

Памятка по подбору edge-зрения на Jetson Orin

Оцените установку «детектор + трекер + сегментер» до следующего заказа железа.

1. Выберите плату — сопоставьте AI-производительность числу камер

Модуль (2026, Super)	Пик AI (INT8)	Подходит для
Orin Nano 8GB	67 TOPS	1 камера: детекция + трекинг, эпизод. сегмент
Orin NX 16GB	157 TOPS	Несколько камер, больше сегментации
AGX Orin 64GB	275 TOPS	Много камер, более крупные модели
AGX Thor	800 TOPS	Робототехника, несколько больших моделей

2. Распределите бюджет кадра — 33 мс при 30 fps (Orin NX, 1 камера)

Стадия	Стоимость	Заметка
Декод (NVDEC)	~2 мс	Аппаратный декодер, параллельно — почти без AI-бюджета
Детекция (YOLO, INT8)	~15 мс	Главная статья. Плату подбирают под эту стадию
Трекинг (ByteTrack)	< 1 мс	Геометрия на CPU. Почти бесплатно
Сегментация (NanoSAM)	~3 мс	Только когда нужно правилу, размазана по кадрам
Правила + наложение	~3 мс	Решить + нарисовать на GPU
Запас	~9 мс	Предохранительная маржа — гасит загрузку + троттлинг

3. Компромисс — работа фиксирована; делите её между камерами и частотой

- ☐ Высокий fps, мало камер — номера у ворот, быстрые события.
- ☐ Низкий fps, много камер — «был ли кто-то здесь?» по площадке.
- ☐ Пакетируйте кадры — запускайте детектор разом по всем камерам.
- ☐ ~40 потоков @ ~5 fps — Orin NX 16GB, малый YOLOv8 INT8 (цифра NVIDIA).
- ☐ Компилируйте INT8 + TensorRT — и проверяйте разрыв точности на видео.
- ☐ Держите кадры на GPU — перемещайте только результаты через границу.

4. Четыре ловушки, которые промахиваются по бюджету кадра

Выдавать бенчмарковый fps за свой

«65 FPS YOLO» — это детектор в одиночку. Ваш конвейер ещё декодирует, трекает, сегментирует, рисует. Решение: бюджет на весь конвейер, на вашем железе и видео.

Сегментировать всё, каждый кадр

Полный SAM 2 на каждом объекте срывает бюджет — он работает около 1 fps на малых устройствах. Решение: облегчённый SAM (NanoSAM), только выбранные треки, переиспользуйте.

Копировать кадр туда-сюда

Пересечение границы CPU/GPU копирует кадр — часто дороже самого AI. Решение: держите кадр на GPU; перемещайте только рамки и маски.

Игнорировать тепло корпуса

Троттлящая плата в горячем корпусе молча роняет fps ниже стендовой цифры. Решение: валидируйте fps в реальном корпусе, при реальной температуре.