

Identificação de mudanças em áreas de mineração a céu aberto por meio de imagens de satélite e aprendizado profundo

Hércules C. S. Pereira
LIAREA e PGCAP
INPE

São José dos Campos, Brazil
hercules.pereira@inpe.br

Valdivino A. Santiago Júnior
LIAREA e PGCAP
INPE

São José dos Campos, Brazil
valdivino.santiago@inpe.br

Elcio H. Shiguemori
LIAREA e PGCAP
IEAv

São José dos Campos, Brazil
elcio@ieav.cta.br

Dener C. Mendes
LIAREA e PGCAP
INPE

São José dos Campos, Brazil
dener.mendes@inpe.br

Caio E. Dias
LIAREA e PGCAP
INPE

São José dos Campos, Brazil
caio.dias@inpe.br

Matheus C. Domingos
LIAREA e PGCAP
INPE

São José dos Campos, Brazil
matheus.domingos@inpe.br

Luísa M. C. Santos
LIAREA e PGCAP
INPE

São José dos Campos, Brazil
luisa.santos@inpe.br

Resumo—A mineração a céu aberto pode gerar impactos ambientais significativos, contaminando solo, ar e corpos hídricos. Embora existam estudos com inteligência artificial (IA) para detectar essas áreas, ainda há carência de abordagens com ampla escala temporal e menor dependência de imagens de alta resolução. Este trabalho visa contribuir para identificar áreas de mineração a céu aberto no Brasil via imagens de satélite, aprendizado profundo e detecção de mudanças. Para isso, foi construída uma base com 2.000 pares de imagens Sentinel-2 e máscaras binárias considerando os estados Minas Gerais, Pará, Amazônia e Goiás, e uma escala de tempo de sete anos. Avaliaram-se quatro modelos: U-Net padrão, redes Siamesas com *encoder* ResNet-18 e ResNet-34, e uma U-Net com *encoder* ResNet-34 e mecanismo de atenção. A U-Net padrão teve melhor desempenho geral, com 92,95% de precisão, 92,43% de *Recall*, F1-Score de 92,66%, *Intersection over Union* (IoU) de 66,94% e latência de 55,88 ms. O Grad-CAM revelou que o modelo foca em áreas com mudanças relevantes, destacando seu potencial para aplicações futuras.

Palavras-chave—Mineração a Céu Aberto, Aprendizado Profundo, Detecção de Mudanças, Segmentação Semântica, Inteligência Artificial Explicável.

I. INTRODUÇÃO

Uma das formas de extração mineral do Brasil é a mineração a céu aberto, dada a viabilidade econômica, embora gere impactos ambientais, causados pelo desmatamento e remoção de solo fértil [1]. Mineração a céu aberto consiste na remoção da cobertura superficial do solo e rochas, e tem o intuito de expor o mineral presente na superfície [2]. Em contraponto a isso, ocorre o crescimento da exploração de minérios ilegais, que resulta em impactos ambientais significativos, tais como

desmatamento, degradação do ecossistema e a contaminação de corpos hídricos resultante da contaminação por mercúrio [3]. Esses impactos são frequentemente associados a práticas de extração informais ou não regulamentadas, cuja atuação difere consideravelmente da mineração industrial formal. Por essa razão, é fundamental distinguir entre os diferentes tipos de atividade minerária, especialmente no que diz respeito à escala, estrutura e legalidade das operações [4].

Nesse sentido, destaca-se a importância de compreender as diferenças entre mineração e garimpo. Enquanto a mineração ocorre em maior escala, com uma estrutura organizacional definida, uso de maquinário pesado e mão de obra especializada, o garimpo caracteriza-se por ser uma atividade de menor escala, conduzida de forma mais manual, com equipamentos rudimentares e menor volume de extração mineral [5].

