

Otimização de pipeline multicritério para detecção de pessoas em dispositivos ultracompactos de borda

João B. A. Santos e Paulo A. L. Rego

¹Universidade Federal do Ceará (UFC)
Departamento de Computação

batistajoaguns@alu.ufc.br, paulo@dc.ufc.br

Abstract. *Edge computing has established itself as a fundamental strategy for computer vision applications that demand low latency and privacy control. However, edge devices face constraints regarding energy consumption and operational costs, making it difficult to choose the ideal configuration for each scenario. This paper proposes a multi-criteria pipeline for ultra-compact devices and compares reCamera, Jetson Xavier NX, and a desktop, using MOT17 and Pedestrian datasets, with external wattmeter and DORI-stratified metrics. Although the desktop achieves the highest accuracy, the reCamera operates at a constant 1.9 W, up to 59× lower than the desktop and 50× cheaper, making it the best alternative for environments with restricted electrical infrastructure. The pipeline indicates 720p at 5 FPS as optimal for low-density scenes and 480p at 30 FPS for real-time deployments.*

Resumo. *A computação de borda consolidou-se como estratégia fundamental para aplicações de visão computacional que exigem baixa latência e controle de privacidade. Contudo, dispositivos de borda enfrentam restrições de consumo energético e custo operacional, dificultando a escolha da configuração ideal. Este artigo propõe um pipeline multicritério para dispositivos ultracompactos e compara reCamera, Jetson Xavier NX e desktop, usando os datasets MOT17 e Pedestrian com wattímetro externo e métricas por faixas DORI. Embora o desktop apresente os melhores resultados, a reCamera opera com 1,9 W constante, até 59× menor que o desktop e 50× mais barata, sendo a melhor alternativa para ambientes com infraestrutura elétrica restrita. O pipeline indica 720p a 5 FPS como configuração ótima para cenas de baixa densidade e 480p a 30 FPS para implantações em tempo real.*

1. Introdução

A evolução da infraestrutura de computação nas últimas décadas tem sido marcada pela busca contínua por maior capacidade de processamento, menor latência e maior flexibilidade arquitetural. Nos anos 1990, o modelo cliente-servidor consolidou-se como paradigma dominante, centralizando o processamento em servidores dedicados. Com a expansão da conectividade de banda larga nos anos 2000, o processamento migrou progressivamente para estruturas em nuvem, tornando plataformas como AWS, Microsoft Azure e Google Cloud referências em escalabilidade para armazenamento e processamento de dados [Dhar 2012].

Contudo, esse modelo centralizado impõe limitações relevantes: a dependência de conectividade contínua, a latência introduzida no trajeto entre dispositivos e *data-centers*, os elevados custos de transmissão de grandes volumes de dados e as restrições de privacidade decorrentes do envio de informações sensíveis tornam a nuvem inadequada para aplicações com requisitos de tempo real. Para endereçar esses desafios, [Bonomi et al. 2012] propuseram o conceito de *fog computing*, introduzindo uma camada intermediária de processamento entre os dispositivos finais e a nuvem. O *edge computing* aprofundou essa proposta ao deslocar o processamento diretamente para os dispositivos finais ou para servidores situados na borda da rede, eliminando a necessidade de transmissão remota para tarefas de baixa tolerância à latência [Shi et al. 2016]. Impulsionado pelo ecossistema de IoT (*Internet of Things*) e pela crescente demanda por inferência de IA local, o mercado global de computação de borda vem registrando uma expansão expressiva nos últimos anos [Cao et al. 2023].

Aplicações de visão computacional na borda, como detecção de pedestres, monitoramento de tráfego e vigilância urbana, impõem requisitos particularmente exigentes: inferência em tempo real, robustez diante de variações de iluminação e de densidade de objetos, e operação com baixo consumo energético. A implantação de modelos de detecção de objetos em dispositivos embarcados é, entretanto, desafiadora, uma vez que tais dispositivos apresentam restrições de poder computacional, de memória e de compatibilidade de hardware, o que dificulta a comparação sistemática entre plataformas e soluções [Alqahtani et al. 2024].

Nesse cenário, uma nova categoria de dispositivos ultracompactos emerge como alternativa promissora: módulos que integram câmera, processamento e comunicação sem fio em um único componente, de baixo custo e de consumo energético reduzido. A reCamera, lançada pela Seeed Studio em 2025 [Seeed Studio 2025], exemplifica essa tendência ao combinar câmera, NPU dedicada com suporte à quantização INT8 e conectividade Wi-Fi em um módulo alimentado por USB 5 V, com consumo máximo de 5 W e custo de aproximadamente USD 50. Apesar do seu potencial, dispositivos dessa categoria carecem de avaliação independente, em especial de estudos comparativos com plataformas embarcadas consolidadas, como a família Jetson da NVIDIA [Ammar et al. 2024, Alqahtani et al. 2024].

Este trabalho endereça essa lacuna por meio de uma avaliação comparativa entre três plataformas: reCamera, Jetson Xavier NX e desktop, com ênfase na identificação do melhor *pipeline* de parâmetros de inferência para a reCamera. A metodologia adota medição elétrica real por meio de um wattímetro externo e o modelo YOLOv11n exportado em diferentes formatos de quantização para cada plataforma. As principais contribuições deste trabalho são:

- Primeira avaliação científica independente da reCamera (2025) em comparação com dispositivos de borda consolidados e ambiente desktop;
- Protocolo experimental com repetições controladas e análise de variabilidade por meio de Mean Absolute Error (Erro Médio Absoluto) (MAE) e Viés médio de detecção (detecção – anotação) (BIAS) segmentados por faixa Detection, Observation, Recognition, Identification (DORI);
- *Pipeline* multicritério de seleção de parâmetros de inferência para a reCamera;
- Adoção do modelo DORI [International Electrotechnical Commission 2015,

Axis Communications 2023] como critério de avaliação de desempenho em dispositivos de borda para aplicações de vigilância urbana.

