



METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS NA DISTRIBUIÇÃO ELÉTRICA USANDO SENTINEL-2, DEEPLABV3+ E VETORIZAÇÃO

Matheus Felipe Gremes¹, Antonio Oseas de C. Filho², Oswaldo Curty da Motta Lima³

¹Doutorando no Programa de Engenharia Química, Departamento de Engenharia Química, Universidade Estadual de Maringá (UEM), Maringá 87020-900, PR, Brasil. matheusfelipegremes@gmail.com

²Co-orientador, professor, pesquisador e doutor no Departamento de Engenharia Elétrica e Ciência da Computação, Universidade Federal do Piauí (UFPI), Teresina 64049-550, PI, Brasil. antoniooseas@ufpi.edu.br

³Orientador, professor, pesquisador e doutor no Programa de Engenharia Química, Departamento de Engenharia Química, Universidade Estadual de Maringá (UEM), Maringá 87020-900, PR, Brasil. ocmlima@uem.br

RESUMO

As perdas não técnicas na distribuição de energia elétrica, especialmente em áreas urbanas densas, elevam custos e impõem desafios operacionais às concessionárias. Este trabalho propõe um fluxo automatizado de monitoramento orbital baseado em imagens Sentinel-2, integrando segmentação semântica via DeepLabv3+, processamento multitemporal, uso de dados da BDGD para identificar áreas fora da cobertura da concessionária e pós-processamento vetorial com dados do OpenStreetMap. Utilizando o conjunto MultiSenGE para treinamento e as máscaras SCL do Sen2Cor para remoção de nuvens, avaliou-se a generalização do modelo treinado na França à região metropolitana do Rio de Janeiro. Os resultados quantitativos destacaram um IoU de 0,9072 na segmentação de áreas urbanas, enquanto estudos de caso demonstraram a eficácia na eliminação de falsos-positivos por nuvens e usos de solo legítimos. A metodologia gera polígonos de alta confiabilidade para indicar prováveis invasões, oferecendo uma ferramenta escalável para priorização de vistorias e mitigação de perdas não técnicas.

PALAVRAS-CHAVE: DeepLabv3+; Perdas não técnicas; Rio de Janeiro; Segmentação semântica; Sensoriamento remoto; Sentinel-2.

1 INTRODUÇÃO

A distribuição de energia elétrica enfrenta desafios persistentes devido às perdas não técnicas (PNT), que englobam furtos, fraudes e imprecisões na medição ou faturamento, comprometendo a eficiência econômica e operacional das distribuidoras (ANEEL, 2008) (CHAVES, TAVARES, *et al.*, 2020). No Brasil, essas perdas atingiram R\$ 8,4 bilhões em 2020, dos quais R\$ 1,8 bilhão referem-se apenas à Light – equivalente a 18,8 % do total de PNT do setor elétrico nacional (ANEEL, 2021). Em áreas urbanas densas e de relevo acidentado, como a região metropolitana do Rio de Janeiro, a identificação de ligações clandestinas torna-se ainda mais complexa, dada a variabilidade sazonal e as interferências atmosféricas.

Avanços no sensoriamento remoto e em técnicas de segmentação de imagens têm sido explorados para mapear automaticamente edificações e áreas urbanas, reduzindo a necessidade de vistorias em campo. Métodos baseados em regras físicas utilizam propriedades espectrais e texturais para identificar construções, mas podem ser sensíveis a ruídos e variações de iluminação (GHANEA, MOALLEM e MOMENI, 2016). Já abordagens de aprendizado profundo, especialmente redes convolucionais como o DeepLabv3+, têm demonstrado alta precisão ao extrair automaticamente características complexas, superando limitações de generalização espacial e sazonal (LI, HUANG, *et al.*, 2022). A integração de informações multitemporais e máscaras de nuvens (SCL do Sen2Cor) (MAIN-KNORN, PFLUG, *et al.*, 2017) aprimora a robustez dessas detecções, enquanto o uso de bases vetoriais como o OpenStreetMap (OSM) (OPENSTREETMAP) e a BDGD (ANEEL, 2025) possibilita filtragem criteriosa de falsos-positivos.



Este estudo propõe um fluxo automatizado que combina segmentação semântica utilizando DeepLabv3+ aplicado a imagens obtidas pela missão Sentinel-2. Adicionalmente, é apresentada uma abordagem baseada em processamento multitemporal e pós-processamento vetorial para subtração de áreas urbanas e de cobertura. Explorando o potencial de generalização de abordagens deep learnings, o modelo foi treinado com o conjunto de dados público MultiSenGE (WENGER; PUISSANT et al., 2022), originalmente desenvolvido para regiões da França. A partir deste treinamento, o modelo foi transferido para a região metropolitana do Rio de Janeiro, permitindo a avaliação de sua capacidade de generalização sem necessidade de retreinamento com dados locais. Uma vantagem em utilizar imagens do Sentinel-2 é a possibilidade de utilização de máscaras para regiões como nuvens, o que possibilita a remoção eficaz destas estruturas. Por fim, as geometrias da BDGD permitem delimitar a faixa de atendimento da rede, objetivando a detecção precisa de prováveis áreas de invasão.

