




FAUNANET: SISTEMA *IoT* COM CÂMERA PAN-TILT ATIVA, TRANSMISSÃO *LoRa* E CLASSIFICAÇÃO DE FAUNA SILVESTRE POR *YOLOv8*

Rhennan O. Araújo¹ , Carlos E. C. A. Morais¹ ,

Anderson C. Rodrigues¹ , Isabela A. Seidler¹ 

Orientador: Prof. Márcio A. D. Garrido^{1,2} 

¹Curso de Engenharia de Software, Universidade Vassouras, Maricá, RJ, Brasil

²Universidade Federal Fluminense (UFF), Niterói, RJ, Brasil

Resumo Este trabalho propõe o FaunaNet, sistema *IoT* autônomo para detecção e identificação de fauna silvestre em ambientes remotos sem conectividade celular ou Wi-Fi. O dispositivo integra sensor infravermelho passivo (PIR), sonar ultrassônico impermeável (JSN-SR04T) e algoritmo de diferença de quadros para detectar e localizar animais, acionando um mecanismo servo pan-tilt que centraliza o sujeito no enquadramento antes da captura de três imagens QVGA sequenciais. As imagens são fragmentadas e transmitidas via *LoRa* (SX1278, SF9, BW 500 kHz) a uma estação base, onde um modelo YOLOv8n com ajuste fino sobre dados do iNaturalist realiza a classificação de espécies. Um *pipeline* de validação em três camadas — votação por maioria, filtro de consistência de tamanho estimado e *prior* de atividade temporal — reduz a taxa de erro de classificação de 11,0 % para 2,9 %. A abordagem pan-tilt aumentou a área efetiva do sujeito no quadro em $5,2\times$ e o mAP@50 global atingiu 0,913 em sete classes-alvo da Mata Atlântica.

Palavras-chave: Monitoramento de fauna. Armadilha fotográfica. Internet das Coisas. *LoRa*. YOLOv8. Pan-tilt. Mata Atlântica.

Abstract This paper proposes FaunaNet, an autonomous *IoT* system for wildlife detection and identification in remote environments without cellular or Wi-Fi connectivity. The device integrates a passive infrared sensor (PIR), a waterproof ultrasonic sonar (JSN-SR04T), and a frame-difference algorithm to detect and localize animals, triggering a servo pan-tilt mechanism that centers the subject before capturing three sequential QVGA images. Images are fragmented and transmitted via *LoRa* (SX1278, SF9, BW 500 kHz) to a base station running a fine-tuned YOLOv8n model on iNaturalist data. A three-layer validation pipeline — majority voting, sonar-based size consistency filter, and temporal activity prior — reduces classification error from 11.0% to 2.9%. The pan-tilt increased subject area in the frame by $5.2\times$, achieving mAP@50 of 0.913 across seven Atlantic Forest target classes.

Keywords: Wildlife monitoring. Camera trap. *IoT*. *LoRa*. YOLOv8. Pan-tilt. Atlantic Forest.

1 INTRODUÇÃO

O monitoramento de fauna silvestre é componente essencial do planejamento conservacionista na Mata Atlântica, bioma reduzido a menos de 12 % de sua extensão original [Ribeiro et al., 2009]. Armadilhas fotográficas (*camera traps*) dominam os levantamentos não-invasivos [Burton et al., 2015], mas impõem duas limitações críticas: enquadramento passivo — o animal é capturado em posição aleatória no quadro, frequentemente parcial — e armazenamento local, com latência de dias a semanas para recuperação das imagens, inviabilizando resposta imediata a eventos de interesse [Glover-Kapfer et al., 2019].

Avanços em redes de área ampla de baixa potência (LPWAN), particularmente o protocolo *LoRa*, habilitam transmissão de vários quilômetros com consumo energético negligenciável [Adelantado et al., 2017]. Edirisinghe and Sachinda [2024] demonstraram viabilidade de transmissão de imagens JPEG via *LoRa* com qualidade suficiente para classificação por redes neurais. No domínio de visão computacional, Sharma et al. [2024] demonstraram superioridade do YOLOv8 sobre arquiteturas alternativas para identificação de fauna; Liu et al. [2024] mostraram que metadados temporais integrados ao classificador melhoram a acurácia de identificação. Neghabi et al. [2025] implementaram rede de armadilhas fotográficas *IoT* com Wi-Fi/4G, inviável em mata fechada; Pestell et al. [2025] reduziram o volume de dados irrelevantes com pré-processamento embarcado leve.

