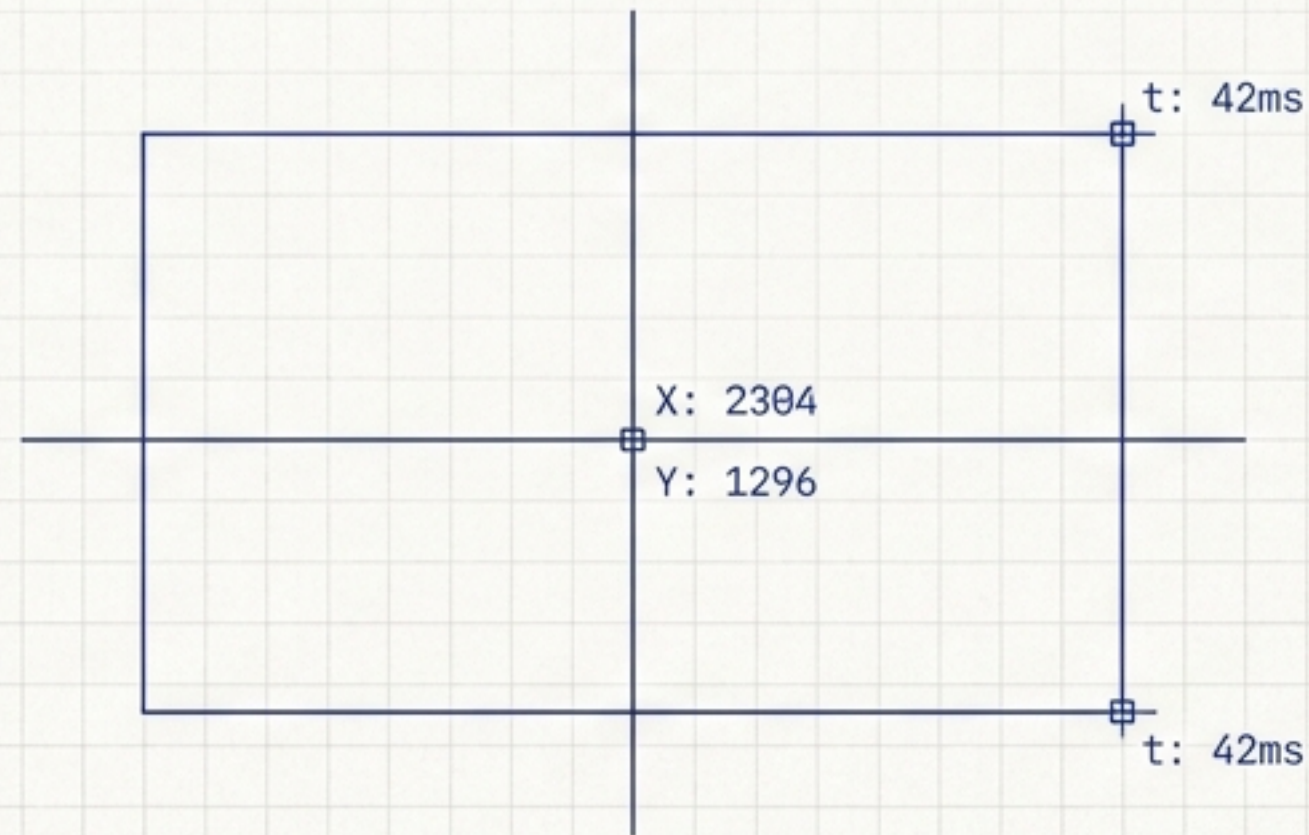
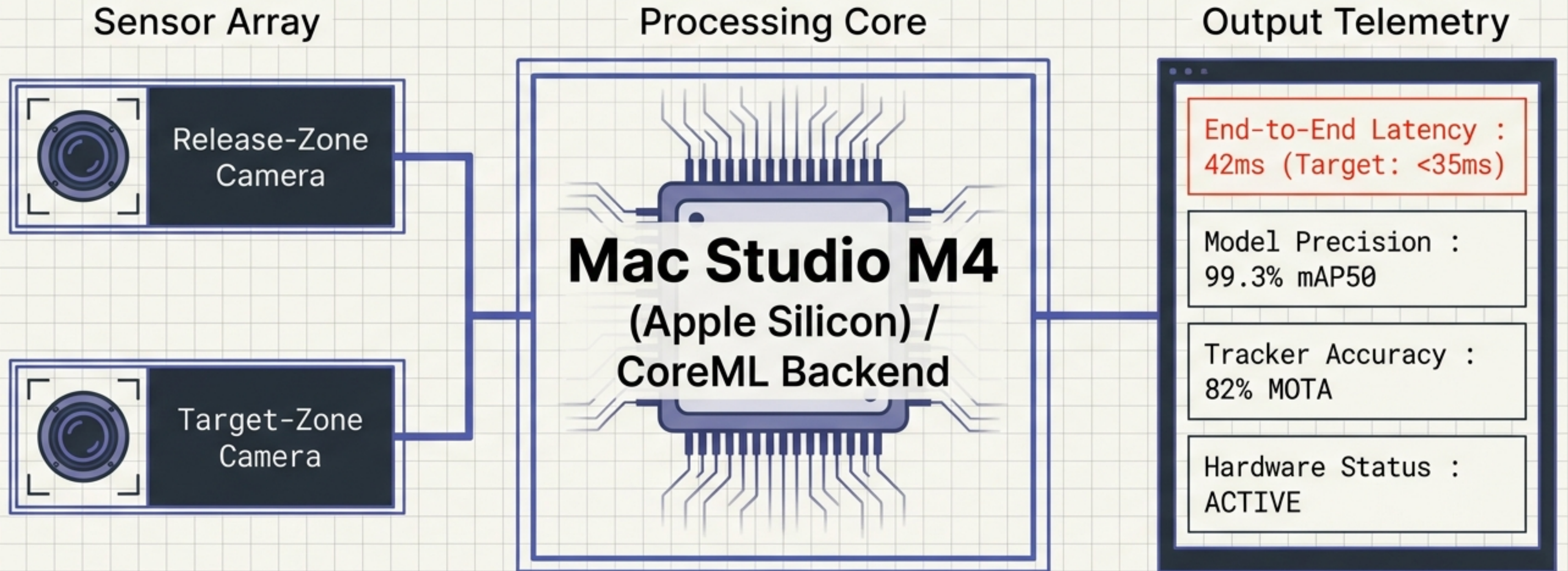