Como contribuição, este artigo propõe um fluxo automatizado que integra segmentação semântica via DeepLabv3+, imagens Sentinel-2, processamento multitemporal e pós-processamento vetorial para a detecção de áreas suspeitas de perdas não técnicas, avalia a capacidade de generalização do modelo MultiSenGE, originalmente treinado na França, à região metropolitana do Rio de Janeiro sem necessidade de retreinamento, disponibiliza um conjunto de dados de alta fidelidade para testes na cidade—baseado em anotações do OpenStreetMap e correções manuais—e introduz um procedimento sequencial de pós-processamento vetorial que reduz significativamente falsos-positivos gerados por nuvens e usos legítimos do solo, oferecendo uma ferramenta potencial para auxiliar concessionárias na priorização de inspeções de campo, otimização de recursos e fortalecimento do combate às perdas não técnicas em contextos urbanos complexos.

2 MATERIAIS E MÉTODOS

Nossa metodologia compreende cinco etapas principais: (i) aquisição de dados a partir de imagens Sentinel-2 disponíveis no repositório Copernicus (COPERNICUS, 2025); (ii) treinamento de uma arquitetura de aprendizagem profunda, especificamente o DeepLabv3+; (iii) construção de uma área de cobertura da concessionária utilizando dados públicos disponibilizados no portal da Base de Dados Geográfica do Distribuidor (BDGD) (ANEEL, 2025); (iv) pós-processamento dos resultados; e, finalmente, (v) avaliação dos resultados obtidos.

Na etapa de aquisição de dados, utilizou-se o conjunto MultiSenGE (WENGER, PUISSANT, et al., 2022), que se destaca na literatura como a única base voltada à segmentação semântica com imagens Sentinel-2 anotadas com elevada precisão e qualidade e que apresenta um volume expressivo de cenas para treinamento. Embora o MultiSenGE abranja amostras da região da França, acredita-se que o modelo seja capaz de extrapolar esses resultados para segmentar adequadamente áreas urbanas em outros contextos geográficos. Além disso, esse conjunto inclui imagens referentes às quatro estações do ano, garantindo que o modelo aprenda a segmentar áreas urbanas de forma consistente, independentemente da variabilidade sazonal. Por fim, não foi identificado nenhum conjunto de dados público com perfil semelhante para regiões do Brasil.

Para a construção do dataset de testes, selecionou-se a cidade do Rio de Janeiro como área de estudo, pois essa região apresenta detalhes ricos que possibilitam a validação abrangente da metodologia por meio de diversos testes. A comparação visual entre as características geográficas do Rio de Janeiro e da área da França contemplada pelo MultiSenGE revela diferenças marcantes. A malha urbana do Rio de Janeiro é



impactada pelo relevo montanhoso acidentado, resultando em layouts urbanos irregulares e níveis de elevação variados. Em contrapartida, a região francesa coberta pelo MultiSenGE exibe em geral terreno mais plano ou suavemente ondulado, o que favorece uma estrutura urbana e malha viária mais homogêneas.

Utilizando uma imagem Sentinel-2 adquirida em 22 de julho de 2023, constatou-se que diferentes áreas, incluindo o município do Rio de Janeiro, estavam contidas em um único tile (unidade espacial específica dos dados capturados pelo satélite Sentinel-2), conforme ilustrado na Figura 1. Na análise preliminar, identificou-se extensa faixa marítima, a qual foi removida da consideração. A área de interesse foi definida de maneira empírica para abranger uma área urbana relevante, conforme mostrado na Figura 1. Ocupando cerca de 4000 km², o ROI foi delimitado pelas coordenadas [629.869, 7.446.400, 709.178, 7.491.689] no sistema de referência espacial CRS EPSG:32723 WGS 84/UTM zone 23S.

As máscaras de cobertura do solo foram elaboradas a partir das anotações disponíveis no OpenStreetMap (OPENSTREETMAP), que serviram como referência (Ground Truth) para o conjunto de imagens de teste. Por meio de tags específicas do OSM — tais como “residential”, “industrial”, “commercial”, “airport”, “road” e “building” —, identificaram-se as regiões correspondentes a cada categoria de uso do solo. Essas etiquetas fornecem uma representação abrangente e detalhada dos diferentes tipos de cobertura, possibilitando a geração precisa das máscaras para análise. Após a criação inicial das anotações de forma automatizada, realizou-se um processo cuidadoso de correção manual, visando aprimorar a qualidade e a exatidão das anotações.

Para a geração do conjunto de dados, foi empregada uma imagem obtida em 22 de julho de 2023, a partir da qual foram extraídos 2232 recortes de 256 × 256 pixels contendo as 12 bandas do Sentinel-2 com resolução espacial de 10 metros. O resultado final constitui um dataset abrangente e preciso que representa as áreas urbanas da região do Rio de Janeiro, e a Figura 1 exemplifica a máscara da imagem obtida a partir das anotações urbanas.



