formal S2 is the continuous predict-before-update stream. M0 is the frozen source reference and M5 is the full adapter configuration. **Probe B** evaluates archived predictions from M0 and three frozen final M5 student states (M5-0/1/2) on five scenes, 200 frames and 286 eligible held-out motorcycle/bicycle observations per condition. B is disjoint from S2, source calibration and adaptation development at the scene level; disjointness from all official pretraining scenes was not established. The synthetic loss of 1, 3 or 6 camera views is nested and does not model weather.

Metrics. R_{H}@0.25 is held-out recall with one-to-one matching at 3D IoU 0.25. $\mathrm{mAP}_{\mathrm{cd}}$ is known-class center-distance mean AP. Eligible GT has positive LiDAR point count. FP/frame@0.5 counts unmatched held-out-branch candidates per evaluated frame at 3D IoU 0.5, not road alarms. The default unknown candidate ceiling is $K = 500$ per frame. Recall and mAP are percentages; differences are percentage points (pp).

Supplementary Table S1: Archived S2 cross-check. Stored evaluator JSON and archived metrics CSV agree on frame count, eligible held-out GT count, R_{H}@0.25 and $\mathrm{mAP}_{\mathrm{cd}}$ for all six method–seed rows. No S2 predictions were regenerated or re-evaluated in this check.

Seed	Frames	GT	M0 R_{H}@0.25 (%)	M5 R_{H}@0.25 (%)	M0 $\mathrm{mAP}_{\mathrm{cd}}$ (%)	M5 $\mathrm{mAP}_{\mathrm{cd}}$ (%)
0	1009	639	17.214	16.432	49.901	49.924
1	1008	588	18.537	18.878	49.823	49.892
2	1008	782	17.903	17.647	44.790	44.864

Supplementary Table S2: Probe B: re-evaluation of frozen predictions joined to GT by sample token. All 16 condition–state rows agree with archived B summaries on frame count, held-out GT count, R_{H}@0.25 and $\mathrm{mAP}_{\mathrm{cd}}$. No model updates were performed.

Condition	State	TP@0.25	R_{H}@0.25 (%)	$\mathrm{mAP}_{\mathrm{cd}}$ (%)	FP/frame@0.5
Clean	M0	43	15.035	39.645	247.865
Clean	M5-0	52	18.182	39.017	248.170
Clean	M5-1	49	17.133	39.377	248.415
Clean	M5-2	47	16.434	39.528	247.860
miss1	M0	45	15.734	37.524	254.910
miss1	M5-0	52	18.182	37.294	255.825
miss1	M5-1	47	16.434	37.190	255.805
miss1	M5-2	45	15.734	37.136	254.880
miss3	M0	47	16.434	33.977	269.985
miss3	M5-0	51	17.832	33.818	270.335
miss3	M5-1	49	17.133	33.628	270.400
miss3	M5-2	48	16.783	33.864	269.555
miss6	M0	42	14.685	30.624	297.405
miss6	M5-0	44	15.385	29.996	297.005
miss6	M5-1	40	13.986	29.925	297.600
miss6	M5-2	39	13.636	29.966	296.545

Supplementary Note: Scene-paired probe results (P0)

Within each condition, M0 and M5 predictions are paired on scene and sample token; the camera mask is held fixed for a given frame and condition. In the Clean probe, scene-0911 contributes 5, 3 and 4 additional held-out matches for M5-0/1/2, respectively. M5-2 also loses two matches on scene-0928. The aggregate gain therefore varies by scene and state.

Supplementary Table S3: Probe B per-scene metrics for Clean and miss6. Each state cell lists TP@0.25 / R_{H}@0.25 (%) above $\mathrm{mAP}_{\mathrm{cd}}$ (%) / FP/frame@0.5. Each scene has 40 frames. Recall is undefined (NA) in scene-0556; see Table S5 for its FP counts.

