Jetson Orin Edge-Vision Scoping Worksheet

Size a detector + tracker + segmenter deployment before the next hardware order.

1. Pick the board — match AI throughput to your camera count

Module (2026, Super)	Peak AI (INT8)	Good for
Orin Nano 8GB	67 TOPS	1 camera: detect + track, occasional segment
Orin NX 16GB	157 TOPS	Several cameras, more segmenting
AGX Orin 64GB	275 TOPS	Many cameras, larger models
AGX Thor	800 TOPS	Robotics, several large models at once

2. Budget the frame — 33 ms at 30 fps (representative Orin NX, 1 camera)

Stage	Cost	Note
Decode (NVDEC)	~2 ms	Hardware decoder, runs in parallel — near-free of AI budget
Detect (YOLO, INT8)	~15 ms	The big cost. Size the board around this stage
Track (ByteTrack)	< 1 ms	CPU-side geometry. Nearly free
Segment (NanoSAM)	~3 ms	Only when a rule needs it, amortised over frames
Rules + overlay	~3 ms	Decide + draw on the GPU
Headroom	~9 ms	Safety margin — absorbs busy frames + thermal throttle

3. The trade — total work is fixed; divide it between cameras and frame rate

- ☐ High fps, few cameras — licence-plate at a gate, fast events.
- ☐ Batch frames — run the detector once over all cameras' frames together.
- ☐ Compile INT8 + TensorRT — and validate the accuracy gap on your footage.
- ☐ Low fps, many cameras — 'was anyone here?' across a site.
- ☐ ~40 streams @ ~5 fps — Orin NX 16GB, small YOLOv8 INT8 (NVIDIA figure).
- ☐ Keep frames on the GPU — move only results across the chip boundary.

4. The four traps that miss the frame budget

Quoting the benchmark frame rate

'65 FPS YOLO' is the detector alone. Your pipeline also decodes, tracks, segments, and draws.
Fix: budget the whole pipeline, on your hardware and footage.

Copying the frame around

Crossing the CPU/GPU boundary copies the frame — often costlier than the AI itself.
Fix: keep the frame on the GPU; move only boxes and masks.

Segmenting everything, every frame

Full SAM 2 on every object blows the budget — it runs near 1 fps on small devices.
Fix: distilled SAM (NanoSAM), selected tracks only, reuse across frames.

Ignoring the enclosure heat

A throttled board in a hot housing silently drops frame rate below the bench number.
Fix: validate fps in the real enclosure, at real ambient temperature.