Figura 1: Região do Rio de Janeiro com áreas urbanas anotadas e corrigidas (área urbana em amarelo).

Cada imagem da missão Sentinel-2, tanto para treinamento quanto para testes, apresenta múltiplas características essenciais para a análise: são capturadas em três diferentes resoluções espaciais (60m, 20m e 10m) e abrangem doze bandas espectrais que fornecem dados multiespectrais variando do infravermelho próximo ao ultravioleta, o que enriquece as informações sobre a superfície terrestre para análises detalhadas. Além disso, conta-se com informações sobre áreas afetadas por nuvens (ou outras características



semelhantes, como gelo) obtidas por meio da ferramenta Sen2Cor (MAIN-KNORN, PFLUG, et al., 2017).

Na etapa de treinamento, antes de iniciar o processo de treinamento do modelo proposto neste trabalho, foi realizada uma etapa de reprojeção das imagens, cujo objetivo foi uniformizar todas as bandas para resolução de 10m, garantindo consistência entre elas e eliminando regiões que pudessem comprometer o resultado final. Em seguida, no pré-processamento, foram excluídas as áreas cobertas por nuvens utilizando-se o arquivo de mapa de classificação do cenário (SCL) do Sentinel-2 (produto derivado que classifica cada pixel em uma entre 11 classes, incluindo nuvens). Esse procedimento possibilita a identificação automática e a remoção de pixels afetados por nuvens, assegurando que apenas dados de alta qualidade sejam empregados na segmentação semântica. Para maior precisão, adotou-se também um limiar de confiança capaz de descartar pixels parcialmente encobertos por nuvens ou sombras de nuvens (MAIN-KNORN, PFLUG, et al., 2017).

Utilizou-se um modelo baseado em técnicas de aprendizagem profunda para identificar áreas construídas em imagens de satélite, empregando a arquitetura DeepLabv3+. Essa rede neural foi concebida para tarefas de segmentação semântica, incorporando convoluções dilatadas e o módulo de pirâmide de pooling espacial (ASPP) para capturar informações em múltiplas escalas, o que possibilita a delimitação precisa de objetos em diferentes resoluções. A capacidade da DeepLabv3+ de preservar detalhes finos e contornos torna-a especialmente adequada para segmentação de paisagens urbanas e identificação de elementos geográficos em dados de sensoriamento remoto. Em comparação com outros métodos, essa abordagem demonstra alta precisão na detecção de áreas construídas, evidenciando sua robustez e eficiência em contextos onde a segmentação acurada de objetos é essencial (WANG, DU, et al., 2021).

Durante o treinamento, foram exploradas diferentes configurações de hiperparâmetros, iniciando pelos otimizadores Adamax e SGD, que foram avaliados em função da taxa de aprendizado variando de 1×10^{-2} a 1×10^{-3} . A função de perda adotada foi a Binary Cross-Entropy com Logit Loss, escolhida por sua adequação a problemas de segmentação binária. Parâmetros adicionais, como o tamanho do lote (batch size) e as dimensões das imagens de entrada, foram definidos com base na disponibilidade de dados e nas limitações de hardware; para este estudo, optou-se por utilizar recortes de 256×256 pixels, conforme disponibilizado no conjunto de dados.

Para aprimorar a capacidade de generalização do modelo, aplicaram-se técnicas de aumento de dados exclusivamente ao conjunto de treinamento. Nesse processo, realizaram-se transformações no nível espacial, incluindo espelhamento vertical e horizontal, rotações aleatórias de 90 graus e transposição das imagens, visando diversificar exemplos sem alterar a semântica subjacente. Além disso, todas as operações de pré-processamento consideraram que, embora o Sentinel-2 forneça até doze bandas espectrais, apenas as três bandas RGB foram utilizadas nesta etapa, pois, em testes iniciais, essa combinação apresentou os melhores resultados na segmentação de áreas construídas.

Por fim, investigou-se o desempenho da DeepLabv3+ variando-se os diferentes encoders como backbones: ResNet, EfficientNet e TIMM_EfficientNet, avaliando suas variantes mais utilizadas na literatura. Todos os experimentos empregaram pesos pré-treinados na ImageNet, estratégia adotada para acelerar a convergência e melhorar a extração de características iniciais, possibilitando assim a identificação mais eficiente de áreas edificadas em imagens de satélite.

Na etapa de construção de uma área de cobertura da concessionária utilizando dados públicos, empregou-se a Base de Dados Geográfica da Distribuidora (BDGD), componente essencial do Sistema de Informação Geográfica Regulatória (SIG-R) das



distribuidoras de energia (ANEEL, 2025). A BDGD foi elaborada para simplificar a representação do sistema elétrico real de cada distribuidora, reunindo informações técnicas e comerciais relevantes em entidades padronizadas. Cada entidade é organizada conforme o Dicionário de Dados ANEEL do SIG-R (DDA), o que assegura consistência e interoperabilidade entre diferentes sistemas e fontes. Os dados cartográficos são fornecidos em formato Geodatabase (.gdb), facilitando o acesso e o gerenciamento das informações georreferenciadas. O Manual de Instruções da BDGD e o Módulo 10 do PRODIST oferecem orientações detalhadas sobre a estrutura da BDGD e o uso adequado de suas entidades e campos.