Diante desse cenário, destacam-se os trabalhos de [6], que usa redes 3DCNNs para detecção temporal de mineração, e [7], que aplica U-Net++ para segmentação e detecção de mudanças em imagens Sentinel. A despeito de tais abordagens serem interessantes, existe carência de estudos que considerem uma grande escala de tempo, para melhor identificar possíveis ilegalidades, assim como também há dependência de imagens de alta resolução e dificuldade de generalização por regiões. Este trabalho visa contribuir para a identificação de áreas de mineração a céu aberto no Brasil via imagens de satélite, aprendizado profundo (AP) e detecção de mudanças. Para isso, uma abordagem é proposta, onde uma base de dados foi elaborada, contendo 2.000 pares de imagens e máscaras binárias, considerando os estados de Minas Gerais, Pará, Amazônia e Goiás e uma escala de tempo de sete anos. Além das bandas, azul (B), verde (G) e infravermelho próximo (NIR) de imagens do satélite Sentinel-2 terem sido usadas, o índice *Soil Adjusted Vegetation Index* (SAVI) [8] também foi

LIAREA - Laboratório de Inteligência Artificial para Aplicações AeroEspaciais e Ambientais; PGCAP - Programa de Pós-Graduação em Computação Aplicada; INPE - Instituto Nacional de Pesquisas Espaciais; IEAv - Instituto de Estudos Avançados.

calculado para apoiar o objetivo da pesquisa.

Foram selecionados quatro modelos de AP para a abordagem: U-Net modificada para detecção de mudança [9] e duas redes Siamesas com *encoder* ResNet-18 e ResNet-34, baseadas em [10], e uma U-Net com *encoder* ResNet-34 [11] junto a um mecanismo de atenção. A U-Net foi escolhida por ser amplamente utilizada em aplicações de sensoriamento remoto, devido à sua capacidade de segmentação precisa e preservação de detalhes espaciais. Já a rede siamesa com *encoder* ResNet-18 e ResNet-34 foi escolhida após análise de trabalhos relacionados, nos quais essa combinação com *decoder* U-Net apresentou bons resultados na detecção de mudanças. Por fim, adotou-se a U-Net com *encoder* ResNet-34 e mecanismo de atenção, visando melhorar a delimitação feita pelo modelo. Além disso, a aplicação do Grad-CAM [12], uma técnica de inteligência artificial explicável (XAI) possibilitou explicar, em parte, as decisões da U-Net, evidenciando sua atenção em áreas com mudanças relevantes.

As principais contribuições desse trabalho são: (i) a criação de uma base de dados relativa à mineração, considerando diversos estados do país e com uma grande escala de tempo, para apoiar a determinação de ações ilegais; (ii) o desenvolvimento de um método apoiado por AP e detecção de mudanças para abordar o problema em questão; e (iii) o uso de XAI para realizar explicabilidade dos resultados gerados por modelo de AP.

II. TRABALHOS RELACIONADOS

Esta seção apresenta trabalhos relevantes relacionados a essa pesquisa. Em [13], os autores abordam uma gama de métodos utilizados para esse problema, os quais variam desde métodos manuais, como a fotointerpretação, até métodos automáticos de interpretação de imagens. No entanto, o trabalho apresenta limitações, como o foco restrito a esses métodos sem aprofundar a viabilidade técnica, a análise ambiental limitada a poucos aspectos e a ausência de consideração dos impactos sociais e de avaliações mais completas.

Um estudo recente [6] investiga o uso de redes neurais convolucionais 3D (3DCNNs) para classificação de cena aplicadas a imagens de satélite com foco em áreas de mineração. Utilizando sequência temporal de imagens, obtidas pela constelação de satélites *Planet*, combinado a dados do MapaBiomass. Os resultados do estudo indicam que o modelo 3DCNNs é adequado para captar informações temporais, sendo superior se comparados a abordagens baseadas unicamente em modelos 2D com CNNs. Contudo, o trabalho apresenta limitações, como a necessidade de alta capacidade computacional para treinamento e inferência do modelo, além da dependência de dados temporais sequenciais, que podem não estar disponíveis ou completos em todas as regiões, limitando a aplicabilidade do método.