Engineering the QC Real-Time Computer Vision Pipeline

A systems architecture and deployment postmortem for bounded-court tracking.



Deployed Architecture and Production Telemetry

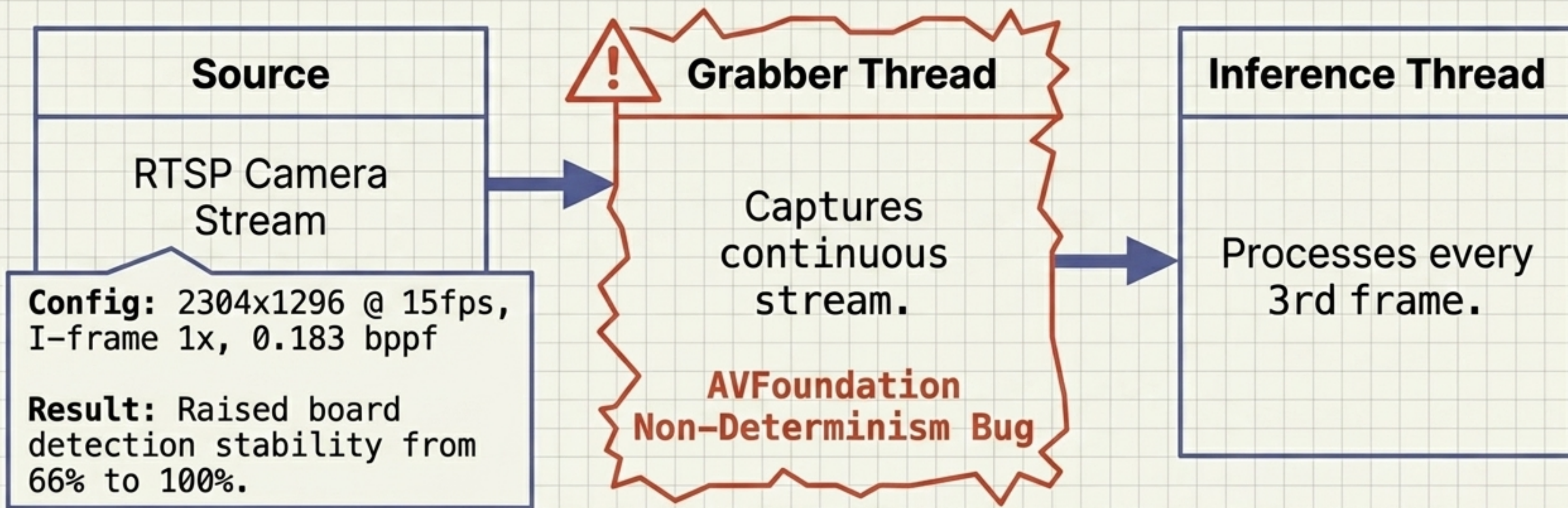


Deployment Constraints Dictate Model Selection

Model Family	Architecture	Post-Processing	Licensing
RT-DETRv2-S (Selected) [~80MB .mlpackage]	Transformer Decoder	NMS-Free (Native Apple Neural Engine Match)	Apache 2.0 (Commercial Friendly)
YOLO11 / YOLO26 (Rejected)	CNN / End-to-End	Requires varying NMS overhead	AGPL-3.0 (Commercial Network Restriction)

