

Detecção de Ativos em Redes de Distribuição Elétrica com Imagens Veiculares Georreferenciadas

Dicson S. Rabelo¹, Jonathan N. Gois^{1,2}, Amaro A. de Lima¹, Alessandro R. L. Zachi¹, Wesley L. Passos^{1,2} e Gabriel M. Araujo¹

Resumo— Este trabalho avalia o uso de visão computacional para a detecção automática de ativos em redes de distribuição elétrica a partir de imagens veiculares georreferenciadas. Para isso, apresenta-se um sistema embarcado de aquisição, composto por câmeras de alta resolução, sensores de posicionamento e unidade de controle, desenvolvido para coletas em campo. A partir dos dados capturados, é construído um conjunto de imagens anotadas em seis classes de ativos. Um modelo YOLO26m é treinado e avaliado, alcançando precisão de 0,70, revocação de 0,34, mAP@0.5 de 0,39 e mAP@0.5:0.95 de 0,23 na etapa de validação, com desempenho substancialmente inferior em um conjunto de teste independente, avaliado por procedimento distinto. Os resultados caracterizam fatores que influenciam a detecção de ativos em cenas urbanas reais, como a distribuição heterogênea das classes, a presença de ativos maiores que o bloco de processamento e a complexidade visual do ambiente.

Palavras-Chave— Ativos elétricos, redes de distribuição elétrica, imagens georreferenciadas, visão computacional.

Abstract— This work evaluates the use of computer vision for automatic asset detection in electric distribution networks using georeferenced vehicle-based images. To this end, an embedded acquisition system composed of high-resolution cameras, positioning sensors, and a control unit is presented for field data collection. From the captured data, an image dataset annotated into six asset classes is constructed. A YOLO26m model is trained and evaluated, achieving a precision of 0.70, a recall of 0.34, an mAP@0.5 of 0.39, and an mAP@0.5:0.95 of 0.23 on the validation set, with substantially lower performance on an independent test set assessed under a distinct procedure. The results characterize factors that influence asset detection in real urban scenes, such as heterogeneous class distribution, assets larger than the processing tile, and visually complex backgrounds.

Keywords— Electric assets, electric distribution networks, georeferenced images, computer vision.

I. INTRODUÇÃO

A gestão de ativos em redes de distribuição elétrica é uma atividade essencial para concessionárias de energia, pois influencia diretamente a confiabilidade operacional, o planejamento de manutenção, a atualização cadastral e a redução de custos associados às inspeções em campo [1], [2], [3]. Em muitos casos, esse levantamento ainda depende de procedimentos manuais ou semiautomáticos, que exigem equipes especializadas,

deslocamentos frequentes e posterior organização das informações coletadas [2], [4]. Esse processo tende a ser oneroso, demorado e suscetível a inconsistências, especialmente em ambientes urbanos densos, nos quais diferentes elementos de infraestrutura compartilham o mesmo espaço visual [5], [6].

Nesse contexto, sistemas embarcados de aquisição de imagens em veículos têm se tornado uma alternativa promissora para a coleta recorrente de dados em campo [5], [7]. Quando combinados com sensores de posicionamento e técnicas de visão computacional, esses sistemas podem contribuir para a criação de bases georreferenciadas de ativos, apoiando a identificação, a localização e a atualização cadastral da rede [6], [7], [8]. No entanto, a detecção automática de ativos nesse domínio ainda envolve desafios relevantes, como objetos pequenos ou alongados, oclusões parciais, variações de iluminação, diferentes ângulos de captura, sobreposição entre redes elétricas e de telecomunicações e semelhança visual entre alguns ativos e o fundo urbano [5], [6], [9], [10], [11].