O restante deste artigo está organizado da seguinte forma: A Seção 2 discute os trabalhos relacionados que contextualizam e motivam este estudo. A Seção 3 detalha a metodologia experimental adotada. A Seção 4 apresenta e discute os principais resultados obtidos. Por fim, a Seção 5 sintetiza as conclusões e aponta direções para trabalhos futuros.

2. Trabalhos relacionados

A literatura recente sobre visão computacional na borda organiza-se em três eixos principais: comparação de plataformas de hardware, eficiência energética e técnicas de otimização, e *benchmarks* com protocolos de avaliação.

[Ammar et al. 2024] demonstram que apenas a análise conjunta de latência, acurácia e consumo fornece base sólida para seleção de plataforma em aplicações de visão em tempo real, sendo insuficiente a avaliação isolada de Frames Per Second (FPS). [González et al. 2025] comparam sete dispositivos de borda, Jetson Orin, Google Coral, Intel NCS2 e FPGA, em inferência de redes convolucionais, evidenciando variação considerável de latência, consumo e precisão entre arquiteturas, mas sem incluir dispositivos ultracompactos com câmera integrada nem pipeline orientado à vigilância.

[Alqahtani et al. 2024] realizam uma avaliação dos modelos YOLOv8, Efficient-Det Lite e Single Shot MultiBox Detector (SSD) em Raspberry Pi 3/4/5, tanto com uso ou não dos aceleradores Graphics Processing Unit (GPU) e Jetson Orin Nano, reunindo consumo via wattímetro, duração da inferência e mean Average Precision (mAP). Os autores concluem que o Jetson Orin Nano é o mais rápido e eficiente em termos de energia, eficiente, mas consome mais em relação aos outros dispositivos em modo ocioso.

[Lema et al. 2024] realizam comparação quantitativa de YOLOv3/v5/YOLOX em cinco dispositivos, RTX 2080Ti, Jetson Nano, Jetson AGX Xavier, i7-9700K e Google Coral, com 20 tamanhos de entrada via PyTorch, TensorRT e TFLite, medindo consumo por wattímetro. A metodologia é próxima ao presente trabalho, porém sem dispositivos ultracompactos de nova geração, sem protocolo de repetição e sem análise DORI. [Choe et al. 2023] demonstram em *benchmark* da família Jetson que o TensorRT acelera inferência em até 4× e reduz uso de Central Processing Unit (CPU)/memória pela metade, com todas as plataformas consumindo em média mais de 80% dos recursos de GPU durante a inferência.

No contexto de pipelines voltados para a quantização e a otimização de inferência, [Ferraz et al. 2025] analisam a quantização de modelos YOLOv8 nos formatos Floating Point 32-bit (FP32), Floating Point 16-bit (FP16) e Integer 8-bit (quantização inteira de 8 bits) (INT8) utilizando os dispositivos Jetson Orin Nano, Jetson Orin NX e Raspberry Pi 5, com medições de consumo utilizando um wattímetro. Os resultados mostram que a quantização em INT8 consegue diminuir o consumo em até 44% em comparação ao FP32, mantendo uma acurácia competitiva, o que é uma evidência relevante para entender o pipeline aplicado na reCamera.

[Desislavov et al. 2023] analisam tendências de consumo em inferência de *deep learning*, estimando que 70 a 80% do custo energético ao longo da vida de um modelo

ocorrem na fase de inferência, não no treinamento, reforçando a relevância de *benchmarks* energéticos em implantações de borda. [Liang et al. 2022] propõem Edge You Only Look Once (YOLO) para cooperação borda e nuvem em veículos autônomos, demonstrando que modelos compactos executando localmente reduzem latência e consumo energético sem sacrifício significativo de acurácia. [Mittal 2024] complementa essa visão com uma revisão sistemática dos modelos de detecção em dispositivos de borda, abrangendo variantes YOLO nano/tiny, MobileNet e SqueezeNet.

Benchmarks consolidados com MOT17 [Milan et al. 2016] e MOT20 [Dendorfer et al. 2020] fornecem base robusta para detecção e rastreamento de pedestres em cenas urbanas desafiadoras. O MOT17, utilizado neste trabalho, inclui anotações de alta qualidade com múltiplos detectores de base, sendo referência para avaliação robusta em condições adversas de oclusão e densidade (quantidade de pessoas). A maioria dos estudos, entretanto, emprega tais bases para comparação algorítmica sem incorporar medições elétricas reais nem análise de variabilidade por repetição [Alqahtani et al. 2024, Lema et al. 2024].

A Tabela 1 resume as particularidades de cada trabalho citados acima. Nenhuma das pesquisas anteriores integra ao mesmo tempo os quatro elementos fundamentais desta proposta: um dispositivo ultracompacto com Neural Processing Unit (NPU) Reduced Instruction Set Computer V (RISC-V) INT8, um protocolo de múltiplas repetições com avaliação de variabilidade MAE e BIAS, dois conjuntos de dados com níveis de dificuldade variados e a estratificação conforme o critério DORI.