Condition	Scene	GT	M0	M5-0	M5-1	M5-2
			16 / 13.33	18 / 15.00	17 / 14.17	16 / 13.33
Clean	0269	120	47.99 / 212.450	47.36 / 212.050	47.42 / 212.050	48.57 / 211.450
			1 / 3.12	2 / 6.25	1 / 3.12	3 / 9.38
Clean	0275	32	32.76 / 321.850	32.68 / 321.525	32.68 / 322.525	32.59 / 322.375
			0 / NA	0 / NA	0 / NA	0 / NA
Clean	0556	0	33.03 / 228.450	33.08 / 229.125	33.21 / 230.200	33.12 / 228.575
			15 / 26.32	20 / 35.09	18 / 31.58	19 / 33.33
Clean	0911	57	26.35 / 222.175	26.42 / 223.975	26.26 / 223.500	26.04 / 223.225
			11 / 14.29	12 / 15.58	13 / 16.88	9 / 11.69
Clean	0928	77	23.74 / 254.400	23.05 / 254.175	23.42 / 253.800	22.92 / 253.675
			15 / 12.50	18 / 15.00	15 / 12.50	11 / 9.17
miss6	0269	120	36.62 / 272.850	36.69 / 270.550	35.31 / 272.475	36.85 / 271.075
			2 / 6.25	3 / 9.38	2 / 6.25	2 / 6.25
miss6	0275	32	29.25 / 365.075	28.96 / 364.800	29.03 / 367.025	29.06 / 363.300
			0 / NA	0 / NA	0 / NA	0 / NA
miss6	0556	0	26.65 / 264.975	26.97 / 265.175	26.65 / 264.475	26.56 / 266.500
			8 / 14.04	8 / 14.04	10 / 17.54	9 / 15.79
miss6	0911	57	22.33 / 277.300	22.22 / 277.300	22.13 / 277.500	22.11 / 276.325
			17 / 22.08	15 / 19.48	13 / 16.88	17 / 22.08
miss6	0928	77	17.94 / 306.825	17.67 / 307.200	17.78 / 306.525	17.75 / 305.525

Supplementary Table S4: Probe B per-scene metrics for miss1 and miss3. Each state cell lists TP@0.25 / R_{H}@0.25 (%) above $\mathrm{mAP}_{\mathrm{cd}}$ (%) / FP/frame@0.5. Each scene has 40 frames. Recall is undefined (NA) in scene-0556; see Table S5 for its FP counts.

Condition	Scene	GT	M0	M5-0	M5-1	M5-2
			18 / 15.00	20 / 16.67	18 / 15.00	18 / 15.00
miss1	0269	120	44.79 / 222.825	44.55 / 222.875	44.96 / 222.625	45.18 / 220.275
			1 / 3.12	1 / 3.12	0 / 0.00	1 / 3.12
miss1	0275	32	32.40 / 329.125	32.27 / 329.850	32.23 / 329.750	32.16 / 329.000
			0 / NA	0 / NA	0 / NA	0 / NA
miss1	0556	0	31.58 / 232.575	31.94 / 234.225	31.97 / 235.175	31.65 / 233.950
			14 / 24.56	18 / 31.58	16 / 28.07	16 / 28.07
miss1	0911	57	25.20 / 228.775	25.20 / 229.975	25.20 / 230.000	25.12 / 230.275
			12 / 15.58	13 / 16.88	13 / 16.88	10 / 12.99
miss1	0928	77	22.98 / 261.250	22.43 / 262.200	22.63 / 261.475	22.02 / 260.900
			18 / 15.00	19 / 15.83	17 / 14.17	19 / 15.83
miss3	0269	120	42.08 / 238.900	42.14 / 238.150	41.99 / 239.350	42.24 / 236.825
			2 / 6.25	2 / 6.25	1 / 3.12	1 / 3.12
miss3	0275	32	30.93 / 342.775	31.11 / 343.375	31.20 / 343.925	31.19 / 342.100
			0 / NA	0 / NA	0 / NA	0 / NA
miss3	0556	0	29.33 / 245.300	29.94 / 244.950	29.29 / 245.425	29.51 / 245.525
			12 / 21.05	14 / 24.56	15 / 26.32	13 / 22.81
miss3	0911	57	23.80 / 244.325	23.86 / 245.250	23.64 / 246.525	23.67 / 245.275
			15 / 19.48	16 / 20.78	16 / 20.78	15 / 19.48
miss3	0928	77	20.48 / 278.625	20.00 / 279.950	20.06 / 276.775	19.95 / 278.050

Candidate FP accounting with zero eligible GT

The scene-only evaluator leaves held-out metrics blank when there is no eligible GT. Recall remains undefined, but every retained held-out candidate is unmatched. For each condition, state and IoU threshold t, the counts add across disjoint frames:

$$F_{0556,t} = F_{\mathrm{all},t} - \sum_{s \neq 0556} F_{s,t}.$$

At IoU 0.25, the archive gives integer inputs directly. At IoU 0.5, counts are recovered as 200 FP/frame for the probe and 40 FP/frame for each scene, using unrounded values and checking integrality. The two residuals agree in every row. Probe-wide TP plus FP equals the number of retained candidates at both IoUs. These checks verify internal count consistency; they do not re-evaluate scene-0556 independently.