Insight: YOLO architectures were bypassed because **AGPL-3.0** triggers network service open-source requirements, making it incompatible with commercial venue deployment. RT-DETR's NMS-free semantics operate significantly faster on the Apple Neural Engine.

Stabilizing the RTSP Video Pipeline

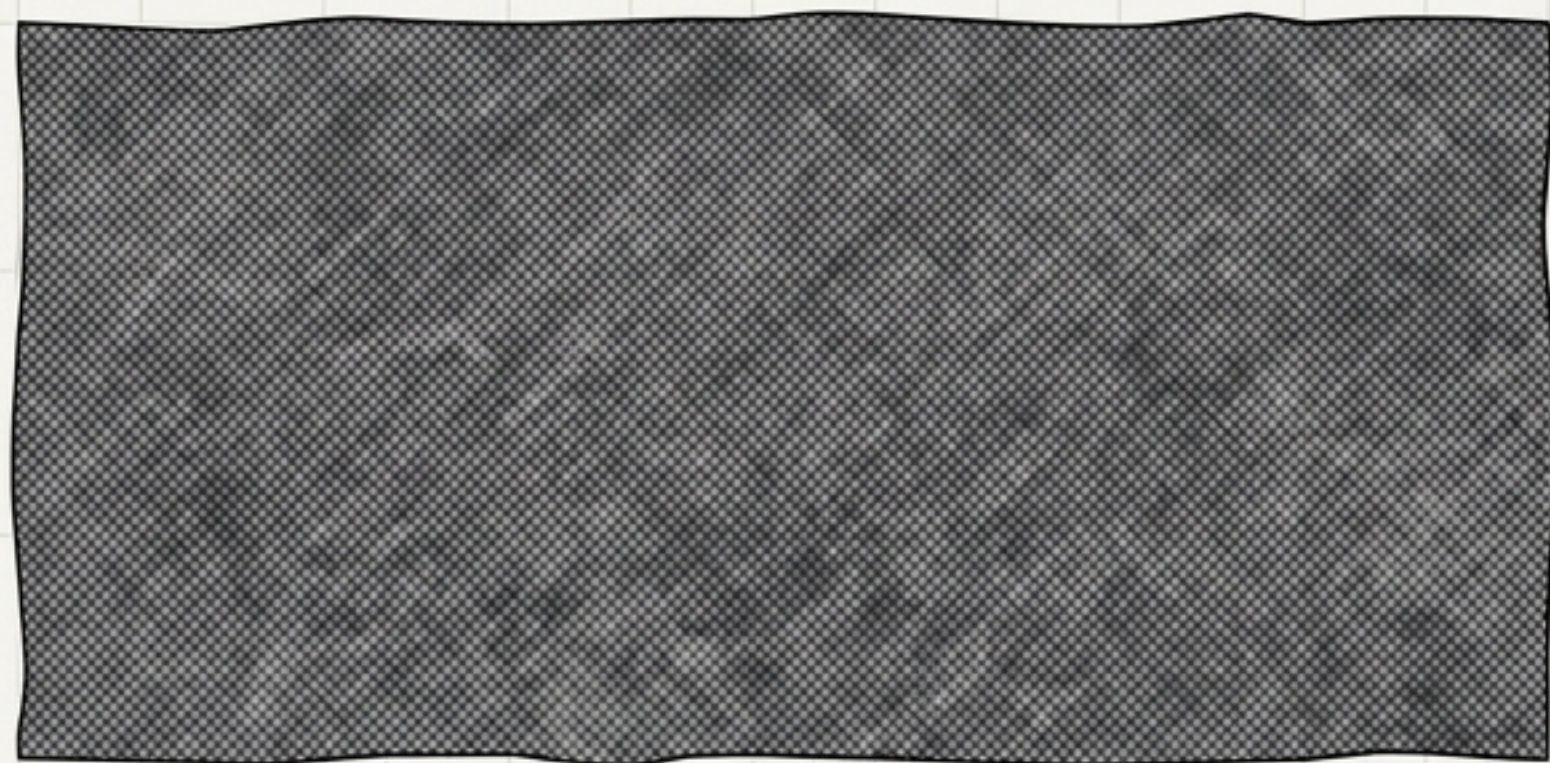


SYMPTOM: macOS AVFoundation silently swaps camera identities at given indices after OS sleep.