O presente trabalho investiga o uso de modelos de detecção de objetos para identificar ativos de redes de distribuição elétrica em imagens veiculares georreferenciadas capturadas em ambiente urbano. Para isso, construiu-se um conjunto de dados composto por 200 imagens obtidas por um sistema de captura embarcado desenvolvido no âmbito do projeto de P&D ANEEL PD-00382-0156/2022. As imagens foram anotadas manualmente considerando seis classes de interesse: ‘rede de baixa tensão’, ‘rede de média tensão’, ‘poste’, ‘iluminação pública’, ‘rede de telecomunicações’ e ‘transformador’. Em seguida, um modelo de detecção de objetos da família *you only look once* (YOLO) [12], [13] é treinado e avaliado.

Diferentemente de cenários controlados, as imagens utilizadas neste estudo apresentam alta resolução original, variação de enquadramento, fundo urbano complexo e distribuição heterogênea entre classes, aspectos frequentemente apontados como críticos em revisões sobre inspeção automatizada e detecção de pequenos objetos [1], [5], [9], [14].

Especificamente, as principais contribuições deste trabalho são:

- descrição de um fluxo de aquisição de imagens veiculares georreferenciadas para ativos de redes de distribuição elétrica;
- construção de um conjunto de dados anotado com múltiplas classes de ativos elétricos e de infraestrutura associada;
- avaliação experimental de um modelo de detecção de objetos para identificação automática desses ativos em imagens reais de campo; e

¹CEFET-RJ, Rio de Janeiro-RJ, ²PESC/COPPE/UFRJ, Rio de Janeiro-RJ. e-mails: dicson.rabelo@aluno.cefet-rj.br, {jonathan.gois, amaro.lima, alessandro.zachi, wesley.passos, gabriel.araujo}@cefet-rj.br. Este trabalho foi parcialmente financiado pelo Programa de Pesquisa, Desenvolvimento e Inovação (P&D) ANEEL, por meio da Light Serviços de Eletricidade S.A (PD-00382-0156/2022). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

- análise dos fatores que influenciam o desempenho do modelo em cenas urbanas reais.

II. TRABALHOS RELACIONADOS

No contexto de infraestrutura elétrica, diferentes estudos têm explorado técnicas de aprendizado de máquina e visão computacional para identificação de equipamentos, classificação de componentes e apoio à gestão de ativos [2], [4], [5], [6], [7]. Parte desses trabalhos concentra-se em ambientes controlados, como subestações, ou em imagens direcionadas para inspeção aérea de componentes específicos, como isoladores, torres e condutores [11], [15], [16]. Esses cenários diferem de imagens veiculares urbanas, nas quais há maior variação de perspectiva, oclusões, fundos complexos e coexistência de diferentes redes de infraestrutura [5], [6], [7]. A disponibilidade de bases públicas voltadas a redes de distribuição em ambientes urbanos permanece limitada, tanto em tamanho quanto em diversidade [1], [5].

Modelos da família YOLO têm sido amplamente utilizados em detecção de objetos em infraestrutura elétrica devido ao equilíbrio entre desempenho e custo computacional [9], [11], [16], sendo atraentes tanto para aplicações embarcadas quanto para processamento de grandes volumes de imagens. Entretanto, em objetos pequenos, finos ou alongados, como cabos e luminárias, o redimensionamento das imagens pode reduzir detalhes visuais relevantes, dificultando a detecção [9], [14]. Diferentemente dos trabalhos voltados a ambientes controlados, este artigo avalia a detecção de múltiplas classes de ativos em imagens veiculares georreferenciadas capturadas em campo, caracterizando os fatores que influenciam o desempenho do detector em cenas urbanas reais [1], [5], [6], [9].

III. SISTEMA DE AQUISIÇÃO E DATASET

A. Sistema de aquisição veicular

O conjunto de imagens utilizado neste trabalho foi obtido por meio de um sistema de captura veicular desenvolvido para o levantamento de ativos em redes de distribuição elétrica. O sistema é composto por quatro câmeras GoPro HERO 11 instaladas em um veículo, sensores de posicionamento global (GNSS/GPS), interface de velocidade veicular via *on-board diagnostics II* (OBD2) e uma unidade embarcada de controle e armazenamento, permitindo a aquisição de imagens georreferenciadas da infraestrutura urbana com cobertura visual das classes de ativos consideradas neste trabalho. As imagens são capturadas originalmente com resolução de 5.568×4.872 pixels, e as informações de posicionamento associadas às capturas permitem a indexação espacial das imagens. A Figura 1 apresenta a visão geral do sistema e a Figura 2 o sistema físico instalado no veículo.