Na abordagem feita em [14] propõe-se uma metodologia para identificar ações ilegais para mineração subterrânea, por meio da detecção de deformações no solo e mudanças visuais na superfície. Os autores combinaram a tecnologia PS-InSAR junto a imagens ópticas de alta resolução para identificação

de extração de minério ilegalmente. Seguindo as abordagens, esta pesquisa propõe uma abordagem próxima para o monitoramento de áreas de mineração a céu aberto, utilizando a detecção de mudanças. Ainda assim, a metodologia apresenta limitações de dados de alta resolução, à cobertura temporal restrita e à validação em área limitada, o que dificulta sua generalização para outros contextos.

No trabalho de Guo et al. [15], é descrito um *benchmark* global voltado à detecção de mudanças resultantes de atividades de mineração, utilizando dois instantes de tempo e alta resolução espectral. O conjunto de dados, composto por 70 mil pares de patches com anotações ao nível de pixel e resolução de 1,2 m, abrange 100 locais de mineração distribuídos em diferentes países. Os autores apresentam a arquitetura ChangeFFT, que utiliza a transformada rápida de *Fourier* com o objetivo de aprimorar a representação de mudanças presentes nas imagens. Além das contribuições relevantes, o trabalho apresenta algumas limitações. A principal é a dependência de imagens de alta resolução. A aplicação do método baseado em *Fourier* pode não conseguir caracterizar mudanças mais sutis em áreas com baixa variação espectral ou com ruídos atmosféricos.

Um estudo sobre detecção de mudança em área de mineração é apresentado no artigo [7], que propõe a metodologia MineCam, que utiliza aprendizado profundo e sensoriamento remoto para segmentação e detecção de mudanças em áreas de mineração. O *dataset* é composto por 2.000 imagens de satélite, cobrindo mais de 400 áreas de mineração, classificadas em dez classes como: escavação, pilhas de estéril e armazenamento de rejeito. Foram empregados dados ópticos (Sentinel-2) e Radar Sentinel-1, ambos com resolução espacial de 10 metros. O modelo utilizado foi uma arquitetura U-Net++, que apresentou resultados de 0,65 a 0,75 — dependendo da classe avaliada — na métrica de performance (provavelmente F1-score ou IoU). Apesar dos resultados satisfatórios, o estudo não realizou uma análise detalhada sobre como as variações de luz e as condições climáticas adversas podem afetar o desempenho do modelo. Além disso, embora os autores tenham sugerido o uso da fusão entre imagens ópticas e SAR para minimizar esses impactos, essa estratégia não foi explorada de forma aprofundada, o que limita a avaliação da eficácia do método nessas situações.

Uma visualização em um contexto de aplicações de inteligência artificial voltado para a indústria de mineração também é relevante, sendo assim o trabalho [16], onde os modelos são aplicados em processos como otimização, manutenção preventiva, monitoramento e segurança. Apresenta uma aplicação em um contexto ético associado à utilização de IA, sobre transparências, responsabilidade e impacto ambiental, com objetivo de aumentar a eficiência em operações e reduzir o impacto socioambiental. Embora o trabalho proponha e apresente a aplicação de IA na indústria de mineração, ele possui duas limitações principais. Primeiramente, o foco é exclusivamente teórico, pois se trata de uma revisão que não inclui experimentos práticos nem validação empírica. Em segundo lugar, apesar de fornecer uma visão geral abrangente

e recomendações valiosas, essas podem não ser diretamente aplicáveis a contextos específicos sem a necessidade de adaptações adicionais.

III. MATERIAIS E MÉTODOS

Neste estudo, utilizamos redes neurais profundas para detectar mudanças em áreas de mineração a céu aberto. A seguir, são detalhados os materiais e métodos.