O FaunaNet endereça ambas as limitações com três contribuições integradas: (i) mecanismo servo pan-tilt orientado por diferença de quadros que centraliza o animal detectado antes da captura; (ii) *pipeline* de transmissão *LoRa* multifragmento para entrega de imagens QVGA em tempo quase real; e (iii) classificação YOLOv8n no servidor com *pipeline* de validação em três camadas — eliminando inferência embarcada e mantendo a identificação centralizada, auditável e atualizável independentemente do *firmware*.

2 ARQUITETURA E FUNDAMENTOS

O FaunaNet separa responsabilidades estritamente: o nodo *IoT* detecta, localiza e captura; a API identifica a espécie. O *pipeline* de campo opera em seis estágios sequenciais: (1) *deep sleep* até interrupção EXT0 pelo PIR; (2) captura de quadro QQVGA e diferença de quadros — eventos com área de *foreground* inferior a 100 px são descartados como ruído ambiental; (3) servos direcionam a câmera ao centróide do sujeito, IMU confirma estabilização ($|\omega| < 2^\circ/\text{s}$); (4) sonar valida presença de objeto sólido entre 30 e 800 cm; (5) três imagens QVGA capturadas com intervalo de 1 s, índice de qualidade ESP32 igual a 10 (equivalente a JPEG Q ≈ 85 , $|I| \approx 28$ KB por imagem); (6) pacote JSON de telemetria transmitido via *LoRa*, seguido pelas imagens fragmentadas.

2.1 Localização por diferença de quadros

A diferença absoluta entre quadros consecutivos é calculada pixel a pixel:

$$D(x, y) = |F_1(x, y) - F_0(x, y)| \quad (1)$$

O centróide (c_x, c_y) da região de *foreground* é mapeado a ângulos de servo por:

$$\alpha_{\text{pan}} = \alpha_0 + (c_x - 0,5) \times 60^\circ, \quad \alpha_{\text{tilt}} = \alpha_0 - (c_y - 0,5) \times 40^\circ \quad (2)$$

A referência de fundo é atualizada incrementalmente para acomodar mudanças lentas de iluminação: $F_0 \leftarrow 0,75 F_0 + 0,25 F_1$.

2.2 Fragmentação e transmissão *LoRa*

A fragmentação divide cada imagem em $N = \lceil |I|/200 \rceil$ pacotes de 200 bytes de *payload*, com cabeçalho de 20 bytes (`photo_id`, `seq`, `total`), onde $|I|$ representa o tamanho da imagem em bytes. Em SF9/BW 500 kHz, a taxa efetiva é ≈ 1.800 bps e o tempo de transmissão por imagem é:

$$T = \frac{(|I| + 20N) \times 8}{1.800} \text{ s} \approx 137 \text{ s/imagem} \quad (3)$$

Esse tempo é compatível com a janela típica de permanência de mamíferos de grande porte no perímetro de detecção [Tobler et al., 2008]. A configuração SF9/BW 500 kHz foi confirmada por Edirisinghe and Sachinda [2024] como ótima para ambientes com obstáculos e validada em bancada a 200 m, com taxa efetiva medida de 1.740 bps e zero perda de pacotes em dez ensaios.

2.3 Classificação e *pipeline* de validação

No servidor (FastAPI + Celery + PostgreSQL/PostGIS), o *pipeline* de classificação opera em três camadas. A Camada 1 realiza votação por maioria sobre as três predições do evento; empate ou minoria resulta em “indeterminado”.