Além de reunir dados técnicos e comerciais, a BDGD disponibiliza informações detalhadas sobre as redes de baixa, média e alta tensão das concessionárias. Esses dados são fundamentais para compreender a infraestrutura elétrica existente em uma determinada região, incluindo a localização e extensão das linhas de transmissão e distribuição. No presente estudo, essas informações foram utilizadas para gerar uma área de cobertura elétrica: aplicou-se um buffer de 70 metros ao redor das linhas de transmissão e distribuição, criando uma zona que representa a faixa potencialmente atendida por essas redes. Essa estratégia permitiu identificar as regiões geográficas mais afastadas da infraestrutura elétrica, as quais podem corresponder a locais com maior probabilidade de ocorrência de invasões ou furtos de energia (perdas não técnicas).

Ao aplicar o buffer de 70 metros às linhas de baixa e média tensão da BDGD, obteve-se uma representação espacial que evidencia claramente a extensão da cobertura elétrica na área de estudo, conforme ilustrado na imagem da esquerda da Figura 2, na qual as faixas de influência são ressaltadas em verde. Esse mapeamento das redes de energia contribui para uma compreensão mais abrangente do cenário elétrico local, fornecendo subsídios valiosos para a análise e a identificação de potenciais áreas de perdas não técnicas.



Figura 2: Área de cobertura elétrica com buffer de 70 m (esquerda) e polígono de cobertura verde sobre máscara urbana amarela (direita).

Após segmentar as áreas urbanas, utilizaram-se os dados da BDGD para gerar um polígono que delimita a área de cobertura da concessionária de energia. Em seguida, aplicou-se a máscara de segmentação urbana sobre esse polígono, subtraindo as áreas urbanas identificadas da extensão de cobertura elétrica. Essa operação evidencia as regiões suscetíveis a invasões, onde podem ocorrer ocupações irregulares e perdas não técnicas, possibilitando a detecção precisa de áreas potencialmente problemáticas para direcionar ações de monitoramento e regularização. Importa ressaltar que o critério para apontar zonas indicativas de invasão baseia-se exclusivamente nessa subtração de polígonos, sem considerar qualquer outro indicador. A imagem da direita da Figura 2 ilustra



esse processo, mostrando o polígono de cobertura em verde sobreposto à máscara de áreas urbanas em amarelo.

Para aprimorar a robustez na detecção de áreas urbanas antes da subtração entre a máscara urbana e a extensão de cobertura da BDGD, aplicou-se um método que utiliza três imagens consecutivas. Inicialmente, realizam-se as detecções de áreas urbanas em cada amostra, contemplando simultaneamente as regiões com nuvens apontadas pelo índice de cobertura de solo (SCL) gerado pelo Sen2Cor. Em seguida, efectua-se a interseção das áreas urbanas identificadas em duas das três amostras, de maneira a descartar detecções isoladas resultantes de nuvens não filtradas ou ruídos pontuais na rede neural. Esse procedimento garante que apenas as regiões detectadas de forma consistente em pelo menos duas imagens sejam consideradas como urbanas, conforme ilustrado na Figura 3, onde se pode visualizar como a sobreposição de datas distintas elimina marcações falsas de nuvens.

Quando uma mesma porção do território está coberta por nuvens em duas das amostras, recorre-se à terceira imagem para manter a detecção de área urbana. Em outras palavras, se uma área aparece como urbana em uma única amostra — mas está encoberta por nuvens em outra ou apresenta ruído de segmentação em outra —, a terceira imagem é utilizada para confirmar ou refutar essa classificação. No exemplo apresentado na Figura 3, a região coberta por nuvens em janeiro de 2019 aparece limpa em dezembro de 2018, e a interseção dessas detecções elimina as falsos-positivos decorrentes de nuvens.