B. Georreferenciamento das imagens

O georreferenciamento das imagens é realizado a partir de leituras periódicas de um sensor GNSS/GPS. Neste trabalho, utilizou-se o módulo U-blox Zed-F9P, capaz de operar com diferentes constelações de satélites, incluindo GPS, GLO-NASS, Galileo e BeiDou, com frequência de atualização de

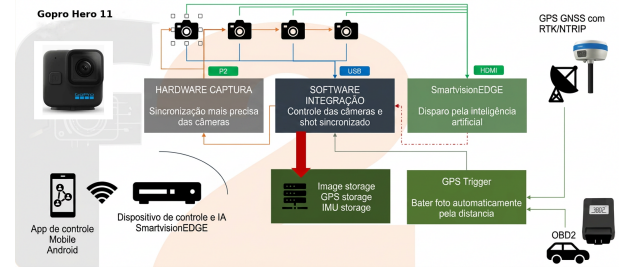


Fig. 1. Conceito do Sistema de Captura de ativos elétricos de distribuição de energia elétrica.

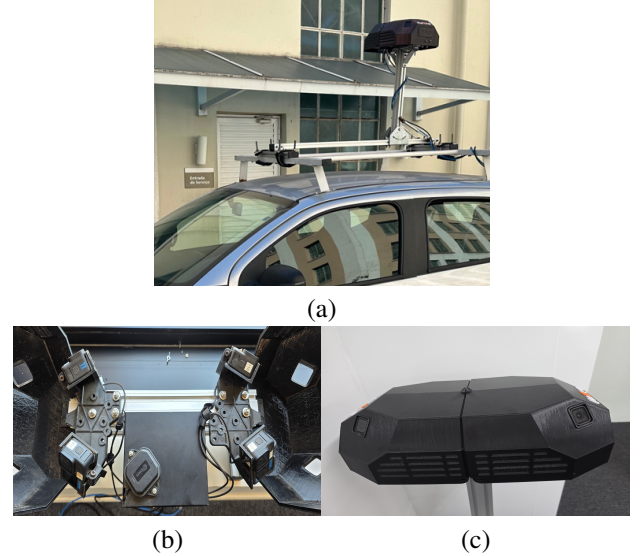


Fig. 2. Sistema de captura veicular. (a) Carenagem instalada no teto do veículo da concessionária; (b) interior da carenagem com câmeras, receptor GNSS e suportes impressos; e (c) caixa de proteção fechada.

até 10 Hz. Como as capturas fotográficas podem ocorrer em instantes distintos daqueles em que as leituras GNSS são disponibilizadas, o sistema estima a posição do veículo no instante de aquisição de cada imagem por meio de interpolação linear entre coordenadas adjacentes. Dado um instante de captura t , tal que $t_i \leq t \leq t_{i+1}$, a posição estimada $\hat{p}(t)$ é obtida por

$$\hat{p}(t) = p_i + \frac{t - t_i}{t_{i+1} - t_i}(p_{i+1} - p_i), \quad (1)$$

onde p_i e p_{i+1} são as coordenadas geográficas nas leituras temporalmente adjacentes à captura, e t_i, t_{i+1} os respectivos instantes em UTC. A Figura 3 ilustra dois cenários de interpolação: (i) com leituras contíguas, a estimativa tende a ser precisa; (ii) com intervalos longos sem leitura, a estimativa pode se afastar da trajetória real, problema agravado em ambientes urbanos por túneis, edificações altas e interferências eletromagnéticas. Nesses casos, o sistema sinaliza leituras ausentes ou de baixa confiabilidade aos demais módulos, podendo interromper temporariamente a captura até que a qualidade do posicionamento seja restabelecida. Cabe destacar que a informação de posicionamento não é utilizada como entrada do detector: ela indexa espacialmente cada imagem, permitindo associar o ativo detectado a uma posição aproximada na rede e viabilizando coletas recorrentes comparáveis