Supplementary Table S5: Scene-0556 candidate FP residuals on 40 frames. Other is the sum over the other four scenes. Residual is the common count at both thresholds.

Condition	State	All@0.25	Other@0.25	All@0.5	Other@0.5	Residual	FP/frame
Clean	M0	49535	40397	49573	40435	9138	228.450
Clean	M5-0	49588	40423	49634	40469	9165	229.125
Clean	M5-1	49640	40432	49683	40475	9208	230.200
Clean	M5-2	49530	40387	49572	40429	9143	228.575
miss1	M0	50943	41640	50982	41679	9303	232.575
miss1	M5-0	51119	41750	51165	41796	9369	234.225
miss1	M5-1	51120	41713	51161	41754	9407	235.175
miss1	M5-2	50937	41579	50976	41618	9358	233.950
miss3	M0	53956	44144	53997	44185	9812	245.300
miss3	M5-0	54022	44224	54067	44269	9798	244.950
miss3	M5-1	54037	44220	54080	44263	9817	245.425
miss3	M5-2	53869	44048	53911	44090	9821	245.525
miss6	M0	59445	48846	59481	48882	10599	264.975
miss6	M5-0	59362	48755	59401	48794	10607	265.175
miss6	M5-1	59484	48905	59520	48941	10579	264.475
miss6	M5-2	59274	48614	59309	48649	10660	266.500

Leave-one-scene-out sensitivity

Each deletion removes the same 40 frames from M0 and all three M5 states in that condition. Recall and known AP were recomputed on the pooled remaining predictions; individual scene APs were not averaged. The three states share the same five scenes. Neither these states nor overlapping deletions constitute independent probe replicates.

Supplementary Table S6: All leave-one-scene-out results on Clean and miss6 (160 remaining frames per row). Differences are M5 minus M0, in percentage points, in state order 0/1/2.

			M0 (%)		ΔR_{H}@0.25 (pp)			$\Delta\mathrm{mAP}_{\mathrm{cd}}$ (pp)		
Condition	Omitted	GT	R_{H}@0.25	$\mathrm{mAP}_{\mathrm{cd}}$	0	1	2	0	1	2
Clean	0269	166	16.265	35.307	+4.217	+3.012	+2.410	-0.387	-0.092	-0.454
Clean	0275	254	16.535	51.978	+3.150	+2.362	+0.787	-0.166	+0.220	-0.181
Clean	0556	286	15.035	39.822	+3.147	+2.098	+1.399	-0.755	-0.390	-0.196
Clean	0911	229	12.227	36.988	+1.747	+1.310	0.000	-0.822	-0.377	-0.202
Clean	0928	209	15.311	42.444	+3.828	+1.914	+2.871	-0.311	-0.013	+0.288
miss6	0269	166	16.265	29.397	-0.602	-1.205	+0.602	-0.359	-0.276	-0.393
miss6	0275	254	15.748	32.255	+0.394	-0.787	-1.181	-0.169	-0.324	+0.300
miss6	0556	286	14.685	31.758	+0.699	-0.699	-1.049	-0.797	-0.765	-0.692
miss6	0911	229	14.847	27.881	+0.873	-1.747	-1.747	-0.625	-0.885	-0.891
miss6	0928	209	11.962	34.345	+1.914	+0.957	-1.435	-0.531	-0.641	-0.588

On Clean, the recall change ranges from +1.747 to +4.217 pp (M5-0), +1.310 to +3.012 pp (M5-1), 0.000 to +2.871 pp (M5-2). The known-class mAP difference becomes positive when omitting scene-0275 for M5-1 or scene-0928 for M5-2. Under miss6 the full-probe recall changes are mixed (+0.699, −0.699, −1.049 pp); the Clean results do not establish recovery under complete camera-view dropout.

Supplementary Note: Shared candidate ceilings (P1)

Cached post-decode held-out candidates were sorted by descending score using stable ordering for ties and truncated to the same per-frame ceiling K for each state. Matching and evaluation were repeated on the truncated lists; the known branch was unchanged. Same K does not imply equal candidate counts or equal FP/frame. No matching independent calibration predictions were available, so these tables are descriptive.

On Clean at $K = 100$, M5-0/1/2 exceed M0's held-out recall by +2.448, +2.098 and +1.049 pp. At $K = 25$, M5-2 falls 0.699 pp below M0 despite equal displayed FP/frame. At $K = 500$, approximately 248 candidates per frame are retained; the higher recall of the three M5 states is accompanied by known mAP that is 0.117–0.629 pp lower.