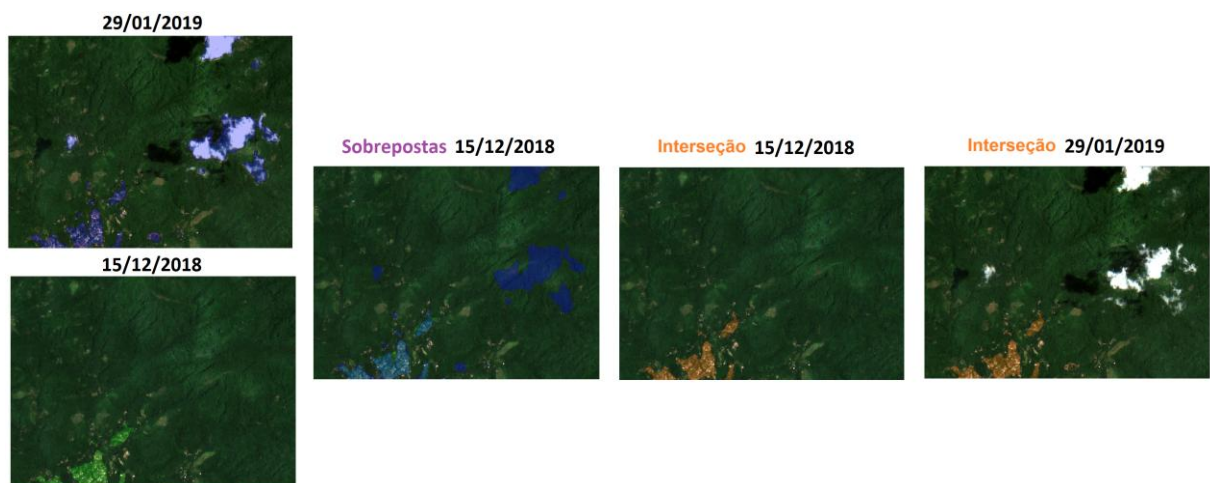


Figura 3: Região com nuvem filtrada a partir da abordagem proposta.

Além disso, o método filtra casos em que a rede neural detecta erroneamente áreas urbanas em posições distintas em duas imagens consecutivas. Ao intersectar as detecções das duas amostras, essas regiões inconsistentes são removidas, permanecendo apenas as áreas que aparecem em locais coincidentes em ambas as imagens. A robustez desse processo pode ser observada na Figura 4, na qual áreas equivocadamente marcadas em imagens consecutivas são filtradas ao cruzar as detecções, mantendo-se apenas as marcações estáveis.



Figura 4: Região com excesso de área urbana filtrada a partir da abordagem proposta.

Após a subtração entre o polígono de cobertura da concessionária e a máscara de áreas urbanas detectadas, aplica-se um refinamento em duas etapas. Primeiramente, excluem-se automaticamente polígonos com área inferior a 400 m² ou que desaparecem após a aplicação de uma erosão morfológica de 5 m, eliminando os chamados “polígonos espaguetificados” que não representam ocupações reais. Em seguida, realiza-se uma filtragem espacial com base em geometrias do OpenStreetMap, removendo todas as regiões que intersectam usos de solo legítimos — como aeroportos, instalações públicas, edificações residenciais e comerciais, hospitais, escolas, áreas industriais, estações de transporte, usinas, subestações, cemitérios, espaços recreativos ou locais religiosos. Essa abordagem integrada assegura que apenas áreas com forte indício de ocupação irregular sejam mantidas, elevando a confiabilidade do indicador de possíveis invasões.

A avaliação de segmentação semântica emprega métricas como acurácia, precisão, recall, F1-score e IoU para comparar arquiteturas de redes e quantificar a capacidade de classificar pixels e delinear regiões. Com esses indicadores, apresentamos um método automatizado para detectar áreas urbanas que ultrapassam a cobertura da concessionária, indicando trechos que devem ser submetidos a análises adicionais.

3 RESULTADOS E DISCUSSÃO

3.1 TREINAMENTO DA REDE

A seguir apresenta-se um resumo dos resultados obtidos. Todos os experimentos listados na Tabela 1 adotaram os mesmos hiperparâmetros definidos na Seção 2, aplicados ao conjunto de teste correspondente à região metropolitana do Rio de Janeiro (Figura 1). Foram avaliados os otimizadores Adamax e SGD, com taxas de aprendizado variando de 1×10^{-2} a 1×10^{-3} , empregando a função de perda Binary Cross-Entropy com Logit Loss, adequada a problemas de segmentação binária. O tamanho de lote foi fixado em recortes de 256×256 pixels — compatível com as capacidades de memória disponíveis — e apenas as três bandas RGB do Sentinel-2 foram utilizadas, pois mostraram melhor desempenho em testes iniciais. Para reforçar a capacidade de generalização, aplicaram-se, exclusivamente ao conjunto de treinamento, aumentos de dados espaciais (espelhamentos vertical e horizontal, rotações de 90° e transposição), preservando a semântica das imagens.

Tabela 1: Resultados do modelo para detecção de área urbana.



Nº	Backbone	Otimizador	LR	Acurácia	F1-Score	IoU	Precisão	Sensibilidade
1	resnet18	Adamax	1e-3	0,9141	0,9292	0,8682	0,9451	0,9141
2	resnet34	Adamax	1e-3	0,9154	0,9254	0,8614	0,9355	0,9154
3	resnet101	Adamax	1e-3	0,7358	0,8439	0,7314	0,9918	0,7358
4	resnet152	Adamax	1e-3	0,6653	0,7961	0,6638	0,9964	0,6653
5	efficientnet-b1	Adamax	1e-3	0,9822	0,9336	0,8758	0,8898	0,9822
6	efficientnet-b3	Adamax	1e-3	0,9451	0,9513	0,9072	0,9577	0,9451
7	efficientnet-b5	Adamax	1e-3	0,8810	0,9204	0,8531	0,9638	0,8810
8	efficientnet-b6	Adamax	1e-3	0,9692	0,9422	0,8908	0,9167	0,9692
9	efficientnet-b7	Adamax	1e-3	0,9030	0,9357	0,8793	0,9708	0,9030
10	t-efficientnet-b1	Adamax	1e-3	0,9717	0,9196	0,8517	0,8731	0,9717
11	t-efficientnet-b1	Adamax	1e-2	0,9838	0,9451	0,8962	0,9095	0,9838
12	t-efficientnet-b3	Adamax	1e-2	0,8886	0,9307	0,8708	0,9774	0,8886
13	t-efficientnet-b3	SGD	1e-2	0,8856	0,9193	0,8510	0,9559	0,8856
14	t-efficientnet-b5	Adamax	1e-2	0,9349	0,9515	0,9077	0,9688	0,9349
15	t-efficientnet-b7	Adamax	1e-2	0,8212	0,8909	0,8038	0,9739	0,8212