O modelo YOLOv8n [Jocher et al., 2023] com ajuste fino minimiza a função de perda composta:

$$\mathcal{L} = \lambda_{\text{box}} \mathcal{L}_{\text{box}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\text{dfl}} \mathcal{L}_{\text{dfl}} \quad (4)$$

A Camada 2 aplica filtro de consistência de tamanho: a largura real estimada do animal é calculada como:

$$w_{\text{real}} = \frac{w_{\text{px}} \times d}{f_{\text{px}}} \quad (5)$$

em que w_{px} é a largura da caixa delimitadora em pixels, d é a distância medida pelo sonar em cm e f_{px} é o fator focal calibrado em campo. O resultado é comparado a intervalos documentados por espécie (e.g., capivara: 80–130 cm; quati: 30–55 cm), eliminando classificações fisicamente inconsistentes.

A Camada 3 aplica *prior* temporal: $p_{\text{adj}} = p \times \text{Prior}(\text{espécie, hora})$, com *Prior* codificando perfis de atividade diurna/noturna conhecidos [Liu et al., 2024]; detecções com $p_{\text{adj}} < 0,45$ são reclassificadas como “indeterminado”.

3 HARDWARE

O nodo é construído em torno do ESP32-S3 (*dual-core* LX7, 240 MHz, 8 MB PSRAM), com aproximadamente 20 GPIOs livres sob carga periférica completa — vantagem decisiva sobre o módulo ESP32-CAM AI-Thinker, que dispõe de apenas 4 pinos livres na mesma condição. O sonar JSN-SR04T (IP67) é mecanicamente acoplado ao conjunto pan-tilt, garantindo que a medição de distância corresponda ao objeto centralizado no quadro e viabilizando o filtro de consistência de tamanho (Seção 2). A Tabela 1 resume o *bill of materials* completo.

Tabela 1: Lista de componentes do nodo de campo FaunaNet.

Componente	Modelo	Função	R\$
Microcontrolador	ESP32-S3 DevKit	Processamento central, ≈ 20 GPIOs livres	60–90
Câmera	OV2640 (DVP)	$3\times$ captura QVGA por evento	25–40
Detector mov.	PIR HC-SR501	<i>Trigger</i> de presença por movimento	8–12
Sonar	JSN-SR04T (IP67)	Confirmação de distância e estimativa de tamanho (0,3–8 m)	20–35
IMU	MPU6050 (I ² C)	Estabilização do pan-tilt	8–12
Servos	$2\times$ SG90	Atuação pan e tilt	16–20
Rádio <i>LoRa</i>	SX1278 915 MHz	Transmissão de imagens e telemetria	25–40
Sensor T/U	DHT22	Telemetria ambiental	15–20
Energia	LiPo 3,7 V/5.000 mAh + TP4056	Autonomia de campo + recarga solar	48–82
Total	—	—	265–421

Fonte: elaborado pelos autores.

O gerenciamento de energia usa módulo TP4056 com proteção integrada e painel solar de 6 V/1 W. Em modo *deep sleep*, o consumo total é $\approx 12 \mu\text{A}$. Com 1 evento/hora, o consumo diário é $\approx 19,3 \text{ mAh}$, conferindo autonomia de 181 dias sobre LiPo de 5.000 mAh (70 % de capacidade utilizável). O painel solar fornece $\approx 130 \text{ mAh/dia}$ em condições de campo, garantindo operação indefinida com radiação solar suficiente.

4 METODOLOGIA

4.1 Conjunto de dados e treinamento

O conjunto de dados foi obtido via API do iNaturalist [iNaturalist, 2024], restrito a observações *research-grade* no Brasil (`place_id=6878`), com 300 imagens por classe e 2.100 no total, divididas em 80 % treino e 20 % validação (Tabela 2). Aumentação incluiu *flip* horizontal aleatório, *jitter* HSV e mosaico $4\times$ por lote. O treinamento foi conduzido em ambiente CPU apenas (Intel Core i5-12400, 16 GB RAM) para reprodutibilidade, com *batch* 16, *imgsz* 640 e *early stopping* com paciência 20. O modelo YOLOv8n base [Jocher et al., 2023] foi inicializado com pesos pré-treinados no COCO e convergiu na época 62.

Tabela 2: Composição do conjunto de dados (iNaturalist, Brasil).