Supplementary Table S7: Clean probe B under all seven shared candidate ceilings (200 frames and 286 eligible held-out observations per state and ceiling). ΔR_{H}@0.25 is relative to M0 at the same ceiling.

K	State	Cand./frame	FP/frame@0.5	TP@0.25	R_{H}@0.25 (%)	ΔR_{H}@0.25 (pp)	$\mathrm{mAP}_{\mathrm{cd}}$ (%)
10	M0	10.000	10.000	2	0.699	0.000	39.645
10	M5-0	10.000	10.000	4	1.399	+0.699	39.017
10	M5-1	10.000	10.000	4	1.399	+0.699	39.377
10	M5-2	10.000	10.000	1	0.350	-0.350	39.528
25	M0	25.000	24.990	10	3.497	0.000	39.645
25	M5-0	25.000	24.990	11	3.846	+0.350	39.017
25	M5-1	25.000	24.990	14	4.895	+1.399	39.377
25	M5-2	25.000	24.990	8	2.797	-0.699	39.528
50	M0	50.000	49.980	28	9.790	0.000	39.645
50	M5-0	50.000	49.980	35	12.238	+2.448	39.017
50	M5-1	50.000	49.975	37	12.937	+3.147	39.377
50	M5-2	50.000	49.985	26	9.091	-0.699	39.528
100	M0	100.000	99.975	43	15.035	0.000	39.645
100	M5-0	100.000	99.970	50	17.483	+2.448	39.017
100	M5-1	100.000	99.970	49	17.133	+2.098	39.377
100	M5-2	100.000	99.975	46	16.084	+1.049	39.528
200	M0	197.940	197.915	43	15.035	0.000	39.645
200	M5-0	198.175	198.145	52	18.182	+3.147	39.017
200	M5-1	198.075	198.045	49	17.133	+2.098	39.377
200	M5-2	198.195	198.170	47	16.434	+1.399	39.528
300	M0	242.855	242.830	43	15.035	0.000	39.645
300	M5-0	243.135	243.105	52	18.182	+3.147	39.017
300	M5-1	243.145	243.115	49	17.133	+2.098	39.377
300	M5-2	242.510	242.485	47	16.434	+1.399	39.528
500	M0	247.890	247.865	43	15.035	0.000	39.645
500	M5-0	248.200	248.170	52	18.182	+3.147	39.017
500	M5-1	248.445	248.415	49	17.133	+2.098	39.377
500	M5-2	247.885	247.860	47	16.434	+1.399	39.528

Supplementary Table S8: miss6 probe B under the three measured ceilings (200 frames and 286 eligible held-out observations per state and ceiling). ΔR_{H}@0.25 is relative to M0 at the same ceiling.

K	State	Cand./frame	FP/frame@0.5	TP@0.25	R_{H}@0.25 (%)	ΔR_{H}@0.25 (pp)	$\mathrm{mAP}_{\mathrm{cd}}$ (%)
50	M0	50.000	49.980	32	11.189	0.000	30.624
50	M5-0	50.000	49.975	31	10.839	-0.350	29.996
50	M5-1	50.000	49.980	31	10.839	-0.350	29.925
50	M5-2	50.000	49.980	31	10.839	-0.350	29.966
100	M0	100.000	99.970	41	14.336	0.000	30.624
100	M5-0	100.000	99.975	42	14.685	+0.350	29.996
100	M5-1	100.000	99.980	39	13.636	-0.699	29.925
100	M5-2	100.000	99.980	38	13.287	-1.049	29.966
500	M0	297.435	297.405	42	14.685	0.000	30.624
500	M5-0	297.030	297.005	44	15.385	+0.699	29.996
500	M5-1	297.620	297.600	40	13.986	-0.699	29.925
500	M5-2	296.565	296.545	39	13.636	-1.049	29.966

Supplementary Note: Additional formal and fixed-probe results

The online S2 stages, seen fixed probe and independent B probe answer different questions. In S2, frame content changes between stages. The 100-frame seen probe overlaps every S2 stream and tests input sensitivity on the same frames. The archived SoftRestore-Consistency-3D-Adapter is a separate external-method-inspired control restricted to the adapters. It does not implement standard CoTTA or teacher-augmentation averaging; it is not among the frozen M0/M5 states analyzed in P0/P1.