Nota: O prefixo t- em t-efficientnet indica que esses modelos foram implementados via timm (PyTorch Image Models); LR se refere a learning rate (taxa de aprendizado).

3.2 DETECÇÃO DE ÁREA URBANA

Observa-se na Tabela 1 que o experimento N° 6 — com backbone EfficientNet-B3, LR (taxa de aprendizado) de 1×10^{-3} e otimizador Adamax — alcançou o melhor compromisso entre as métricas, apresentando acurácia de 0,9451, F1-Score de 0,9513 e IoU de 0,9072. Ao aplicar a rede DeepLabv3+ treinada no conjunto MultiSenGE à região do Rio de Janeiro, obtivemos resultados promissores, com IoU de 0,9072 no conjunto de imagens de teste. É importante destacar que esse modelo foi originalmente ajustado com dados da França (MultiSenGE) e, neste estudo, foi transferido para um contexto geográfico distinto (Rio de Janeiro, Brasil), o que possibilita avaliar sua capacidade de generalização em ambientes variados. Na Figura 5, apresenta-se a predição da rede para a região de interesse (ROI). Como pode ser observado, os valores reportados na Tabela 1 ratificam o resultado qualitativo evidenciado pela metodologia proposta.



Figura 5: Resultados qualitativos da predição da rede (em vermelho) na área delimitada para a região do Rio de Janeiro em 22 de julho de 2023.



3.3 REDUÇÃO DE FALSOS-POSITIVOS

Aplicou-se a abordagem descrita na Metodologia utilizando três imagens de datas próximas, todas submetidas à rede de segmentação treinada. Inicialmente, efetuou-se a detecção de áreas urbanas em cada cena, levando em conta as regiões nubladas identificadas pelo índice SCL do Sen2Cor. Em seguida, intersectaram-se os polígonos resultantes de duas das três imagens para descartar detecções isoladas causadas por nuvens ou ruídos, garantindo que apenas as áreas identificadas de forma consistente em pelo menos dois instantes fossem mantidas. Quando uma mesma porção do território encontrava-se coberta por nuvens em duas cenas, recorreu-se à terceira imagem para confirmar a presença urbana. Esse procedimento assegura a robustez da máscara antes da subtração com o polígono de cobertura da BDGD, reduzindo falsos-positivos decorrentes de nuvens ou falhas de segmentação.

3.4 DETECÇÃO DE ÁREAS DE INVASÃO

Para aprofundar os achados, realizamos estudos de caso usando um buffer de 70 m em torno das linhas de distribuição para identificar possíveis invasões. Verificou-se a presença de ligações clandestinas dentro desse raio, tornando o método eficaz em cenários críticos como o do Rio de Janeiro. Vale notar que a largura do buffer afeta o número de trechos sinalizados — buffers menores geram mais segmentos — e, por isso, deve ser ajustada segundo as características locais.

Aplicando o refinamento descrito na Metodologia (remoção automática de polígonos espaguetificados e filtragem espacial via OSM), houve uma redução significativa nos falsos-positivos, concentrando a detecção em regiões com forte indício de ocupação irregular. Na Figura 6, observam-se dois cenários de possíveis áreas de invasão. À esquerda, as zonas em vermelho estão fora do buffer de 70 m da BDGD, indicando suspeita de ligações clandestinas não monitoradas. À direita, essas áreas exibem padrões típicos de ocupação irregular, como fragmentação e dispersão, enquanto o amarelo representa a faixa do buffer. As visualizações reforçam a presença de ocupações não autorizadas e a necessidade de ações de verificação e regularização.