Fig. 3. Dois cenários de interpolação de coordenadas GNSS ao longo de uma trajetória: com dados contíguos (esquerda) e com dados faltantes (direita).

ao longo do tempo. O uso da posição e da velocidade veicular como informação *a priori* para o detector é uma possibilidade para trabalhos futuros.

C. Construção do dataset

A partir das imagens capturadas, selecionou-se um conjunto de 200 imagens para anotação manual e avaliação experimental. O conjunto foi dividido em 140 imagens para treinamento, 40 para validação e 20 para teste. Para os experimentos deste trabalho, foram selecionadas seis classes de interesse para o mapeamento de ativos de distribuição: rede de baixa tensão, rede de média tensão, poste, iluminação pública, rede de telecomunicações e transformador, conforme apresentado na Tabela I.

A seleção das classes considerou sua frequência no conjunto anotado e sua relevância operacional para o inventário de ativos. Embora a classe transformador apresente menor número de instâncias, ela foi mantida por sua importância para a rede de distribuição e por possuir características visuais mais delimitadas em relação a estruturas como cabos e postes. A Figura 4 apresenta exemplos de imagens do dataset com as anotações sobrepostas, enquanto a Figura 5 ilustra exemplos das seis classes consideradas nos experimentos.

IV. METODOLOGIA EXPERIMENTAL

A. Pré-processamento

As imagens originais possuem resolução de 5.568×4.872 pixels. Para o treinamento do modelo, elas foram recortadas em blocos de tamanho 640×640 pixels, resolução compatível com configurações usuais da família YOLO e com os recursos computacionais disponíveis. Cada imagem origina uma grade de 9×8 blocos, totalizando 10.080 blocos de treinamento, 2.880 de validação e 1.440 de teste. A divisão foi realizada no nível da imagem original, de modo que blocos de uma mesma captura não são compartilhados entre conjuntos.

TABELA I

CLASSES ANOTADAS. AS CONTAGENS REFEREM-SE ÀS IMAGENS ORIGINAIS, ANTES DO RECORTE EM BLOCOS.

Classe	Descrição	#Instâncias
Rede BT	Rede baixa tensão (até 1kV)	188
Rede MT	Rede média tensão (13,8kV)	122
Poste	Poste	199
Ilum. Pub.	Iluminação Pública	189
Rede. Tel.	Rede de telecomunicações	182
Trafo	Transformador	41

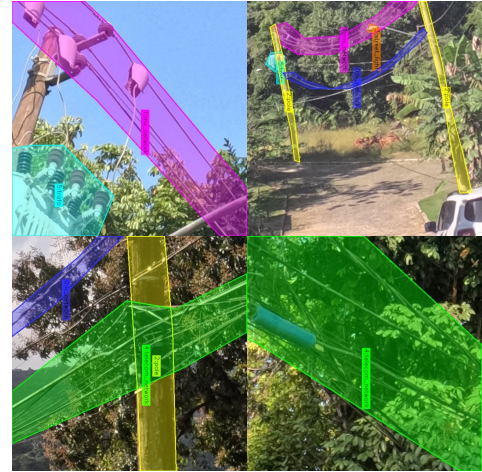


Fig. 4. Exemplos de imagens do conjunto de dados com as anotações manuais sobrepostas.