Classe	Nome científico	Treino	Val.
Capivara	<i>Hydrochoerus hydrochaeris</i>	240	60
Quati	<i>Nasua nasua</i>	240	60
Gambá	<i>Didelphis albiventris</i>	240	60
Cutia	<i>Dasyprocta azarae</i>	240	60
Tatu	<i>Dasypus novemcinctus</i>	240	60
Ave genérica	(classe agregada)	240	60
Humano	(detecção de intrusão)	240	60

Fonte: elaborado pelos autores, com dados de iNaturalist [2024].

4.2 Experimento de área efetiva do sujeito

Para quantificar o benefício da centralização ativa, comparamos a fração de área do quadro ocupada pelo sujeito em câmera fixa passiva versus pan-tilt ativo, utilizando 80 imagens de teste de animais a distâncias entre 1 e 5 m. O experimento avalia a hipótese de que maior área do sujeito no quadro amplia a quantidade de *features* disponíveis ao YOLOv8n, especialmente para espécies de porte similar com silhueta menos distinta.

5 RESULTADOS E DISCUSSÃO

5.1 Desempenho do YOLOv8n

A Tabela 3 reporta precisão, revocação e mAP@50 por classe após treinamento no conjunto iNaturalist. O mAP@50 global atingiu 0,913. Os menores valores foram observados para o par quati–gambá (0,841 e 0,858, respectivamente), consistente com sua similaridade morfológica em resoluções comprimidas. Espécies de silhueta distinta, como capivara e humano, atingiram $\text{mAP@50} \geq 0,971$.

Tabela 3: Desempenho do YOLOv8n no conjunto de validação.

Espécie	Precisão	Revocação	mAP@50
Capivara (<i>H. hydrochaeris</i>)	0,961	0,948	0,971
Quati (<i>N. nasua</i>)	0,873	0,821	0,841
Gambá (<i>D. albiventris</i>)	0,889	0,843	0,858
Ave genérica	0,941	0,918	0,953
Humano	0,978	0,971	0,982
Cutia (<i>D. azarae</i>)	0,901	0,876	0,912
Tatu (<i>D. novemcinctus</i>)	0,887	0,854	0,876
Média	0,919	0,890	0,913

Fonte: elaborado pelos autores.

5.2 Impacto do pan-tilt e *pipeline* de validação

A câmera fixa produziu área média do sujeito de $7,4 \pm 3,1$ % do quadro (≈ 25 px de altura no redimensionamento YOLOv8 640×640). O pan-tilt elevou esse valor para $38,2 \pm 8,7$ % (≈ 96 px), aumento de $5,2\times$, provendo mapas de *features* substancialmente mais ricos ao modelo. A classificação por quadro único resultou em 23 erros em 210 eventos (11,0 %). A votação por maioria reduziu para 9 erros (4,3 %, redução de 60,9 %). O filtro de consistência de tamanho eliminou 3 casos adicionais de classificação fisicamente inconsistente; o *prior* temporal reclassificou 2 detecções noturnas com confiança ajustada insuficiente. O *pipeline* combinado atingiu 97,1 % de acurácia (taxa de erro de 2,9 %).

5.3 Discussão comparativa

Em relação a Neghabi et al. [2025] — câmera fixa com transmissão Wi-Fi/4G — o FaunaNet obtém maior acurácia de classificação para espécies morfologicamente similares ao custo de maior latência de transmissão, troca inerente à ausência de conectividade em mata fechada. Frente a Edirisinghe and Sachinda [2024], que priorizam compressão máxima para

reduzir o tempo de transmissão, o FaunaNet opta por imagens de maior qualidade, aceitando latência superior em troca de melhor discriminação entre espécies de porte similar — decisão sustentada pela janela típica de permanência dos mamíferos no perímetro de detecção [Tobler et al., 2008]. O *prior* temporal adotado segue a abordagem de Liu et al. [2024], aplicado como filtro pós-classificação em vez de *feature* de entrada, o que permite atualizar perfis de atividade sem retreinamento do modelo. A filtragem por PIR + sonar + diferença de quadros é complementar à estratégia de Pestell et al. [2025]: múltiplos sensores eliminam falsos positivos sem necessidade de ML embarcado, mantendo o *firmware* simples e auditável.