A. Conjunto de dados

A base de dados CHANGE-MINING2025, desenvolvida nesse trabalho, é composta por dois tempos de imagens e máscaras binárias. A Fig 1 ilustra o processo feito para elaboração da base de dados. As etapas desse processo são: 1: A aquisição das imagens por meio do satélite Sentinel-2, dispostas na plataforma Copernicus Browser [17]; 2: A seleção dos instantes de tempo T1 e T2 (2017 e 2024) e o cálculo do índice espectral SAVI; 3: Cálculo da mudança por meio da diferença entre os dois índices SAVI ($T2 - T1$) e refinamento criando máscara de forma manual; 4: Geração da máscara de mudança; 5: Recorte no software QGIS da dimensão de 128×128 pixels; 6: O conjunto de dados é gerado e disposto em três pastas: "T1", "T2", "Mudança". A máscara de mudança é composta por valores binários, em que o valor "0" representa regiões sem alteração e o valor "1" indica áreas onde houve mudança. O dataset encontra-se publicamente disponível CHANGE-MINING2025 no GitHub [18].

O SAVI foi escolhido pela sua sensibilidade em destacar solo exposto, sendo uma das características presentes em áreas de mineração a céu aberto. Além do SAVI, as bandas verde (G), azul (B) e infravermelho próximo (NIR) foram também usadas.

O *Copernicus Open Access Hub*, também conhecido como Copernicus Browser [17], é uma plataforma da Agência Espacial Europeia (ESA) que fornece acesso gratuito a dados dos satélites Sentinel, permitindo visualizar e baixar imagens de satélites tais como: Sentinel-1, Sentinel-2 e Sentinel-3. Tais dados podem ser usados em aplicações de monitoramento ambiental e uso e cobertura do solo. O QGIS é um software de código aberto que permite analisar e editar dados georreferenciados de diferentes formatos e uso. O QGIS permite a criação de máscaras, raster, vetores, pontos, linhas e polígonos [19].

B. Detecção de mudanças

A detecção de mudanças é uma tarefa utilizada em sensoriamento remoto, que tem o objetivo de identificar e analisar mudanças presentes na superfície terrestre, levando em consideração diferentes períodos. O estudo dessa mudança pode ser utilizado em aplicações tais como desmatamento, urbanização, expansão agrícola e exploração de minério. A detecção de mudanças consiste na comparação de imagens obtidas por sensores de satélite, com características que sejam compatíveis, tendo objetivo de identificar alterações no uso de

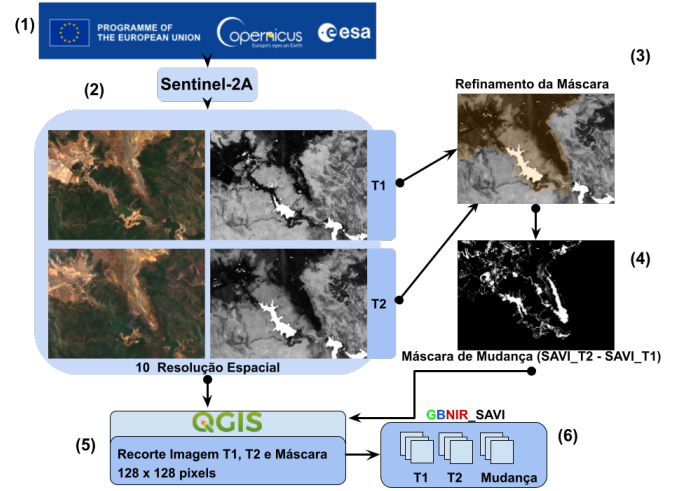


Figura 1. Fluxograma do processo de geração do conjunto de dados CHANGE-MINING2025.

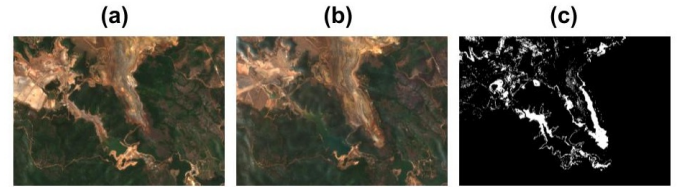


Figura 2. a) primeiro instante de tempo T1, b) segundo instante de tempo e c) máscara da mudança.

cobertura do solo ao longo do tempo. A mudança é o resultado da diferença entre dois instantes de tempo [20].