RESOLUTION: Hardcoded static RTSP IP assignments (192.168.2.2 and 192.168.2.3), entirely bypassing AVFoundation index references.

The Data Quality Paradox in Training

Iteration v11



4,398 noisily-labeled frames



72.5% mAP50

Iteration v15



505 clean, hand-annotated frames



99.3% mAP50

Insight: A 10x improvement in data quality produced drastically better results than a 12x increase in dataset size using the exact same RT-DETR-S architecture.

Incident Reports: Sensor Blindness and Environmental Noise

INCIDENT 01: SENSOR COVERAGE FAILURE

[Symptom]

Model generalized to the wrong camera during early deployment.

[Root Cause]

Severe data imbalance. Only **42** out of **4,148** training images were captured from the actual production sensor.

[Resolution]

Mandatory sensor matching in all future training datasets.

INCIDENT 02: THE HARD NEGATIVE REQUIREMENT

[Symptom]

System generated phantom projectile detections.

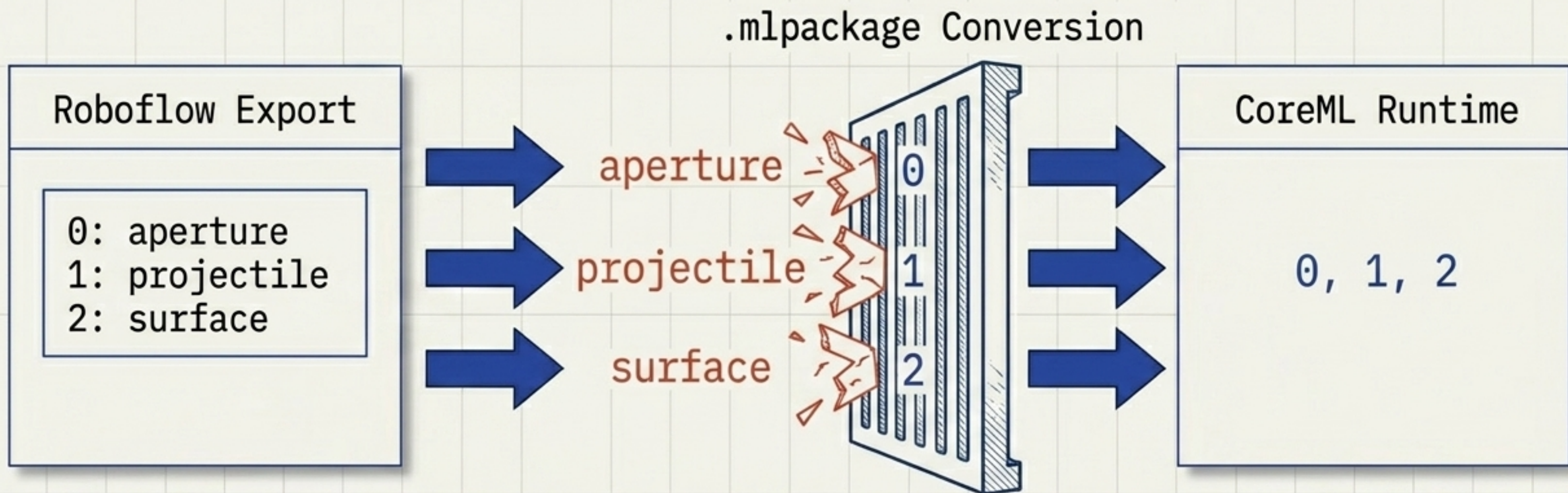
[Root Cause]

The bounded court features heavily decorated scoring surfaces.

[Resolution]

Explicit labeling of court patterns as a 'surface' class. Hard negative examples are required to prevent trackers from locking onto environmental noise.

