Stage 11 — Step 5 Breakthrough (One■Pager)

Why this helps as a banner for the new thread:

• Sets shared context in one glance (what changed, why it worked).
• Anchors all follow■ups to a stable baseline (S5■3) for apples■to■apples comparisons.
• Prevents regression to pre■Step■5 ideas (e.g., median aggregation).
• Creates a crisp success metric target for the team (hold recall=1.0, push hallucination <0.26).

What changed in Step 5:

• Prototype redesign: multi■scale half (halfN/halfW), hinges (L/R), derivative.
• Aggregation: switched to softmin ($\tau \approx 0.40–0.45$) with consensus ($k \geq 2$, $\varepsilon \approx 0.10–0.12$).
• Orthogonalization: reduce cross■family leakage (flip_v ■ {flip_h, rotate}).
• Per■primitive gates: raw_floor & residual_ceiling only where needed (flip_v).

Key Results (50 samples each):

Sweep	Accuracy	Precision	Recall	F1	Jaccard	Halluc.	Margins (μ / min)
S5■1	0.36	0.707	1.00	0.792	0.707	0.293	1.67 / 0.90
S5■2	0.24	0.620	1.00	0.732	0.620	0.380	1.56 / 0.89
S5■3	0.36	0.713	1.00	0.796	0.713	0.287	1.88 / 1.03

Takeaways:

• First structural drop in hallucination (~0.34 → ~0.29) with recall held at 1.0.
• Softmin + consensus is essential; median aggregation regresses (see S5■2, S6■4).
• flip_v hallucination finally below the stubborn wall (~0.28–0.29).
• Margins improved ($\mu \approx 1.9$, $\min \approx 1.0$) — wells sharpened without brittleness.

Targets for the new thread:

• Keep S5■3 as the baseline; document any changes against it.
• Phantom suppression: aim Hallucination ≤ 0.26 overall; flip_v ≤ 0.24, Recall = 1.0.
• Use surface dumps + ridge plots to localize phantom basins.