Tabela 1. Comparativo entre trabalhos relacionados e o presente trabalho.

Trabalho	Wattímetro real	Ultra-compacto	2 Data-sets	Pipeline otimizado	Aplicação	Métricas
[Ammar et al. 2024]	✓	✗	✗	✗	Visão em tempo real	Latência, acurácia, W
[González et al. 2025]	✗	✗	✗	✗	Inferência geral	Latência, consumo, erro
[Alqahtani et al. 2024]	✓	✗	✗	✗	Detecção de objetos	mAP, FPS, W
[Lema et al. 2024]	✓	✗	✗	✗	Detecção de objetos	AP, FPS, W
[Choe et al. 2023]	✗	✗	✗	✗	Detecção 3D (robótica)	FPS, CPU, memória
[Mao et al. 2024]	✗	✗	✗	✗	Edge AI (survey)	Consumo teórico
[Liang et al. 2022]	✗	✗	✗	✓	Veículos autônomos	FPS, acurácia
[Ferraz et al. 2025]	✓	✗	✗	✓	Detecção de objetos	FPS, W, mAP, J/inf.
[Liu et al. 2023]	✗	✗	✗	✓	Detecção de objetos	mAP, tam. modelo
Este trabalho	✓	✓	✓	✓	Vigilância de pedestres	Exact Match, MAE, BIAS, FPS, W, DORI

3. Metodologia

Esta seção detalha os procedimentos metodológicos adotados para avaliar a eficiência energética e o desempenho das diferentes plataformas na tarefa de detecção de pessoas. Inicialmente, é descrito o arranjo experimental geral, definindo os parâmetros do modelo e as resoluções avaliadas. Na sequência, detalham-se as configurações específicas do *pipeline* de inferência e de medição elétrica para cada dispositivo: Jetson Xavier NX, Desktop GPU e reCamera. Por fim, são apresentadas as métricas de avaliação adotadas, seguidas pela explicação de como a norma DORI foi incorporada para estratificar os resultados, vinculando a qualidade das detecções a distâncias operacionais em cenários reais de vigilância.

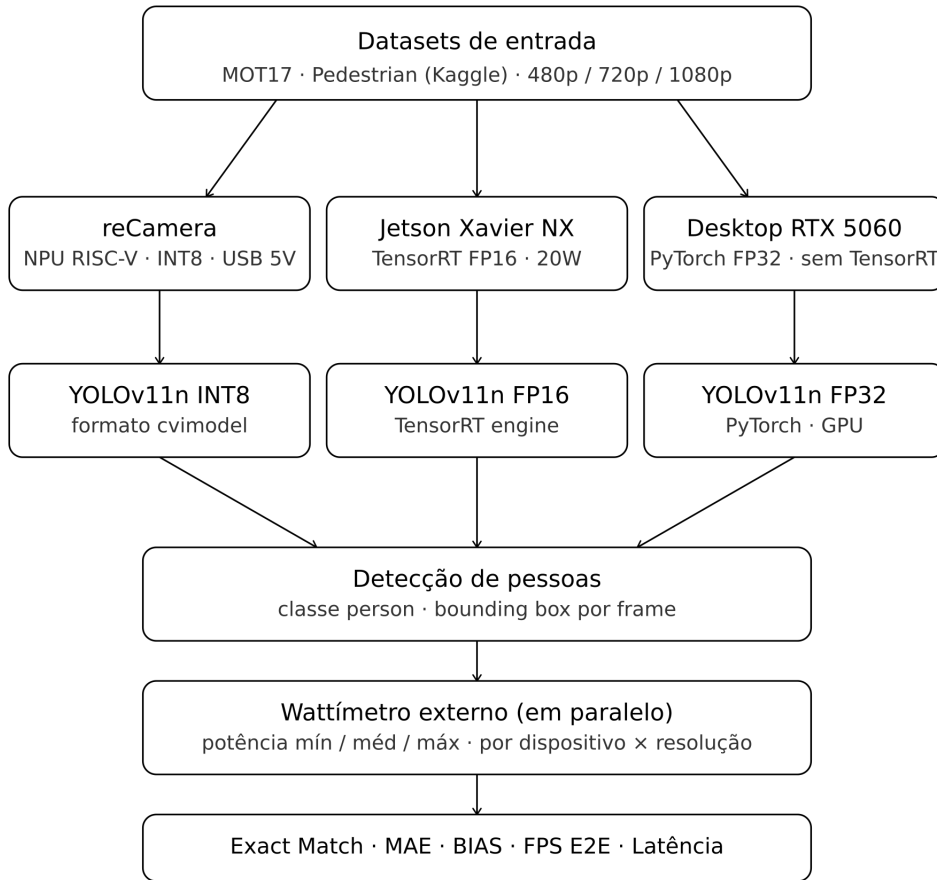


Figura 1. Pipeline de avaliação por dispositivo.

3.1. Experimento

O experimento foi projetado para garantir comparabilidade entre plataformas e de variabilidade. Cada configuração de parâmetros (dispositivo \times resolução \times dataset) foi executada cinco vezes de forma independente, com duração de um minuto por execução, totalizando cinco minutos de inferência por configuração.

A *threshold* de confiança de detecção foi fixado em 0,45 para todos os dispositivos, e apenas a classe *person* foi considerada. Esse limiar equilibra taxa de detecção com taxa de falsos positivos, seguindo prática comum na literatura de detecção de pedestres [Zhao et al. 2023]. Resoluções avaliadas: 480p (640 \times 480), 720p (1280 \times 720) e 1080p (1920 \times 1080). Modelos YOLOv11n foram exportados separadamente para cada resolução.