The “Ground Zero” Class Mapping Bug

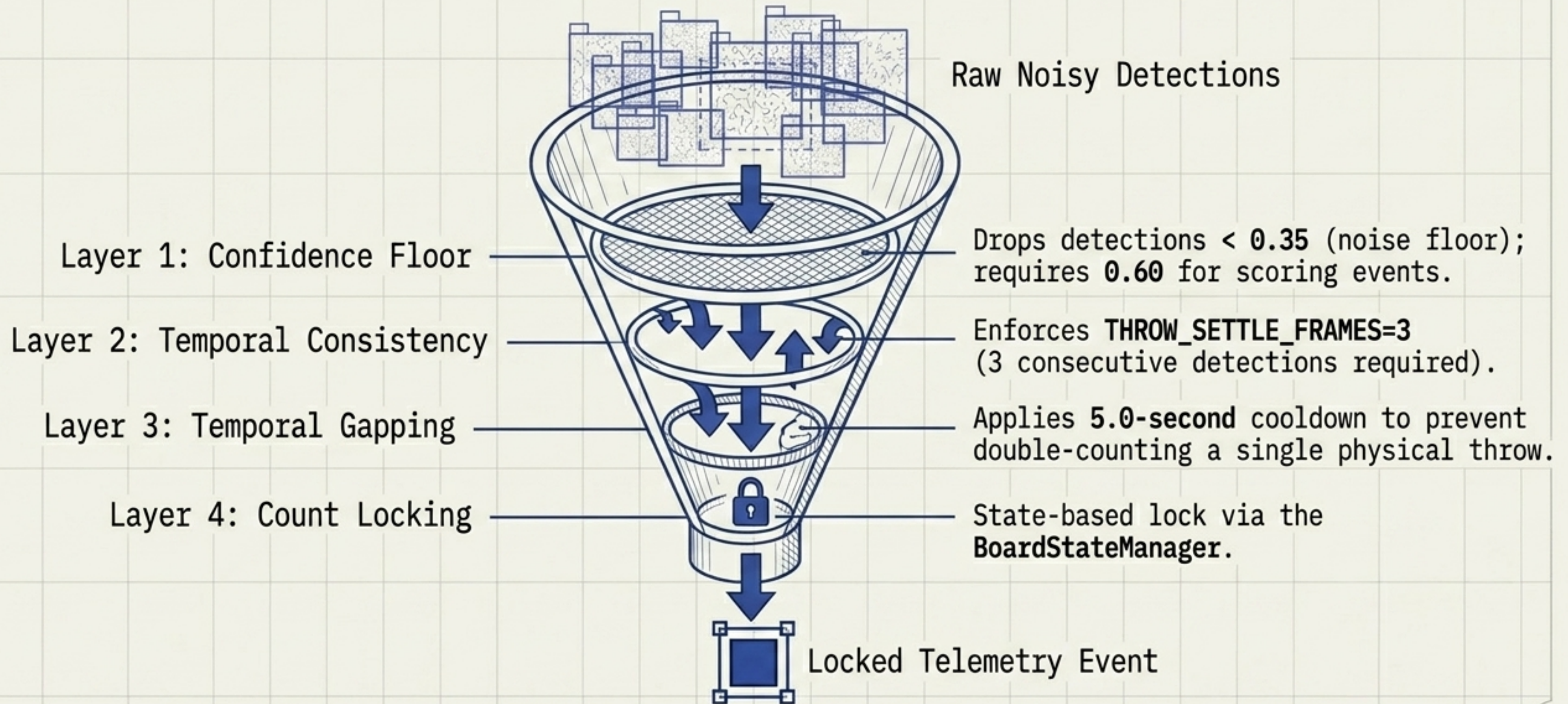


INCIDENT: Version 6 models exhibited fundamental class confusion (scoring aperture detected as a thrown projectile).

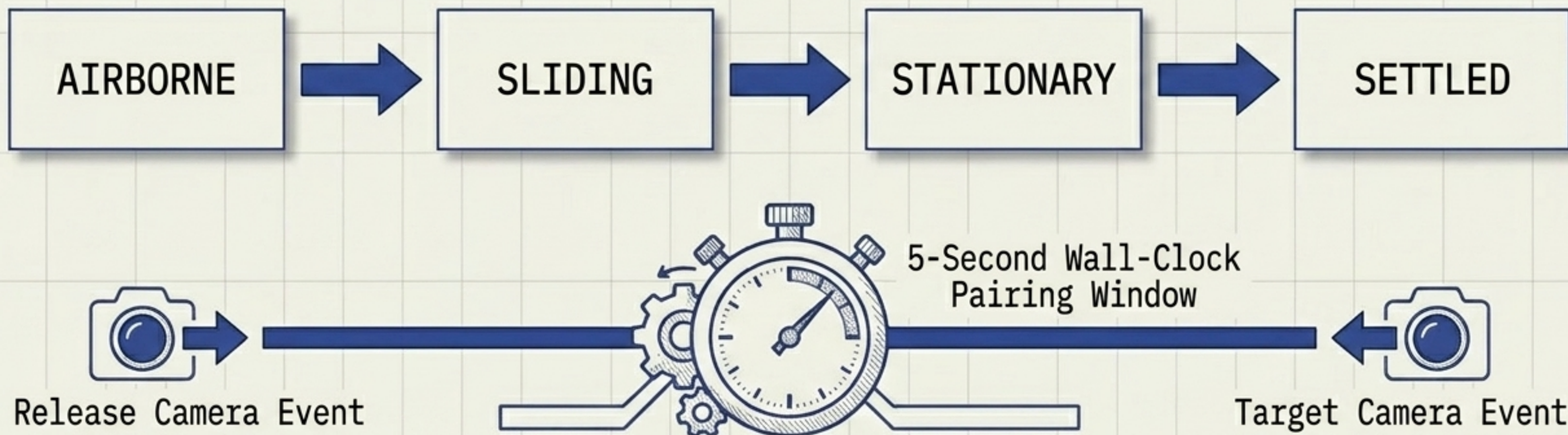
ROOT CAUSE: CoreML .mlpackage models do not expose class names at runtime. The deployed model used alphabetical ordering; inference code assumed legacy ordering.

