

Análise de Custos e Infraestrutura

Monitoramento Inteligente por Visão Computacional e IA
NeurAffect · Sistemas Inteligentes

Sumário

1. Carga de processamento
2. Capacidade por GPU
3. Arquitetura recomendada
4. Custo total de propriedade (por unidade, 3 anos)
5. Conclusão

1. Carga de processamento

Cada câmera em 1080p a 15 fps exige 15 inferências de detecção por segundo. Em CPU pura, o processamento é inviável em tempo real para múltiplas câmeras; portanto, a GPU é componente obrigatório da arquitetura.

Modelo	GPU (ms/quadro)	CPU pura (ms/quadro)
YOLOv8n (nano)	2–4	60–120
YOLOv8s (small)	5–8	150–300
YOLOv8m (medium)	10–15	400–800

2. Capacidade por GPU

GPU	Câmeras simultâneas	Consumo	Preço
NVIDIA Jetson Orin NX 16GB	8–16	15–40 W	R\$ 4.500–6.500
NVIDIA Jetson AGX Orin 64GB	20–40	30–60 W	R\$ 11.000–15.000
RTX 4060 8GB (mini PC)	15–30	115 W	R\$ 6.500–9.500
RTX 4090 24GB (servidor)	60–120	350–450 W	R\$ 25.000–40.000

3. Arquitetura recomendada

Arquitetura híbrida: processamento na borda (dispositivo local com GPU em cada unidade) e gestão central em servidor próprio, que recebe apenas alertas e metadados. O vídeo permanece na unidade, em conformidade com a LGPD.

4. Custo total de propriedade (por unidade, 3 anos)

Componente	Edge (Jetson)	Edge (RTX 4060)	Nuvem (100%)
CAPEX de hardware	R\$ 6.000	R\$ 9.000	R\$ 0
OPEX (energia + manutenção, 3 anos)	R\$ 3.000–4.500	R\$ 6.500–9.500	—
Instâncias GPU (3 anos)	—	—	R\$ 58.000–76.000
Total por unidade (3 anos)	R\$ 9.000–10.500	R\$ 15.500–18.500	R\$ 58.000–76.000

5. Conclusão

A arquitetura de borda é de seis a oito vezes mais econômica que o processamento integral em nuvem no horizonte de três anos, além de oferecer menor latência, independência de banda e conformidade com a LGPD.