A Fig 2 apresenta um exemplo de imagens de cenas de mineração, em que T1 representa o ano de 2017 e T2 representa o ano de 2024. A máscara de mudança corresponde a uma matriz binária que indica as regiões onde há alterações entre os instantes de tempo T1 e T2.

C. Modelos de AP selecionados

1) *U-Net*: O modelo de rede neural convolucional U-Net, realiza segmentação de imagens, com sua arquitetura em formato de U. Foi desenvolvida por [9], para o desafio de segmentação semântica aplicado a imagens microscópicas para elétrons de tecido neurológico, foi proposto para o desafio do Simpósio Internacional de Imagens Biomédicas (ISBI) em 2012. A Fig 3, ilustra a arquitetura do modelo U-Net, apresentado no artigo [9] a arquitetura é composta por duas partes, *encoder* e *decoder*. O *encoder* é um modelo de rede neural tradicional.

Sendo composta por convoluções 3×3 sem preenchimento de borda, seguido de *max pooling* 2×2 passo 2 realizando *downsampling*. A cada *downsampling* o número de canais de características dobra. A cada etapa do *decoder* é aplicado um aumento de resolução *upsampling* sobre o mapa de características, seguido por uma convolução 2×2 *up-convolution*, realizando uma redução pela metade do número de canais de

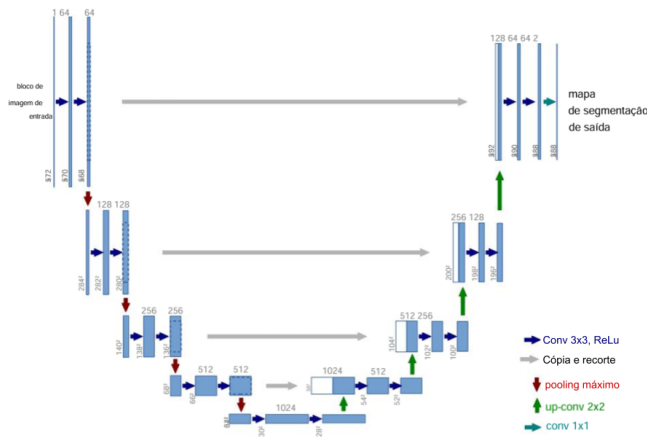


Figura 3. Arquitetura do modelo U-Net.

característica. Em seguida, é feita uma concatenação com o mapa de característica, e depois duas convoluções 3×3 , seguidas de uma ReLU. Por fim, se aplica uma convolução 1×1 a fim de mapear o vetor de características.

2) *Siamesa (U-Net Encoder Resnet-18)*: A arquitetura do modelo Siamesa utilizada neste trabalho é baseada em [10]. A arquitetura emprega duas redes convolucionais idênticas aplicadas em paralelo às imagens de entrada T1 e T2, com compartilhamento de pesos, a fim de garantir uma extração simétrica e consistente de características para ambos os instantes temporais. O *encoder* adotado é uma versão modificada da ResNet-18, pré-treinada na base de dados ImageNet. A principal modificação consiste na adaptação da primeira camada convolucional para aceitar quatro bandas espectrais, o que permite a utilização de imagens multiespectrais do satélite Sentinel-2, além do tradicional uso de imagens RGB.

As imagens são redimensionadas para 256×256 pixels, e o *encoder* é truncado na camada *Layer4*, o que reduz a profundidade da rede e gera um feature map com 512 canais e resolução espacial de 8×8 . Essa modificação visa equilibrar o poder de extração de características com a complexidade computacional. Após a extração, é realizada uma operação de subtração entre os vetores de características das imagens T2 e T1, com o objetivo de destacar as mudanças ocorridas entre os dois períodos.

O *decoder*, inspirado na arquitetura U-Net, aplica sucessivas camadas de upsampling e convoluções, incorporando conexões de skip para preservar detalhes espaciais e melhorar a reconstrução do mapa de mudanças. O resultado final é um mapa de segmentação binária que identifica as regiões de alteração entre as imagens temporais.