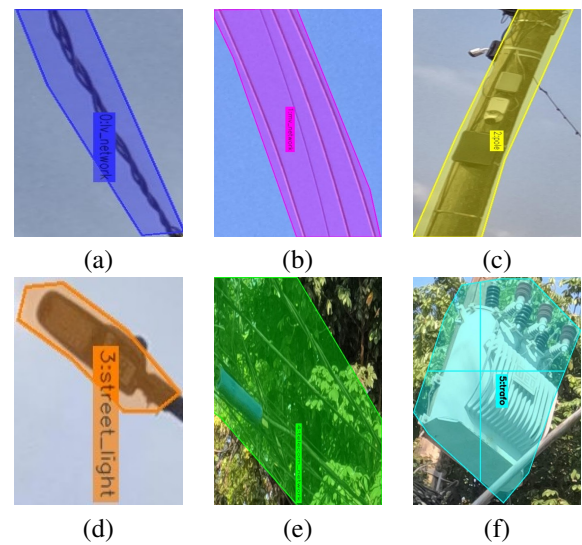


Fig. 5. Exemplo de anotações das 6 classes utilizadas nos experimentos: (a) 'rede de baixa tensão (BT)', (b) 'rede de média tensão (MT)', (c) 'poste', (d) 'iluminação pública', (e) 'rede de telecomunicações', e (f) 'transformador (trafo)'.

B. Modelo e treinamento

Foi utilizado o modelo YOLO26m [13], variante intermediária (20,4 M parâmetros e 68,2 GFLOPs) de uma família recente cuja cabeça de detecção opera de forma *end-to-end*, dispensando a etapa de supressão de não máximos. A variante *m* foi escolhida pelo compromisso entre acurácia e custo computacional na GPU disponível. O treinamento foi realizado por até 200 épocas, com tamanho de entrada de 640 pixels, lotes de 8 amostras e critério de parada antecipada com janela de 35 épocas. Foram empregadas técnicas de aumento de dados, incluindo variação de matiz, saturação e brilho, rotação, translação, alteração de escala, cisalhamento, perspectiva, espelhamento horizontal e vertical, *mosaic*, *mixup*, *copy-paste* e apagamento aleatório. Essas estratégias foram adotadas para ampliar a variabilidade das amostras durante o treinamento e favorecer a adaptação do modelo às variações visuais presentes nas imagens de campo.

O treinamento foi executado em uma GPU NVIDIA GTX 1650. As métricas utilizadas na etapa de validação foram precisão, revocação, F1-score e a precisão média (mAP, do inglês *mean average precision*), que resume a curva precisão–revocação de cada classe, reportada nos limiares mAP@0.5 e mAP@0.5:0.95 [17]. Essas métricas permitem avaliar tanto a confiabilidade das detecções realizadas quanto a capacidade do modelo de recuperar os objetos anotados.

C. Avaliação de desempenho

O desempenho do modelo foi avaliado em duas etapas. Inicialmente, foram analisadas as métricas obtidas no conjunto de validação ao longo do treinamento, com o objetivo de acompanhar a convergência do modelo e selecionar a melhor configuração observada. Em seguida, o modelo treinado foi aplicado ao conjunto de teste, composto por imagens não utilizadas nas etapas de treinamento e validação.

No conjunto de teste, as predições foram comparadas às anotações manuais considerando simultaneamente a correspondência de classe e a sobreposição espacial entre caixas delimitadoras. A sobreposição é quantificada pela Interseção sobre União (IoU), definida como:

$$\text{IoU}(B_{\text{pred}}, B_{\text{gt}}) = \frac{|B_{\text{pred}} \cap B_{\text{gt}}|}{|B_{\text{pred}} \cup B_{\text{gt}}|}, \quad (2)$$

onde B_{pred} e B_{gt} denotam, respectivamente, a caixa delimitadora predita e a anotação de referência. Uma detecção é considerada um verdadeiro positivo (TP) quando apresenta $\text{IoU} \geq 0,5$ em relação à anotação de referência da mesma classe. Caso contrário, as predições e anotações foram contabilizadas como falsos positivos (FP) ou falsos negativos (FN), conforme o caso.