3.2. Configuração por Dispositivo

A Figura 1 apresenta o pipeline experimental adotado, ilustrando o fluxo desde os datasets de entrada até as métricas de saída, passando pela inferência em cada dispositivo com seu respectivo modelo e formato quantizado, mostrando o consumo elétrico capturado em paralelo por wattímetro externo e as detecções avaliadas pelas métricas estratificadas por faixa DORI.

O Jetson Xavier NX usa modo de energia: 20W (6 núcleos de CPU ativos). Modelo YOLOv11n exportado via TensorRT FP16, essa quantização é escolhida por manter o meio termo entre velocidade e precisão, o que torna o Jetson um dispositivo de borda a nível intermediário. A inferência é executada via *script* Python personalizado que processa os *frames* do dataset, registra *timestamps*, contagem de detecções e seus tamanhos em pixel por pessoa. O consumo elétrico é capturado por wattímetro externo em paralelo.

No Desktop, foi usado o modelo YOLOv11n em PyTorch FP32 na GPU NVIDIA RTX 5060. *Script* Python lê *frames* sequencialmente, registra *timestamps* e detecções por *frame*. Wattímetro externo monitora o sistema completo. A ausência de otimização TensorRT é intencional para representar o teto de acurácia FP32 como referência de comparação.

Já a reCamera, o pipeline utiliza exclusivamente a câmera embutida como fonte de entrada para as imagens dos datasets. As mesmas foram apresentadas em monitor de 34 polegadas a 1440p, com a câmera posicionada a 50 cm da tela. O brilho foi calibrado para nível médio a fim de minimizar saturação e reflexos, garantindo que as capturas mantenham resolução e proporção compatíveis com as demais plataformas, ou seja, o tamanho real da imagem, sendo a única alternativa viável para avaliação controlada do dispositivo sem modificações de firmware, analogia baseada em [Hu et al. 2016]. O modelo YOLOv11n é compilado no formato `cvimodel INT8` e seus resultados são salvos na memória do dispositivo. O consumo também é monitorado por wattímetro externo.

A reCamera¹ [Seed Studio 2025] é um dispositivo ultracompacto lançado pela Seed Studio em 2025. Combina câmera integrada, NPU dedicada para inferência INT8, processador RISC-V, conectividade WiFi e alimentação USB 5V com máximo de 5W em um único módulo com dimensões reduzidas (4cm x 4cm x 3,6cm). O ambiente de desenvolvimento é baseado no Node-RED v0.2.2, que simplifica o pipeline câmera → processamento → transmissão. O modelo YOLOv11n é exportado no formato `cvimodel INT8` específico para o NPU do dispositivo.

3.3. Métricas de Avaliação

As métricas de Exact Match, MAE e BIAS não apresentam desvio padrão nas tabelas por serem determinísticas: visto que o modelo YOLOv11n é fixo e os frames são processados sequencialmente a partir de imagens estáticas, as cinco repetições independentes produziram resultados idênticos em todas as configurações, confirmando a estabilidade do pipeline de inferência em cada plataforma. O desvio padrão é apresentado apenas para as medições de consumo de energia elétrica, pois essa é a única grandeza efetivamente sujeita a variações ambientais entre diferentes execuções.

- **Exact Match (%)**: porcentagem de *frames* em que a contagem de detecções coincide exatamente com o número anotado;
- **MAE**: desvio médio entre contagem detectada e anotada por faixa;
- **BIAS**: diferença média (detecção – anotação), indicando detecção positiva ou negativa;
- **FPS**: taxa de quadros;
- **Latência de inferência (ms)**: média por configuração;

¹<https://www.seedstudio.com/reCamera-2002w-64GB-p-6249.html>

- **Consumo (W):** mínima, média e máxima por wattímetro externo em tempo de inferência por conjunto de configuração.

O **Exact Match** é adotado como métrica primária por ser diretamente alinhado ao objetivo deste trabalho: avaliar se cada dispositivo estima corretamente a quantidade de pessoas presentes em cada imagem. Diferentemente do mAP, que penaliza desvios de Intersection over Union (IoU) entre bounding boxes detectadas e anotadas, sendo mais adequado para tarefas de localização precisa, o Exact Match captura diretamente a acurácia de contagem, que é o critério operacional relevante em sistemas de vigilância baseados em densidade.

3.4. Aplicação do Modelo DORI

O padrão DORI, definido pela norma IEC EN62676-4 [International Electrotechnical Commission 2015], especifica densidades de pixels por metro (px/m) necessárias para cada nível operacional: 25 px/m (Detecção), 62,5 px/m (Observação), 125 px/m (Reconhecimento) e 250 px/m (Identificação). Para converter essas densidades em limiares de tamanho de *bounding box* em pixels, informação diretamente observável nas saídas do modelo, adota-se a altura corporal de referência de 1,7 m pela classe `Pessoa`, valor amplamente utilizado em sistemas de vigilância [Axis Communications 2023]. O limiar de cada faixa é dado por:

$$t_{\text{faixa}} = d_{\text{faixa}} \times h_{\text{ref}} \quad (1)$$

onde d_{faixa} é a densidade (px/m) e $h_{\text{ref}} = 1,7$ m. Aplicando a Equação 1: $25 \times 1,7 \approx 42$ px (D), $62,5 \times 1,7 \approx 107$ px (O), $125 \times 1,7 \approx 213$ px (R) e $250 \times 1,7 = 425$ px (I), conforme a Tabela 2. Esse mapeamento vincula a qualidade da detecção a distâncias operacionais reais de vigilância, permitindo avaliar cada dispositivo em condições comparáveis às de sistemas de vigilância em campo de uso.