RESOLUTION: A three-line index mapping fix resolved four days of catastrophic mislabeled detections.

4-Layer Temporal Filter Stack



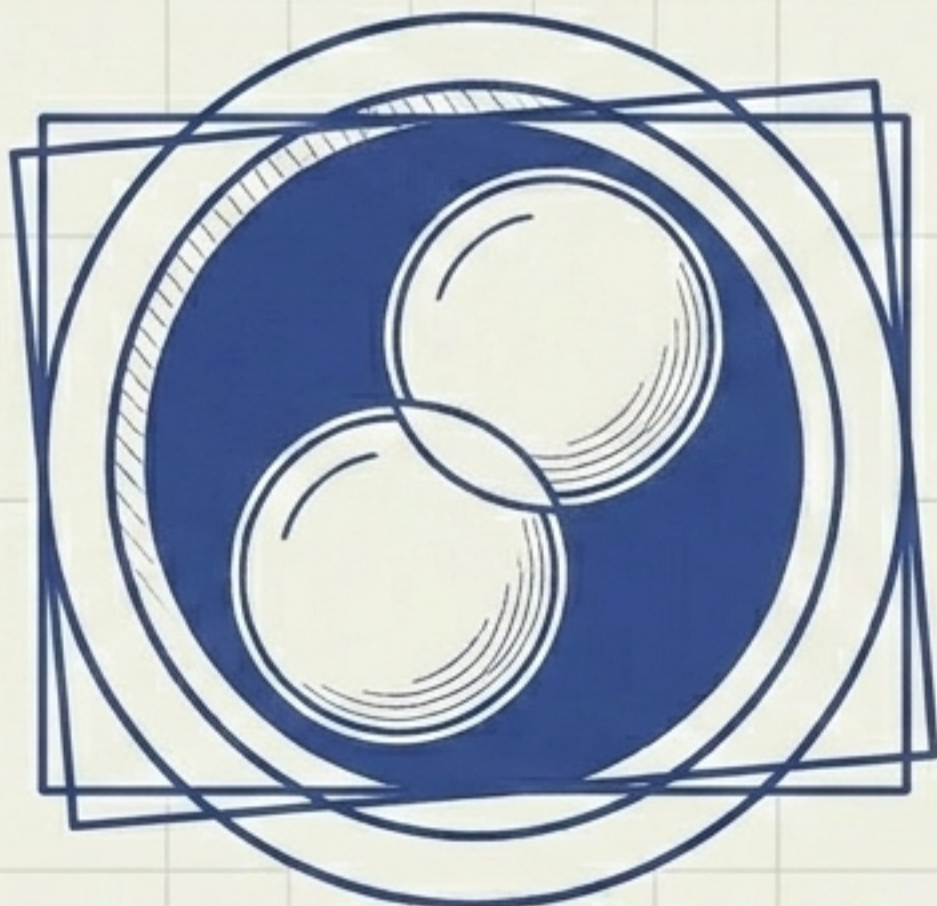
Projectile Settlement State Machine



- **State Trigger:** The SETTLED state only fires after N consecutive STATIONARY frames.
- **The Pairing Window:** Correlating throw events with settlement events requires a 5-second wall-clock window.
- **Critical Constraint (Warning):** This timer must be **wall-clock independent** and isolated from the camera frame loop; unexpected camera stalls will prevent the window from expiring.