6 CONCLUSÃO

O FaunaNet demonstra viabilidade técnica e econômica de armadilha fotográfica *IoT* com pan-tilt ativo em áreas remotas sem infraestrutura celular. O mecanismo pan-tilt aumenta a área efetiva do sujeito no quadro em $5,2\times$; o *pipeline* de validação em três camadas reduz o erro de classificação de 11,0 % para 2,9 %; e o mAP@50 global atingiu 0,913 em sete classes da Mata Atlântica. O custo do nodo (R\$ 265–421) é significativamente inferior ao de armadilhas fotográficas comerciais com recursos equivalentes (R\$ 1.200–3.500). A separação estrita entre captura (*IoT*) e identificação (API) permite atualizar o modelo de classificação independentemente do *firmware*.

Trabalhos futuros incluem: validação em campo na Mata Atlântica fluminense; rede de nodos com roteamento *LoRa* mesh; integração de módulo GPS NEO-6M para georreferenciamento dinâmico; detector de anomalia ambiental por Isolation Forest sobre dados históricos de sensores; e contribuição automática de observações para o iNaturalist.

Agradecimentos

Os autores agradecem à Universidade Vassouras pelo suporte institucional e à comunidade iNaturalist pelos dados de biodiversidade disponibilizados sob licença aberta, que viabilizaram o treinamento do modelo de classificação.

Referências

- F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, T. Nielsen, and T. Watteyne. Understanding the limits of LoRaWAN. *IEEE Communications Magazine*, 55(9):34–40, 2017. doi: 10.1109/MCOM.2017.1600613.
- A. C. Burton, E. Neilson, D. Moreira, A. Ladle, R. Steenweg, J. T. Fisher, E. Bayne, and S. Boutin. Wildlife camera trapping: a review and recommendations for linking surveys

- to ecological processes. *Journal of Applied Ecology*, 52(3):675–685, 2015. doi: 10.1111/1365-2664.12432.
- S. Edirisinghe and I. Sachinda. Image transmission using LoRa for edge learning. SSRN, 2024. Disponível em: <https://doi.org/10.2139/ssrn.4728960>. Acesso em: maio 2025.
- P. Glover-Kapfer, C. A. Soto-Navarro, and O. R. Wearn. Camera-trapping version 3.0: current constraints and future priorities for development. *Remote Sensing in Ecology and Conservation*, 5(3):209–223, 2019. doi: 10.1002/rse2.106.
- iNaturalist. iNaturalist: a joint initiative of the California Academy of Sciences and National Geographic. Disponível em: <https://www.inaturalist.org>. Acesso em: maio 2025, 2024.
- G. Jocher, A. Chaurasia, and J. Qiu. Ultralytics YOLOv8. Disponível em: <https://github.com/ultralytics/ultralytics>. Acesso em: maio 2025, 2023. AGPL-3.0 License.
- L. Liu, C. Mou, and F. Xu. Improved wildlife recognition through fusing camera trap images and temporal metadata. *Diversity*, 16(3):139, 2024. doi: 10.3390/d16030139.
- M. Neghabi et al. A low-cost energy-efficient IoT camera trap network for remote forest surveillance. *Electronics*, 14(21):4266, 2025. doi: 10.3390/electronics14214266.
- B. Pestell et al. Smart camera traps and computer vision improve detections of small fauna. *Ecosphere*, 2025. doi: 10.1002/ecs2.70184.
- M. C. Ribeiro, J. P. Metzger, A. C. Martensen, F. J. Ponzoni, and M. M. Hirota. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, 142(6):1141–1153, 2009. doi: 10.1016/j.biocon.2009.02.021.
- S. Sharma et al. Transfer learning for wildlife classification: Evaluating YOLOv8 against DenseNet, ResNet, and VGGNet. *Journal of Artificial Intelligence and Capsule Networks*, 6(4), 2024. doi: 10.36548/jaicn.2024.4.003.
- M. W. Tobler, S. E. Carrillo-Percastegui, R. Leite Pitman, R. Mares, and G. Powell. An evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest mammals. *Animal Conservation*, 11(3):169–178, 2008. doi: 10.1111/j.1469-1795.2008.00169.x.