Supplementary Table S9: Stage-wise changes in continuing online S2 (M5 minus M0, pp). Pairs use identical frames within a stream and stage; stages differ in frame content.

	$\Delta\mathrm{mAP}_{\mathrm{cd}}$ (pp)			ΔR_{H}@0.25 (pp)		
Stage	Seed 0	Seed 1	Seed 2	Seed 0	Seed 1	Seed 2
Initial Clean	-0.015	-0.067	-0.228	-0.901	+0.690	-0.478
miss1	-0.266	+0.021	+0.088	+5.263	+0.000	-1.418
miss3	+0.452	-0.046	-0.274	-1.515	+2.041	+0.000
miss6	-0.210	+0.020	+0.155	+1.266	-1.020	+2.703
Terminal Clean	-0.312	+0.137	+0.276	-3.604	-1.709	-1.796

Supplementary Table S10: Seen fixed-probe results with no online updates (100 frames, three scenes). M5 entries are final states 0/1/2. Held-out matches are true positives at 3D IoU 0.25 out of 59 eligible observations.

Condition	M0 mAP (%)	M5 mAP, states 0/1/2 (%)	M0 matches	M5 matches
Clean	56.882	56.875/56.922/56.963	14	16/14/14
miss1	55.611	55.329/55.671/55.904	10	10/13/13
miss3	52.882	53.192/53.235/53.084	10	7/10/10
miss6	48.477	48.846/48.262/48.762	6	8/6/6

Supplementary Table S11: Formal M5 update funnel. All means and contraction percentages are conditional on successful updates. Before-cap pseudo-labels already passed confidence and protection filtering; the 50-box cap precedes query pairing.

Seed	Successful updates	Before cap	After cap	Paired queries	Contractions
0	243	37.09	32.24	31.71	134 (55.1%)
1	239	35.10	31.14	30.74	133 (55.6%)
2	243	35.72	31.38	30.84	133 (54.7%)

Supplementary Note: Mechanism and state provenance

Diagnostic replay A. The corrected archive contains 293 diagnostic observations, with a SHA-256 index for 586 sparse mask/candidate and eligible-GT files. The arrays support offline recomputation of center-in-mask coverage, candidate counts and separate one-to-one matching of candidate sets. Rematching deleted candidates does not trace which true positives were lost in formal inference. Seven formal successful updates did not yield diagnostic dumps: replay rejected five for non-finite gradients and two for excessively large gradients. The trajectories differ; post-query-pairing true positives were not measured.

Limited hook comparison. A separate 100-frame stream-seed-0 check used Python, NumPy, Torch and CUDA seed 12345 in both arms. Counting hooks wrapped filtering and mask operations, without exercising the full dump path. Frame order and update decisions agreed (21 successful updates per arm). Logged losses differed on 20 frames and known pseudo-label counts on three. Detection arrays were not extracted, and the combined-log fingerprint does not verify predictions or full dump-path non-interference.

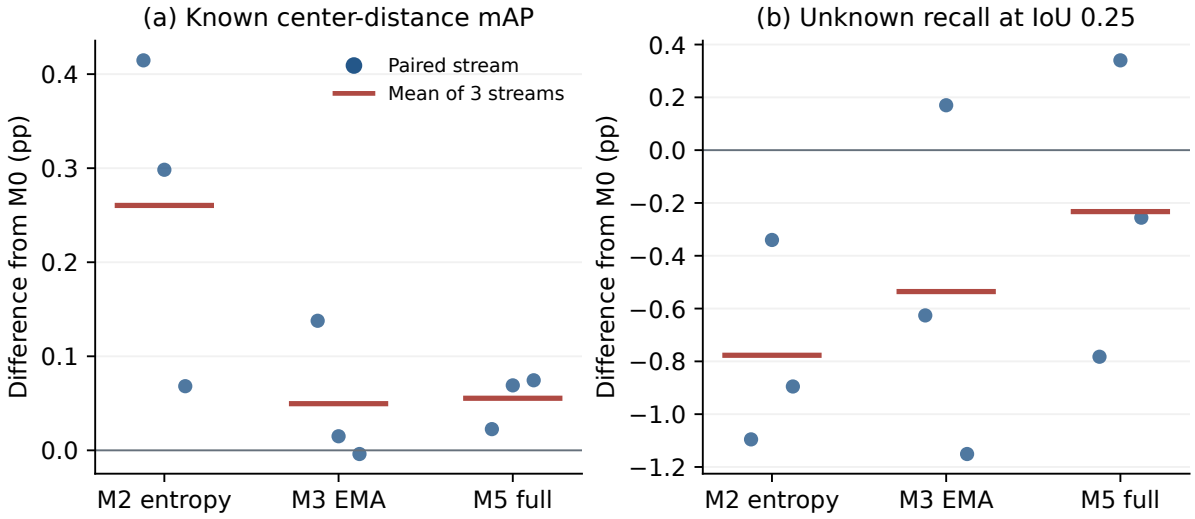
Probe B state provenance. The replay script names the original v2b M5 run path, separate from diagnostic replay. Six retrospectively exported formal/replay state-file hashes are verifiable; formal and diagnostic final adapter hashes differ for all three seeds. This distinguishes preserved files and the documented load path from an in-memory checksum at probe inference. The frozen-probe logs record 800 forwards per state and no adapter or normalization-state change.