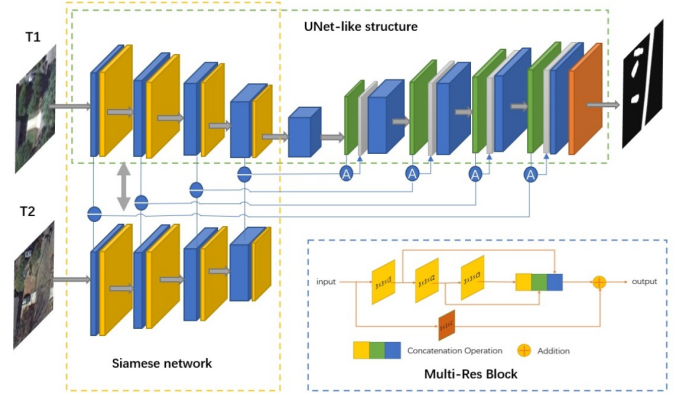


Figura 4. Arquitetura de referência para os modelos de rede siamesa

A Fig 4 ilustra uma arquitetura análoga do modelo. No exemplo, são apresentadas duas imagens como entradas redimensionadas para 256×256 pixels, processadas por uma ResNet-18, e o decodificador se baseia no *decoder* de uma U-Net tradicional, que gera a segmentação da imagem.

3) *Siamesa (U-Net Encoder Resnet-34)*: O modelo baseia-se na arquitetura U-Net, combinada com a ResNet-34 com *encoder*. A camada de entrada foi modificada para aceitar 8 canais espectrais, correspondentes à concatenação de dois patches multitemporais, ou seja, 4 bandas por imagem. O *encoder* utiliza os blocos convolucionais da ResNet-34 até a camada *layer4*, preservando a estrutura residual da rede original. Na parte da decodificação, são aplicadas operações de upsampling seguidas de concatenação com os mapas correspondentes do *encoder skip connection*, possibilitando a recuperação de detalhes espaciais. Cada etapa de *decoder* é composta por blocos convolucionais com duas camadas *Conv-BN-ReLU*. Um bloco adicional de convoluções é aplicado após o último nível do *encoder bottleneck*, expandindo a profundidade para 1024 canais. A saída final é uma convolução 1×1 com 2 canais, representando as classes de mudança e não mudança, com a resolução ajustada por interpolação bilinear para coincidir com a entrada.

4) *Siamesa (U-Net Encoder ResNet-34 + Attention)*: Nesta variante, foi adotada uma arquitetura U-Net com ResNet-34 como codificador *encoder* e a adição de módulos de atenção espacial no decodificar. A entrada do modelo consiste na concatenação de duas imagens multiespectrais (t1 e t2), totalizando 8 bandas espectrais, processadas por uma camada convolucional adaptada para acomodar essa dimensão. O *encoder* segue a estrutura da ResNet-34 até a *layer4*, extraindo representações hierárquicas profundas.

Na fase de decodificação, são aplicadas operações de upsampling seguidas por *Attention Gates (AGs)*, que atuam sobre os mapas de ativação do *encoder* antes de serem combinados nas conexões de atalho *skip connections*. Cada AG recebe como entrada o mapa de ativação do *encoder* e um sinal de gating do *decoder*, aprendendo a suprimir regiões irrelevantes e realçar regiões discriminativas para a tarefa de detecção de mudanças. Essa filtragem adaptativa é feita por meio de

operações convolucionais seguidas de uma função sigmoide que gera um mapa de atenção aplicado ao mapa do *encoder*.

Cada etapa do *decoder* é composta por um bloco convolucional do tipo Conv-BN-ReLU, e a saída final é obtida por uma convolução 1×1 com 2 canais (mudança/não mudança), interpolada para coincidir com a resolução original da entrada. A introdução dos módulos de atenção visa melhorar a delimitação espacial das regiões alteradas, reduzindo ruído e melhorando a consistência da segmentação.