A partir dessas contagens, as métricas de precisão (P), revocação (R) e F1-score ($F1$) são definidas por:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (3)$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (4)$$

$$F1 = 2 \frac{P \cdot R}{P + R}. \quad (5)$$

Para a etapa de inferência no conjunto de teste, foram avaliados limiares de confiança $\tau = \{0,25, 0,35, 0,60\}$. O valor $\tau = 0,35$ foi adotado por apresentar o melhor compromisso observado entre detecções corretas e falsas detecções no conjunto avaliado. Além das métricas quantitativas, foi realizada uma análise qualitativa dos principais fatores que influenciam o desempenho do detector, considerando a escala dos objetos, a sobreposição entre classes, a presença de estruturas alongadas e a complexidade visual do fundo urbano.

V. RESULTADOS E DISCUSSÃO

A. Resultados de validação

Na etapa de validação, o modelo alcançou precisão de 0,70, revocação de 0,34, mAP@0.5 de 0,39 e mAP@0.5:0.95 de 0,23. A diferença entre precisão e revocação indica um detector seletivo sob os limiares adotados: as predições realizadas

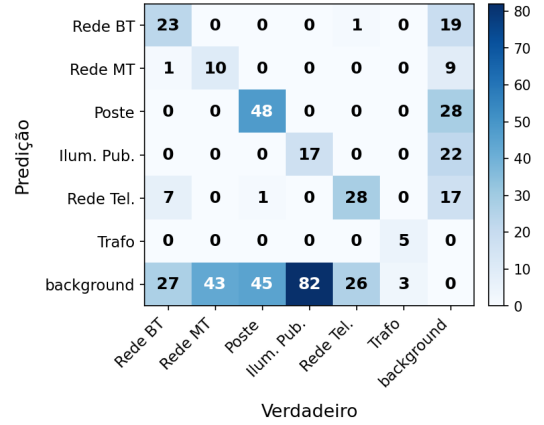


Fig. 6. Matriz de confusão no conjunto de validação, no nível dos blocos de 640×640 pixels, com limiar de confiança $\tau = 0,35$.

são relativamente confiáveis, mas parte dos objetos anotados não é recuperada.

A matriz de confusão, apresentada na Figura 6, indica que parte das instâncias não recuperadas está associada à classe *background*. Esse comportamento é esperado em aplicações com cenas urbanas complexas, nas quais os ativos podem aparecer em pequena escala relativa, parcialmente ocluídos ou com fronteiras visuais pouco definidas. Em particular, classes compostas por estruturas finas ou alongadas, como redes aéreas e postes, tendem a apresentar maior sensibilidade à sobreposição com elementos do ambiente, como vegetação, fachadas, sombras e outras estruturas verticais.

Por outro lado, a classe transformador não registrou confusão com outras classes na validação: todas as suas predições foram corretas e cinco de suas oito instâncias foram recuperadas. Esse resultado, contudo, apoia-se em um número muito pequeno de instâncias e não se sustentou no conjunto de teste (Tabela II), não sendo, portanto, estatisticamente informativo. A análise por classe é aprofundada a seguir.

B. Resultados no conjunto de teste

A Tabela II apresenta os resultados no conjunto de teste em termos de TP, FP, FN, precisão, revocação e F1-score, considerando limiar de confiança $\tau = 0,35$ e correspondência baseada em $\text{IoU} \geq 0,5$, valor selecionado por apresentar o melhor compromisso entre detecções corretas e falsas detecções. Diferentemente da validação, conduzida sobre os blocos e comparada às anotações no mesmo nível, a inferência de teste funde as detecções dos blocos adjacentes e as compara às anotações da imagem original. Trata-se, portanto, de procedimentos de avaliação distintos, o que desaconselha a comparação numérica direta entre a Figura 6 e a Tabela II.