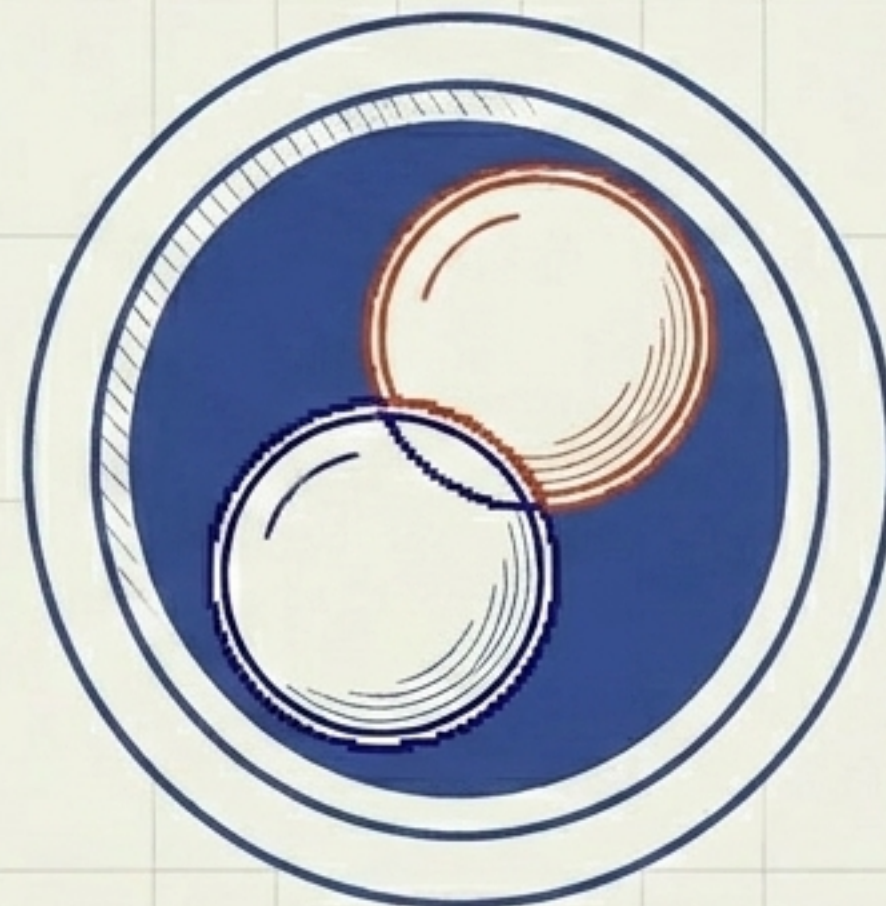
Resolving the Stacked Projectile Problem

Current State



Standard Detection:
Bounding Box Failure

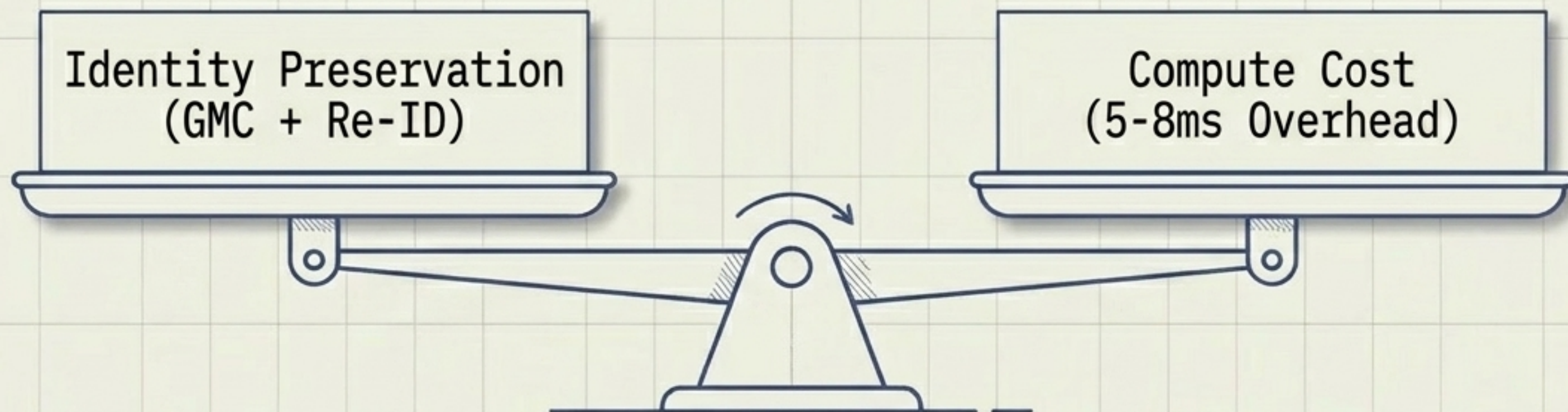
Future State



SAM 2.1 VideoPredictor:
End-of-Turn Disambiguation

- When multiple objects share the same physical boundary, bounding boxes fail to disambiguate. Future architecture triggers a **SAM 2.1** zero-shot segmentation pass after all motion ceases to resolve the final scoring state.