D. Inteligência artificial explicável (XAI)

A inteligência artificial explicável (XAI) tem como objetivo entender os resultados obtidos pelos modelos de aprendizado de máquina. Ao compreender como o modelo chega a determinada decisão, é possível agregar confiabilidade aos resultados gerados. Dado que compreendemos como o modelo chega ao resultado, é possível confiar mais em suas previsões e identificar eventuais falhas ou vieses no processo decisório. Neste estudo, utiliza-se a técnica *Gradient-weighted Class Activation Mapping* (Grad-CAM) proposta por [12]. Trata-se de uma abordagem de interpretabilidade/explicabilidade que permite visualizar quais regiões da imagem mais influenciam as decisões de um modelo de rede neural convolucional. A partir dos gradientes das ativações de uma determinada camada convolucional, o método gera mapas de ativação que destacam as áreas relevantes para a predição do modelo. Como o modelo processa dois instantes de tempo combinados (T1 e T2), um software foi implementado para desagregar e examinar cada momento separadamente. Considerando o modelo U-Net que, conforme será visto adiante (Seção IV) foi o que teve o melhor desempenho, foi selecionada a última camada de convolução do quarto bloco do codificador (enc4[-3]), pois ela proporciona um equilíbrio satisfatório entre a abstração semântica e a preservação da resolução espacial. O Grad-CAM foi calculado a partir do pixel central da classe “mudança” na saída, o que permitiu visualizar as áreas das imagens que mais impactaram a previsão em cada período. A técnica foi aplicada após o treinamento do modelo para gerar mapas de ativação sobre as imagens do conjunto de teste.

E. Métricas

Em problemas de classificação com desequilíbrio entre classes, é comum utilizar o parâmetro *average=weighted* para calcular métricas globais. Essa abordagem pondera o resultado de cada classe de acordo com sua frequência no conjunto de dados, garantindo que classes mais representativas tenham maior influência nas métricas finais.

1) *Precision (Weighted)*: A métrica *Precision* [21] avalia a proporção de acertos entre todas as previsões positivas feitas pelo modelo. Em outras palavras, mede quantas das áreas identificadas como “mudança” realmente sofreram alterações.

$$\text{Precision}_{\text{weighted}} = \frac{\sum_{i=1}^C w_i \cdot \text{precision}_i}{\sum_{i=1}^C w_i}$$

2) *Recall (Weighted)*: O *Recall* [21] representa a capacidade do modelo em encontrar todas as ocorrências reais da classe de interesse. No contexto da detecção de mudanças, indica quantas das alterações reais foram corretamente detectadas.

$$\text{Recall}_{\text{weighted}} = \frac{\sum_{i=1}^C w_i \cdot \text{recall}_i}{\sum_{i=1}^C w_i}$$

3) *F1-Score (Weighted)*: O *F1-Score* [21] é a média harmônica entre *Precision* e *Recall*, sendo útil especialmente em cenários desbalanceados. Ele fornece uma medida única que leva em conta tanto os falsos positivos quanto os falsos negativos.

$$F1_{\text{weighted}} = \frac{\sum_{i=1}^C w_i \cdot F1_i}{\sum_{i=1}^C w_i}$$

Onde: C é o número total de classes, w_i representa a quantidade de amostras reais da classe i , e precision_i , recall_i e $F1_i$ são os valores individuais obtidos para cada classe.