Tabela 2. Mapeamento das faixas DORI para tamanho de bounding box (altura de referência: 1,7 m).

Faixa	Tamanho (px)	Distância estimada	Nível operacional
D	42–107	Longa	Detecção
O	107–213	Média-longa	Observação
R	213–425	Média-curta	Reconhecimento
I	425+	Proximidade	Identificação

Métricas de *Exact Match*, MAE e BIAS são calculadas separadamente por faixa, permitindo identificar em quais condições operacionais cada dispositivo/resolução apresenta melhor desempenho.

4. Resultados

Esta seção apresenta os resultados organizados em quatro partes: consumo energético, Dataset Pedestrian, Dataset MOT17 e Considerações Finais. Para cada dispositivo, são reportadas as métricas estratificadas por faixa DORI.

4.1. Consumo Energético

A Tabela 3 apresenta os dados de consumo para os três dispositivos medidos por wattímetro externo. A reCamera mantém 1,9 W em todas as resoluções, comportamento esperado para um NPU INT8 com carga fixa independente da dimensão de entrada. O Jetson Xavier NX varia de 7,5 W a 10,3 W conforme aumento de resolução, refletindo maior demanda de GPU com entradas maiores. O desktop oscila de 78,9 W a 112,7 W com picos de 117 W por conta da GPU sem otimização TensorRT. A razão Desktop/reCamera vai de $41\times$ em 480p a $59\times$ em 1080p; em operação contínua de 24 h, isso representa 45,6 Wh/dia para a reCamera contra até 2.705 Wh/dia para o Desktop.



(a) Pedestrian (Kaggle) - cenas simples, baixa densidade [Karthika and Saravanan 2020]. (b) MOT17 - cenas densas, oclusões frequentes [Dendorfer et al. 2021].

Figura 2. Exemplos representativos dos datasets utilizados.

Tabela 3. Consumo elétrico por dispositivo e resolução (wattímetro externo).

Dispositivo	Resolução	Mín (W)	Méd (W)	Máx (W)
reCamera	480-1080p	1,8	1,9	2,0
Jetson	480p	6,2	7,5	8,1
	720p	7,8	8,4	9,3
	1080p	9,2	10,3	13,6
Desktop	480p	71,3	78,9	117,1
	720p	97,4	102,3	107,5
	1080p	102,8	112,7	117,4

4.2. Dataset Pedestrian

O dataset Pedestrian² contém cenas simples com baixa densidade de pessoas e baixa sobreposição de *bounding boxes*, conforme ilustrado na Figura 2(a).

²<https://www.kaggle.com/datasets/karthika95/pedestrian-detection>

O Desktop opera com PyTorch FP32 sem otimização TensorRT, configuração deliberada que representa o teto de acurácia para referência, não o estado da arte para plataformas desktop. Ambos os dispositivos apresentam acurácia elevada nas faixas D e O, como visto na Tabela 4. O Desktop atinge 83,2% na faixa D com 480p a 43,5 FPS, enquanto o Jetson chega a 73,4% na mesma faixa com 27,7 FPS e menos da metade do consumo. A faixa I, relacionada a pessoas próximas, ≥ 425 px, apresenta queda expressiva com o aumento de resolução: no Jetson, 1080p reduz a faixa I de 56,6% para 12,7%, indicando sobredetecção por fragmentação de *bounding boxes* grandes. Nas faixas intermediárias O e R, o Desktop mantém acurácia mais estável, 66,7-66,9% em O e 57,3-67,5% em R, enquanto o Jetson apresenta queda mais sensível, especialmente em R, com 47,4% em 480p contra 22,0% em 1080p.

Tabela 4. Métricas por resolução e faixa DORI - Desktop e Jetson Xavier NX, Pedestrian.

Dispositivo	Res.	FPS	Lat. (ms)	D (%)	O (%)	R (%)	I (%)
Desktop	480p	43,5	19,4	83,2	66,9	67,5	39,0
	720p	36,3	23,8	80,2	66,9	67,5	43,0
	1080p	33,2	26,4	77,9	66,7	57,3	36,2
Jetson Xavier NX	480p	27,7	12,1	73,4	51,7	47,4	56,6
	720p	17,7	20,3	65,7	46,4	32,0	26,7
	1080p	11,0	24,8	66,4	48,1	22,0	12,7

A reCamera foi avaliada nos dois datasets com seis configurações cada: duas frequências de captura, 5 e 30 FPS, combinadas com as três resoluções. O FPS de inferência da NPU mantém-se em proximadamente 10 FPS com latência média de 100 ms em todas as resoluções e datasets, confirmando que o tempo de inferência não depende da dimensão de entrada, comportamento característico de aceleradores INT8 com pipeline fixo. O consumo permanece constante em 1,9 W independentemente da configuração, indicando que a NPU opera em regime de carga fixa e que variações de resolução ou FPS de entrada não alteram o perfil energético do dispositivo.