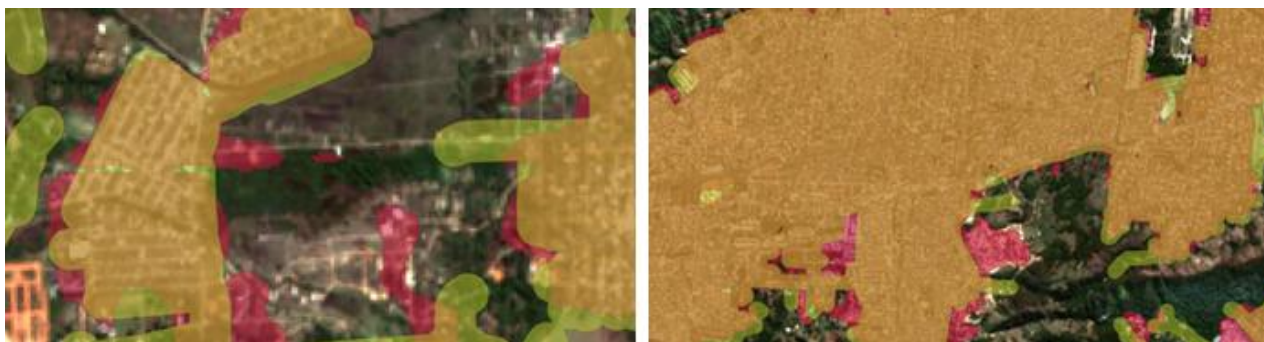


Figura 6: Exemplos de áreas sugestivas de invasão - em vermelho, região provável de ligação clandestina; em amarelo, cobertura da BDGD com buffer de 70 m

Na Figura 7, observa-se o efeito da etapa de refinamento geométrico na máscara de áreas urbanas: os polígonos com área inferior a 400 m² ou sensíveis à erosão morfológica de 5 m foram eliminados automaticamente. As regiões em vermelho correspondem aos “polígonos espaguetificados” descartados por não atenderem aos critérios mínimos de extensão ou robustez topológica, enquanto as áreas em verde representam os



componentes mantidos para análise, garantindo a remoção das áreas serrilhadas resultantes da subtração entre o polígono de cobertura e a máscara de áreas urbanas.

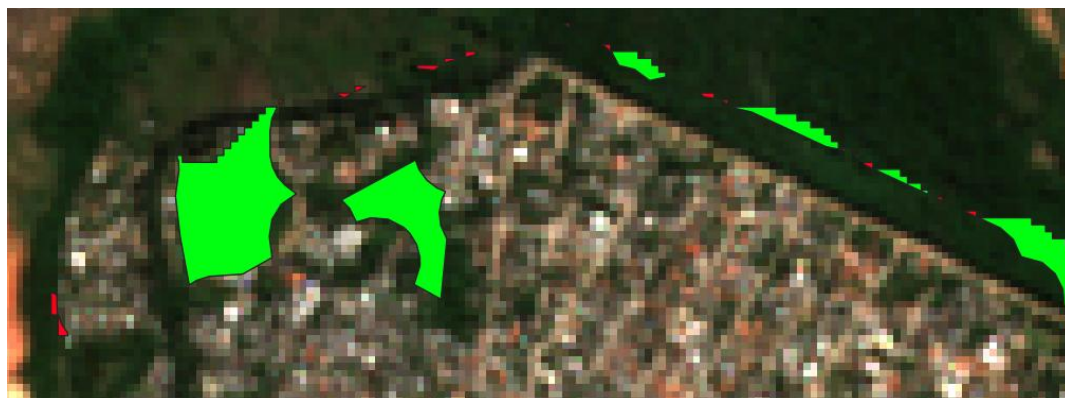


Figura 7: Remoção de polígonos espaguetificados por área (< 400 m²) ou erosão (5 m). (Verde: mantidos; vermelho: removidos.)

Na Figura 8, ilustra-se o resultado da filtragem espacial baseada em geometrias do OpenStreetMap, aplicada sobre o conjunto previamente refinado. As áreas em rosa destacam as feições do OSM associadas a usos de solo legítimos (aeroportos, edificações, instalações públicas, entre outros), que foram cruzadas com a máscara vetorial; as regiões em vermelho, então, indicam as intersecções removidas, ao passo que as porções em verde correspondem às áreas de indício irregular preservadas. Essa etapa assegura que apenas zonas sem sobreposição com infraestruturas permitidas sejam mantidas no indicador de invasão.



Figura 8: Filtragem espacial com OpenStreetMap para remoção de áreas sobre usos de solo legítimos. (Verde: mantidas; vermelho: removidas; rosa: OSM.)

3.5 DISCUSSÕES

Por fim, destaca-se que a combinação de múltiplas etapas — desde a detecção robusta de áreas urbanas em cenas consecutivas, passando pelo refinamento geométrico até a filtragem espacial com OSM — resulta na seleção de regiões com elevada probabilidade de invasão. Esse encadeamento de procedimentos garante que apenas polígonos validados em diferentes datas sejam considerados e que falsos-positivos associados a usos de solo legítimos sejam efetivamente removidos. Assim, obtêm-se áreas



realmente suscetíveis a ocupações irregulares, fornecendo subsídios confiáveis para planejamento de ações de monitoramento e intervenção.

4 CONCLUSÃO

Conclui-se que a metodologia proposta, ao integrar a detecção de áreas urbanas em múltiplas datas, o refinamento geométrico e a filtragem espacial com base em geometrias do OpenStreetMap, apresentou elevada robustez na identificação de ocupações irregulares. O desempenho quantitativo foi destacado por um IoU de 0,9072 na segmentação de áreas urbanas no conjunto de teste evidenciando a eficácia da rede DeepLabv3+ neste novo contexto geográfico.