Upgrading Association: ByteTrack to BoT-SORT



▪ Current Limitation:

- ByteTrack relies on motion-only association (Kalman filters), resulting in identity swaps during projectile-on-projectile occlusion when motion vectors overlap.

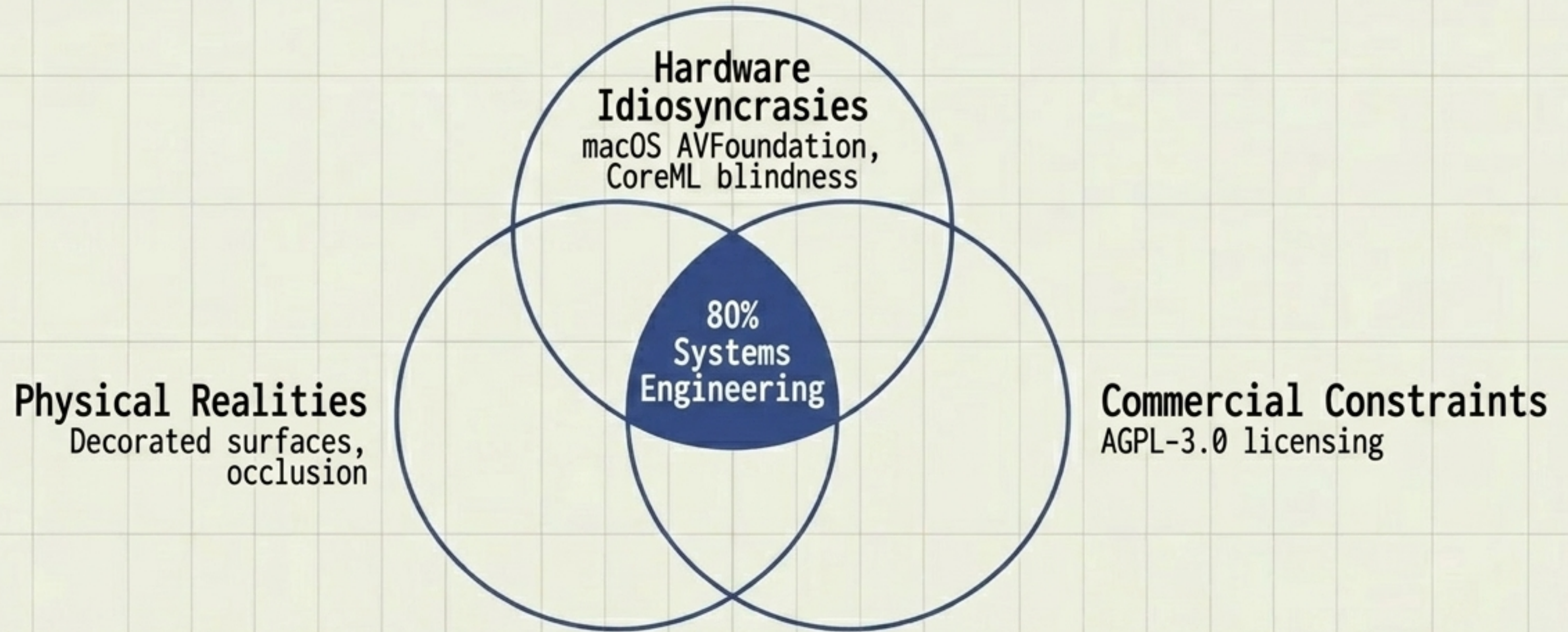
▪ Target Architecture:

- BoT-SORT introduces appearance-based Re-ID and Global Motion Compensation (GMC) to maintain identities through visual occlusion.

▪ Compute Constraint:

- Implementing BoT-SORT costs **5-8ms** per frame. The fully NMS-free **RT-DETRv2** backbone is a strict prerequisite to afford this overhead while maintaining the **<35ms** latency target.

Synthesis: Deployment Reality Outweighs Benchmarks



1. Data over Architecture:
500 clean frames will always beat 4,000 noisy frames.

2. Beware the Runtime: Conversion to edge deployment formats (CoreML) destroys training-time metadata. Validate index mappings manually.

3. Licensing is Architecture: Raw model capability is irrelevant if the license restricts commercial network deployment.