A reCamera apresenta desempenho estável entre as configurações, com 55,9-59,5% na faixa D e 28,3-34,4% nas demais faixas, como visto na Tabela 5. A configuração FPS5/720p destaca-se com 59,5% em D, enquanto FPS30/720p entrega 34,4% na faixa I, a maior taxa entre todos os pares de FPS/resolução para pessoas próximas. O uso de CPU permanece abaixo de 5,4% em todas as configurações, indicando que o processamento é efetivamente direcionado à NPU, com mínima interferência do processador RISC-V. O consumo de RAM estável entre 87-91 MB confirma que o pipeline não sofre pressão de memória, mesmo em 1080p, reforçando a viabilidade de execução contínua. Note-se que o dispositivo possui apenas 256 MB de RAM total.

4.3. Dataset MOT17

O MOT17³, que estende o MOT16 [Milan et al. 2016], contém imagens com oclusões frequentes e alta densidade de pessoas, constituindo cenário significativamente mais de-

³<https://motchallenge.net/data/MOT17/>

Tabela 5. Métricas por resolução e faixa DORI - reCamera, Pedestrian.

Config.	D (%)	O (%)	R (%)	I (%)	CPU (%)	RAM (MB)
FPS5/480p	57,2	32,4	25,2	32,2	4,6	87,8
FPS5/720p	59,5	32,2	26,0	30,4	4,6	89,3
FPS5/1080p	55,3	33,1	22,8	28,3	5,3	90,3
FPS30/480p	57,0	33,0	26,8	33,8	4,5	88,7
FPS30/720p	56,2	34,0	27,3	34,4	5,4	89,6
FPS30/1080p	55,9	34,1	22,3	33,5	4,5	90,2

Consumo: 1,9 W constante. Temperatura: 46,0-46,8 °C. FPS inferência: ≈ 10 . Latência: ≈ 100 ms.

safiador que o Pedestrian, como demonstrado na Figura 2(b).

A acurácia cai consideravelmente em ambos os dispositivos em relação ao Pedestrian, Tabela 6. O Desktop atinge até 73,5% na faixa I com 480p, e o Jetson até 47,3% na faixa I com 720p. O BIAS é negativo em quase todas as configurações, indicando subdetecção em cenas de alta sobreposição de *bounding boxes*. O aumento de resolução tem efeito positivo na faixa D do Jetson: 1080p melhora D de 22,0% para 32,2%, mostrando que maior resolução auxilia a detecção de pessoas distantes em cenas densas. O efeito oposto ocorre na faixa I do Jetson em 1080p, que cai de 47,3% para 28,0%: em cenas densas, a maior resolução amplifica as sobreposições de *bounding boxes*, aumentando a taxa de subdetecção.

Tabela 6. Métricas por resolução e faixa DORI - Desktop e Jetson Xavier NX, MOT17.

Dispositivo	Res.	FPS	Lat. (ms)	D (%)	O (%)	R (%)	I (%)
Desktop	480p	36,9	21,3	43,6	40,5	44,7	73,5
	720p	32,4	25,7	45,1	46,6	44,3	73,1
	1080p	30,2	27,1	46,2	53,8	49,6	49,6
Jetson Xavier NX	480p	26,0	10,0	15,5	22,3	18,9	25,0
	720p	18,3	13,7	22,0	30,7	23,9	47,3
	1080p	10,3	25,2	32,2	14,8	20,8	28,0

Os resultados da reCamera no MOT17 são inferiores, diretamente relacionados à complexidade das cenas, como mostrado na Tabela 7. A acurácia na faixa D fica entre 20,1% e 25,8%, e na faixa I entre 21,1% e 28,6%. O BIAS é negativo em todas as faixas, chegando a $-2,05$ na faixa I, indicando subdetecção. A NPU INT8 perde detecções em cenas com múltiplas sobreposições. A temperatura sobe para 50-51 °C, acima dos 46-47 °C registrados no Pedestrian, refletindo maior atividade de inferência por frame mais denso, ainda dentro do limite seguro de 70 °C especificado pelo fabricante.

A queda de desempenho entre Pedestrian e MOT17 é maior na reCamera, 59,5% \rightarrow 25,8% na faixa D, do que no Desktop, 83,2% \rightarrow 43,6% e no Jetson, 73,4% \rightarrow 22,0%, sugerindo que a quantização INT8 do `cvimodel` é mais sensível à complexidade visual das cenas do que as implementações FP16 e FP32. Esse comportamento é esperado em

Tabela 7. Métricas por resolução e faixa DORI - reCamera, MOT17.

Config.	D (%)	O (%)	R (%)	I (%)	CPU (%)	RAM (MB)
FPS5/480p	20,1	14,8	15,2	24,6	4,2	88,1
FPS5/720p	22,3	13,3	14,4	24,3	4,7	89,2
FPS5/1080p	23,1	15,2	16,3	21,1	4,6	90,9
FPS30/480p	24,6	17,0	14,4	24,3	4,0	89,3
FPS30/720p	25,4	14,8	17,8	28,6	4,2	87,7
FPS30/1080p	25,8	16,7	16,7	25,1	5,0	86,9

Consumo: 1,9W constante. Temperatura: 50,2–51,2 °C. FPS inferência: ≈ 10 . Latência: ≈ 100 ms.

implementações de precisão reduzida sem mecanismo de recalibração por cena: regiões com múltiplas sobreposições produzem ativações intermediárias que, após quantização, ficam abaixo do limiar de confiança e são descartadas, amplificando a subdetecção em cenas densas independentemente de FPS ou resolução. A Tabela 8 apresenta o BIAS médio por faixa DORI para a reCamera nos dois datasets. O MAE permanece estável em todas as configurações, entre 0,7 e 1,2 no Pedestrian e entre 2,0 e 2,2 no MOT17, indicando que o limitador é a capacidade do modelo quantizado, não os parâmetros de entrada. O BIAS no Pedestrian é negativo na faixa D, $-0,3$ a $-0,4$, e aproxima-se de zero nas faixas O e R, com leve tendência positiva a 30 FPS; na faixa I recua a $-0,8$, possivelmente por fragmentação de *bounding boxes* grandes. No MOT17, o BIAS é negativo em todas as faixas, atingindo $-2,1$ na faixa I, evidenciando que a quantização INT8 amplifica a subdetecção em cenas densas com oclusão, independentemente de FPS ou resolução.