Control naming and timing. Historical SoftRestore archive names include a CoTTA-inspired alias. The 100-frame smoke test logged no restore; the three complete runs logged 1, 3 and 3 events. A matched timing remeasurement was unavailable; no acceleration claim is made.

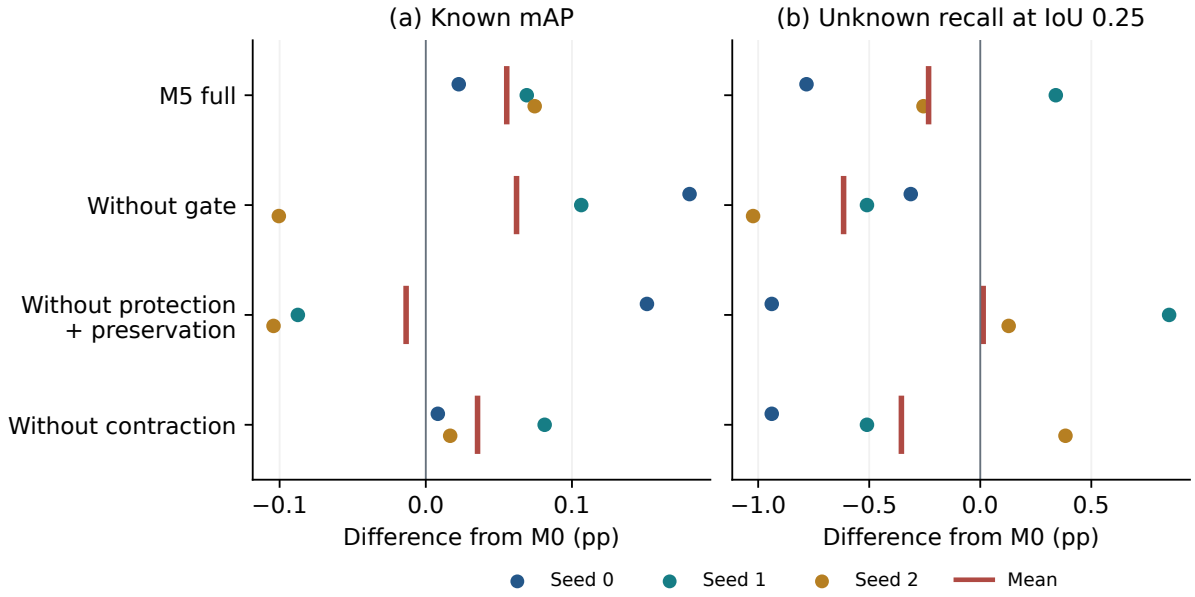
Evidence files. The experiment archive contains the following machine-readable records:

- S2 cross-check and B re-evaluation: `verify/verify_report.json`.
- Scene pairing and LOSO: `p0/scene_paired.csv`, `p0/leave_one_scene_out.csv`.
- Clean budgets: `p1/candidate_budget.csv`, `p1/candidate_budget_deltas.csv`.
- miss6 budgets: `p1/candidate_budget_miss6_subset.csv`.

Supplementary Note: Formal per-stream comparisons



Supplementary Figure S1: Whole-stream paired differences relative to M0 on identical within-seed schedules. Dots show individual streams and horizontal red segments show arithmetic means. The horizontal zero line is the frozen reference. Differences are percentage points. *Unknown recall* denotes held-out recall here and in Supplementary Fig. S2.



Supplementary Figure S2: Paired module differences relative to M0 across all three S2 streams. Colored dots identify seeds and red vertical segments show arithmetic means. The gating ablation removes permission and scales jointly. The protection ablation removes both mask filtering and preservation loss. These comparisons do not isolate the paired operations individually.