As classes poste e iluminação pública apresentaram os melhores F1-scores individuais (0,17 e 0,10), por possuírem estruturas visualmente mais recorrentes no conjunto, embora ainda sujeitas a variações de escala, perspectiva e oclusão. As classes rede de baixa tensão e rede de telecomunicações registraram verdadeiros positivos, mas concentraram alto número de falsos positivos, indicando maior sensibilidade a elementos visualmente semelhantes no fundo urbano. As classes rede de

TABELA II

DESEMPENHO POR CLASSE NO CONJUNTO DE TESTE (20 IMAGENS ORIGINAIS), OBTIDO COM RECORTE VIRTUAL EM BLOCOS E ASSOCIAÇÃO DAS DETECÇÕES ÀS ANOTAÇÕES DA IMAGEM ORIGINAL, CONSIDERANDO LIMIAZ DE CONFIANÇA $\tau = 0,35$.

Classe	TP	FP	FN	Precisão	Revocação	F1-score
Poste	14	67	67	0,17	0,17	0,17
Ilum. Pub.	3	31	23	0,09	0,12	0,10
Rede BT	3	237	83	0,01	0,03	0,02
Rede Tel.	2	382	59	0,01	0,03	0,01
Rede MT	0	8	56	0,00	0,00	0,00
Trafo	0	10	7	0,00	0,00	0,00

média tensão e transformador não apresentaram verdadeiros positivos sob o critério de IoU adotado.

Esses resultados evidenciam que a detecção em cenas urbanas é fortemente influenciada pela escala dos objetos e pela semelhança visual entre as classes e o fundo. A eles soma-se um efeito de escala introduzido pelo recorte: ativos maiores que um bloco — das redes aéreas, que atravessam a imagem, a transformadores, um dos quais exigiu a fusão de seis blocos — nunca são observados integralmente durante o treinamento, e sua reconstituição no teste depende da fusão das detecções entre blocos adjacentes. O casamento por IoU contra a anotação na imagem original avalia, portanto, o conjunto formado pelo detector e pela etapa de fusão, e não o detector isoladamente. Classes cuja extensão se aproxima da de um bloco, como poste, são as menos afetadas e apresentaram o melhor F1-score do conjunto.

Três limitações condicionam a leitura destes resultados. Primeiro, o conjunto de teste possui apenas 20 imagens, o que limita a estabilidade estatística das métricas por classe. Segundo, o conjunto anotado é fortemente desbalanceado, com o transformador reunindo 41 instâncias. Terceiro, as classes de rede possuem fronteiras visuais ambíguas entre si, o que introduz incerteza na própria anotação de referência.

VI. CONCLUSÃO

Este trabalho apresentou uma avaliação experimental do uso de modelos de detecção de objetos para a identificação de ativos em redes de distribuição elétrica a partir de imagens veiculares georreferenciadas. Para isso, foi utilizado um conjunto de 200 imagens capturadas em campo por um sistema embarcado, com anotações de seis classes de interesse. Os resultados de validação indicaram que o modelo aprende padrões visuais associados às classes avaliadas, enquanto a avaliação no conjunto de teste, conduzida por procedimento distinto, mostrou desempenho substancialmente inferior, concentrado nas classes compostas por estruturas alongadas. O trabalho se posiciona, portanto, como a caracterização de uma linha de base sobre um conjunto de dados inédito, e não como a proposta de um detector otimizado.

Como trabalhos futuros, pretende-se ampliar a diversidade e a representatividade do dataset, revisar e padronizar o protocolo de anotação, avaliar diferentes resoluções de entrada e investigar estratégias de fusão das detecções entre blocos adjacentes, das quais depende a reconstituição de ativos maiores que o bloco. Também serão exploradas etapas sistemáticas

de ajuste de hiperparâmetros, múltiplos limiares de IoU e a comparação com outras variantes da família YOLO, com arquiteturas alternativas de detecção e com bases públicas de infraestrutura elétrica. Por fim, a integração entre detecção visual e informação georreferenciada poderá permitir a associação automática entre ativos detectados e sua localização aproximada na rede de distribuição.