Tabela 8. BIAS médio por faixa DORI - reCamera, Pedestrian e MOT17.

Config.	D	O	R	I
<i>Pedestrian</i>				
FPS5/480p	$-0,3$	$+0,0$	$+0,2$	$-0,8$
FPS5/720p	$-0,4$	$-0,1$	$+0,3$	$-0,7$
FPS5/1080p	$-0,3$	$+0,0$	$+0,2$	$-0,7$
FPS30/480p	$-0,3$	$+0,1$	$+0,2$	$-0,8$
FPS30/720p	$-0,4$	$+0,0$	$+0,3$	$-0,8$
FPS30/1080p	$-0,3$	$-0,0$	$+0,4$	$-0,8$
<i>MOT17</i>				
FPS5/480p	$-0,7$	$-0,4$	$-1,3$	$-2,0$
FPS5/720p	$-0,7$	$-0,4$	$-1,4$	$-2,1$
FPS5/1080p	$-0,6$	$-0,3$	$-1,4$	$-2,1$
FPS30/480p	$-0,7$	$-0,3$	$-1,3$	$-2,0$
FPS30/720p	$-0,8$	$-0,4$	$-1,3$	$-2,0$
FPS30/1080p	$-0,6$	$-0,0$	$-1,3$	$-2,1$

4.4. Considerações Finais

Os resultados confirmam três perfis distintos de dispositivo. O Desktop lidera em acurácia, atingindo até 83,2% de *Exact Match* na faixa D do Pedestrian em 480p e

43,5 FPS, porém com consumo de 79-113 W. O Jetson Xavier NX oferece equilíbrio entre acurácia e consumo, 7,5-10,3 W, sendo adequado para borda embarcada de médio custo. A reCamera destaca-se pelo consumo constante de 1,9 W, sendo a única opção viável para monitoramento com infraestrutura elétrica restrita. Em termos de custo de aquisição, a reCamera está disponível por USD 50, o Jetson Xavier NX por aproximadamente USD 500 e o Desktop com RTX 5060 por aproximadamente USD 2.500, diferenças de $10\times$ e $50\times$ em relação à reCamera, respectivamente, reforçando seu posicionamento para implantações em larga escala. A Tabela 9 resume os dados médios dos três dispositivos.

Tabela 9. Síntese comparativa dos três dispositivos.

Dispositivo	Dataset	Match méd. (%)	FPS méd.	W méd.	Config. recomendada
Desktop	MOT17	54,6	33,2	95,4	480p — máximo desempenho
	Pedestrian	67,6	37,7	95,4	
Jetson Xavier NX	MOT17	27,0	18,2	8,7	480p — equilíbrio borda
	Pedestrian	57,9	18,8	8,7	
reCamera	MOT17	19,7	10	1,9	FPS30/720p
	Pedestrian	36,2	10	1,9	FPS5/720p ou FPS30/480p

A análise DORI estratificada revelou comportamentos mascarados por métricas globais: a faixa I beneficia todas as plataformas com maior *Exact Match*, a faixa D é a mais sensível à resolução de entrada e o BIAS negativo comum no MOT17 confirma subdetecção em cenas densas com sobreposição de pedestres. Para a reCamera, o pipeline ótimo prioriza acurácia operacional mínima na faixa D com consumo constante de 1,9 W e 720p a 5 FPS oferecendo a melhor acurácia em D, 59,5%, para cenas de baixa densidade, e 480p a 30 FPS é o padrão recomendado para ambientes com fluxo contínuo de pessoas. A resolução 1080p não se justifica em nenhum dispositivo, pois não apresenta ganho de acurácia proporcional ao custo computacional adicional. Para cenas densas como o MOT17, nenhuma configuração da reCamera supera 25,8% na faixa D, indicando que o modelo quantizado em INT8 é o principal limitante nesse cenário, não os parâmetros de entrada.

Uma limitação específica da avaliação da reCamera decorre de uma restrição arquitetural do dispositivo: o firmware atual não suporta injeção direta de *stream* de vídeo via código ou protocolo RTSP, exigindo que as imagens do dataset sejam apresentadas em monitor externo para captura pela câmera integrada. Embora os parâmetros de brilho e posicionamento tenham sido calibrados para minimizar artefatos, esse método pode introduzir perdas de contraste e saturação que não ocorrem nas demais plataformas, que processaram as imagens diretamente da memória.

5. Conclusões e trabalhos futuros

Este artigo apresentou avaliação comparativa entre reCamera (INT8), Jetson Xavier NX (TensorRT FP16) e Desktop com RTX 5060 (PyTorch FP32) para detecção de pessoas na borda, experimento unificado com cinco repetições por configuração, wattímetro externo e estratificação por faixas DORI. Os três dispositivos confirmaram perfis complementares de desempenho, consumo e custo: o Desktop como teto de acurácia (até 83,2% na

faixa D), o Jetson como alternativa de borda de médio custo e a reCamera como única opção viável sob restrição elétrica severa, operando a 1,9 W constante, até 59× menor que o Desktop e 50× mais barata. O pipeline multicritério resultante indica 720p a 5 FPS para cenas de baixa densidade e 480p a 30 FPS para implantações em tempo real, ambos mantendo o consumo de 1,9 W.