A estratégia de remoção de nuvens combinou o uso das máscaras SCL geradas pelo Sen2Cor com um procedimento multitemporal que cruza três datas consecutivas: somente as áreas detectadas como urbanas em pelo menos duas imagens são mantidas, enquanto detecções isoladas (seja por nuvens não filtradas ou ruídos de segmentação) são descartadas, e, quando duas amostras estão nubladas, recorre-se à terceira para confirmar ou refutar a classificação. Em seguida, as anotações do OpenStreetMap servem de referência para excluir automaticamente qualquer polígono que se sobreponha a usos de solo legítimos (aeroportos, vias, edificações, áreas industriais etc.), assegurando que os polígonos finais representem com alta confiabilidade apenas prováveis invasões.

Adicionalmente, o método demonstrou capacidade de generalização: treinado com dados do MultiSenGE na França, foi aplicado com sucesso à região metropolitana do Rio de Janeiro sem necessidade de re-treinamento. Essa característica ressalta o potencial de escalabilidade da abordagem para diferentes cenários urbanos.

A construção da área de cobertura elétrica a partir dos dados da BDGD, por meio de buffer de 70m, foi crucial para delimitar as regiões direta e indiretamente atendidas pela rede. As áreas mapeadas fora dessa faixa, validadas pelos estudos de caso, confirmaram a capacidade do método de identificar trechos com alta probabilidade de invasão.

Por fim, o fluxo completo de processamento — da segmentação semântica ao pós-processamento vetorial — fornece uma ferramenta confiável para apoiar ações de monitoramento e planejamento urbano em contextos de distribuição elétrica, contribuindo para a mitigação de perdas não técnicas e para a gestão de riscos geográficos.

Como trabalhos futuros, sugerimos explorar a integração de dados de consumo energético em nível de bairro para correlacionar detecções de invasão com alterações no padrão de demanda, além de testar outras arquiteturas de redes neurais para segmentação semântica e comparar seu desempenho e robustez em distintos cenários urbanos; também sugerimos o uso de imagens de maior resolução e a mudança do foco de segmentação de áreas urbanas para telhados, permitindo a identificação mais precisa de pontos de conexão clandestina.

AGRADECIMENTOS

Os autores agradecem à Agência Nacional de Energia Elétrica (ANEEL) e ao seu Programa de Pesquisa e Desenvolvimento pelo apoio financeiro por meio do PD 00382 0160/2023. Além disso, os autores reconhecem o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).



REFERÊNCIAS

ANEEL. **Nota técnica nº 342 Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica segundo ciclo de revisão tarifária periódica das concessionárias de distribuição.** Superintendência de Regulação Econômica (SER). Brasília. 2008.

ANEEL. **Perdas de energia elétrica na distribuição.** Superintendência de Gestão Tarifária (SGT). [S.l.]. 2021. Edição 1/2021; Período dos dados: janeiro/2008 a dezembro/2020; dados obtidos por meio do Sistema de Acompanhamento de Informações de Mercado – SAMP Balanço (julho 2021).

ANEEL. Base de Dados Geográfica da Distribuidora – BDGD. **Dados Abertos ANEEL,** 2025. Disponível em: <<https://dadosabertos.aneel.gov.br/>>. Acesso em: 21 Julho 2025. SRD/ANEEL; Mantenedor: EGDI/ANEEL; Última Atualização: 26 de abril de 2023; Criado: 20 de abril de 2023; Frequência de Atualização: Anual.

CHAVES, A. C. et al. **As Perdas Não Técnicas no Setor de Distribuição Brasileiro: Uma Abordagem Regulatória.** [S.l.]: Grupo de Estudos do Setor Elétrico (GESEL), 2020.

COPERNICUS. **Copernicus Sentinel-2,** 2025. Disponível em: <<https://www.copernicus.eu/>>. Acesso em: 21 Julho 2025.

GHANEA, M.; MOALLEM, P.; MOMENI, M. Building extraction from high-resolution satellite images in urban areas: Recent methods and strategies against significant challenges. **International Journal of Remote Sensing,** v. 37, p. 5234–5248, 2016. ISSN 21.

LI, J. et al. A review of building detection from very high resolution optical remote sensing images. **GIScience & Remote Sensing,** v. 59, n. 1, p. 1199–1225, 2022.

MAIN-KNORN, M. et al. **Sen2Cor for sentinel-2.** Image and Signal Processing for Remote Sensing XXIII. [S.l.]: SPIE. 2017. p. 37–48.

OPENSTREETMAP. About OpenStreetMap. Disponível em: <https://wiki.openstreetmap.org/wiki/About_OpenStreetMap>. Acesso em: 21 Junho 2025.

WANG, C. et al. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. **Computers and Electronics in Agriculture,** v. 189, p. 106373, 2021.

WENGER, R. et al. Multisenge: A multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,** 3, 2022. 635–640.