REFERÊNCIAS

- [1] A. A. Faisal, I. Mecheter, Y. Qiblawey, J. H. Fernandez, M. E. H. Chowdhury, and S. Kiranyaz, "Deep learning in automated power line inspection: A review," *Applied Energy*, vol. 385, p. 125507, 2025.
- [2] G. L. Rajora, M. A. Sanz-Bobi, L. B. Tjernberg, and J. E. Urrea Cabus, "A review of asset management using artificial intelligence-based machine learning models: Applications for the electric power and energy system," *IET Generation, Transmission & Distribution*, vol. 18, no. 12, pp. 2155–2170, 2024.
- [3] A. Alquraidei and M. Awad, "Physical asset management for critical utilities: A systematic literature review," *IEEE Access*, vol. 12, pp. 90 644–90 659, 2024.
- [4] G. Lal Rajora, P. C. Báscones, C. M. Domingo, M. A. Sanz-Bobi, R. Palacios Hielscher, M. Martin, D. Vrbčić Tendera, and H. Keko, "Application of machine learning techniques for asset management and proactive analysis in power systems," in *IEEE International Conference on Power Systems Technology*, 2022, pp. 1–8.
- [5] P. Sharma, S. Saurav, and S. Singh, "Object detection in power line infrastructure: A review of the challenges and solutions," *Engineering Applications of Artificial Intelligence*, vol. 130, p. 107781, 2024.
- [6] P. D. Rocha, F. J. P. Lopes, and L. A. d. S. Cruz, "Automating electrical grid asset inspection: From current challenges to future directions," *IEEE Access*, vol. 13, pp. 201 392–201 438, 2025.
- [7] P. Massicotte, L.-A. Leclaire, M. Gaha, G. Houle, and C. Buteau, "Automated inventory of electrical distribution assets based on image recognition and ground LiDAR," in *32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 486–490.
- [8] M. F. da Silva, M. F. dos Santos, J. E. N. Masson, P. M. R. Alves, W. R. Silva, G. M. C. Martins, V. H. G. da Costa, V. B. Schettino, and M. E. Stivanello, "AI-powered automated inspection for optimized asset management in electrical distribution networks," *Journal of Control, Automation and Electrical Systems*, vol. 36, pp. 938–955, 2025.
- [9] B. Khalili and A. W. Smyth, "SOD-YOLOv8: Enhancing YOLOv8 for small object detection in aerial imagery and traffic scenes," *Sensors*, vol. 24, no. 19, 2024.
- [10] W. Shi, X. Lyu, and L. Han, "An object detection model for power lines with occlusions combining CNN and transformer," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–12, 2025.
- [11] E. Bellou, I. Pisica, and K. Banitsas, "Aerial inspection of high-voltage power lines using YOLOv8 real-time object detector," *Energies*, vol. 17, no. 11, 2024.
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, USA, 2016, pp. 779–788.
- [13] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu, "Ultralytics YOLO26: Unified real-time end-to-end vision models," <https://arxiv.org/abs/2606.03748>, 2026.
- [14] H. Zhang, G. Li, D. Wan, Z. Wang, J. Dong, S. Lin, L. Deng, and H. Liu, "DS-YOLO: A dense small object detection algorithm based on inverted bottleneck and multi-scale fusion network," *Biomimetic Intelligence and Robotics*, vol. 4, no. 4, p. 100190, 2024.
- [15] A. A. P. Gomes, F. I. S. Ganacim, F. G. S. Magrin, N. Bobko, L. G. Fernandes, A. Pombeiro, and E. F. R. Romaneli, "A semantically annotated 15-class ground truth dataset for substation equipment to train semantic segmentation models," *Data*, vol. 8, no. 7, 2023.
- [16] S. Panigrahy and S. Karmakar, "Real-time condition monitoring of transmission line insulators using the YOLO object detection model with a UAV," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–9, 2024.
- [17] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics*, vol. 10, no. 3, pp. 1–28, 2021.