Como trabalhos futuros, pretende-se: (i) avaliar outros métodos como rastreamento e identificação de objeto em tempo real na reCamera; (ii) realizar análise de sensibilidade ao limiar de confiança, especialmente para imagens densas como no MOT17; (iii) testar o pipeline em datasets de vigilância externos com variação de iluminação artificial; (iv) comparar modelos YOLOv11s e YOLOv11m na reCamera, avaliando o *trade-off* entre acurácia e consumo para versões maiores do modelo; (v) incorporar medição de latência de transmissão Wi-Fi para avaliação fim a fim do pipeline; e (vi) estender a avaliação para um pipeline distribuído da reCamera em conjunto à nuvem.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Disponibilidade de artefatos

As métricas utilizadas e resultados encontrados estão disponíveis neste link⁴

Referências

- Alqahtani, D. K., Cheema, A., and Toosi, A. N. (2024). Benchmarking deep learning models for object detection on edge computing devices. arXiv preprint arXiv:2409.16808.
- Ammar, O., Benyahya, S., and El Fkihi, S. (2024). Energy-efficient edge AI inference for real-time vision applications. *Sensors*.
- Axis Communications (2023). Densidade de pixels e DORI: atender aos requisitos operacionais no vídeo em rede. Technical Report T10176489/PT/M2.3/2305, Axis Communications AB.
- Bonomi, F., Milito, R., Zhu, J., and Addepalli, S. (2012). Fog computing and its role in the Internet of Things. In *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, MCC '12*, pages 13–16. ACM.
- Cao, K., Liu, Y., Meng, G., and Sun, Q. (2023). A survey of AI-enabled edge computing: architectures, applications and challenges. *IEEE Access*, 11:10240–10257.
- Choe, C., Choe, M., and Jung, S. (2023). Run your 3D object detector on NVIDIA Jetson platforms: a benchmark analysis. *Sensors*, 23(8):4005.
- Dendorfer, P., Osep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., and Leal-Taixé, L. (2021). Motchallenge: A benchmark for single-camera multiple target tracking: P. dendorfer et al. *International Journal of Computer Vision*, 129(4):845–881.

⁴<https://drive.google.com/drive/folders/1QPXqaqK52OW6piBjDfyW4d8SZHFkbBx6?usp=sharing>

- Dendorfer, P., Rezatofghi, H., Milan, A., Shi, J., Cremers, D., Reid, I., Roth, S., Schindler, K., and Leal-Taixé, L. (2020). Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv preprint arXiv:2003.09003*.
- Desislavov, R., Martínez-Plumed, F., and Hernández-Orallo, J. (2023). Compute and energy consumption trends in deep learning inference. *Patterns*, 4(1):100i.
- Dhar, S. (2012). From outsourcing to cloud computing: evolution of IT services. *Management Research Review*, 35(8):664–675.
- Ferraz, L. et al. (2025). A performance analysis of YOLO-based models for quantized object detection on edge devices. *arXiv preprint arXiv:2502.15737*.
- González, M. L. et al. (2025). Deep learning inference on edge: a preliminary device comparison. In *Intelligent Data Engineering and Automated Learning – IDEAL 2024*, volume 15346 of *Lecture Notes in Computer Science*, pages 247–258. Springer.
- Hu, Y., Liu, H., Pfeiffer, M., and Delbruck, T. (2016). Dvs benchmark datasets for object tracking, action recognition, and object recognition. *Frontiers in neuroscience*, 10:405.
- International Electrotechnical Commission (2015). IEC EN62676-4: video surveillance systems for use in security applications – Part 4: application guidelines. International standard, IEC.
- Karthika, N. J. and Saravanan, C. (2020). Addressing false positives in pedestrian detection. In *International Conference on Electronic Systems and Intelligent Computing (ESIC 2020)*.
- Lema, D. G., Usamentiaga, R., and García, D. F. (2024). Quantitative comparison and performance evaluation of deep learning-based object detection models on edge computing devices. *Integration*, 95:102127.
- Liang, S., Wu, H., Zhen, L., Hua, Q., Garg, S., Kaddoum, G., Hassan, M. M., and Yu, K. (2022). Edge YOLO: real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(12):25345–25360.
- Liu, Y., Zhang, H., et al. (2023). Q-YOLO: efficient inference for real-time object detection. *arXiv preprint arXiv:2307.04816*.
- Mao, Y., You, C., Zhang, J., Huang, K., and Letaief, K. B. (2024). Green edge AI: a contemporary survey. *Proceedings of the IEEE*, 112(7):880–918.
- Milan, A., Leal-Taixé, L., Reid, I., Roth, S., and Schindler, K. (2016). MOT16: a benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*.
- Mittal, P. (2024). A comprehensive survey of deep learning-based lightweight object detection models for edge devices. *Artificial Intelligence Review*, 57(9):242.
- Seed Studio (2025). reCamera – intelligent edge vision device. Página do produto. Acesso em: 19 mar. 2026.
- Shi, W., Cao, J., Zhang, Q., Li, Y., and Xu, L. (2016). Edge computing: vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646.
- Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., and Chen, J. (2023). DETRs beat YOLOs on real-time object detection. *arXiv preprint arXiv:2304.08069*.