4) *Intersection over Union*: *Intersection over Union* (IoU) [22] mede a sobreposição entre área prevista (*bounding box*) e área real.

$$\text{IoU} = \frac{\text{Área da Interseção}}{\text{Área da União}} \quad (1)$$

5) *Latência*: É calculado pela média do tempo de inferência sobre todas as imagens presentes no conjunto de teste [23].

$$\text{Latência} = \frac{1}{N} \sum_{i=1}^N t_i$$

IV. EXPERIMENTAÇÃO E RESULTADOS

A seleção das áreas de estudo considerou o contexto histórico de regiões brasileiras com intensa atividade mineradora, como os estados de Minas Gerais, Pará, Amazonas e Goiás. Foi adotada uma escala temporal de sete anos, visando capturar as diferentes fases do processo minerário, desde a abertura das frentes de lavra até possíveis estágios de estabilização ou abandono. Já a avaliação dos resultados da detecção de mudanças utilizou imagens selecionadas com base em áreas que apresentavam tanto alterações abruptas quanto sutis, incluindo regiões com cavas de mineração, barragens de rejeito e pilhas de estéril. Essa escolha permitiu considerar diferentes escalas de mudanças na paisagem, avaliando a capacidade do modelo em identificar variações de magnitudes distintas associadas à atividade mineradora. A detecção dessas mudanças é importante considerando os potenciais impactos ambientais relacionados à mineração a céu aberto, tais como alterações na qualidade do solo, da água e do ar. Além disso, há uma carência de estudos que abrangem períodos temporais mais longos, o que dificulta a compreensão completa da dinâmica das áreas mineradas e a generalização dos resultados em diferentes regiões.

Foram utilizados os seguintes parâmetros/hiperparâmetros: 2.000 imagens (T1 2017, T2 2024 e máscara binária) que indicam se houve ou não mudanças nos estados selecionados; tamanho de lote = 16; otimizador AdamW; 50 épocas; taxa de aprendizado de $1e-5$; e peso de classe (*class weights*) de $[0,2, 0,8]$ para lidar com o desbalanceamento das classes. O otimizador AdamW [24] foi utilizado em vez do Adam tradicional devido à sua capacidade de regulagem dos pesos. Enquanto o Adam incorpora o *weight decay* diretamente nos gradientes, o AdamW evita interferências nos momentos de atualização na capacidade de generalização do modelo, sendo mais eficaz em trabalhos de segmentação com classes desbalanceadas.

Tabela I
DESEMPENHO DE TODOS OS MODELOS. EM NEGRITO, OS MELHORES VALORES.

Modelo	Precision	Recall	F1-Score	IoU	Latência (ms)
U-Net	0,9295	0,9243	0,9266	0,6694	55,88
Siamesa U-Net encoder ResNet34	0,9106	0,9120	0,9113	0,6120	72,49
Siamesa U-Net ResNet34 encoder + Attention	0,9127	0,9023	0,9069	0,6122	85,62
Siamesa U-Net encoder ResNet18	0,8959	0,8780	0,8860	0,5597	57,09

A Tabela I apresenta os resultados obtidos pelos modelos. O modelo U-Net apresentou o melhor desempenho global, com valores superiores em todas as métricas: *Precision* (0,9295), *Recall* (0,9243), *F1-Score* (0,9266) e *IoU* (0,6694), além da menor latência média (55,88 ms). Esses resultados indicam que a U-Net foi capaz de identificar mudanças com alta acurácia, mantendo um equilíbrio entre falsos positivos e falsos negativos, além de ser computacionalmente eficiente. A versão Siamesa com *encoder* ResNet34 apresentou desempenho próximo (*F1-Score* de 0,9113 e *IoU* de 0,6120), mas com aumento da latência (72,49 ms). A adição de um mecanismo de atenção a essa arquitetura não resultou em ganhos expressivos em termos de acurácia (*F1-Score* de 0,9069), mas aumentou ainda mais a latência (85,62 ms), o que pode limitar sua aplicabilidade em cenários que exigem inferência em tempo real. Já a versão Siamesa com ResNet18 apresentou os piores resultados, com menor capacidade de detecção (*F1-Score* de 0,8860 e *IoU* de 0,5597), ainda que com latência semelhante à U-Net.

A análise dos resultados obtidos mostra que o modelo U-Net com *encoder* ResNet34 alcançou desempenho consistente, com precisão de 0,9106, revocação de 0,9120 e *F1-score* de 0,9113, além de um *IoU* médio de 0,6120 e latência de 72,49 ms. Já a versão Siamesa U-Net com o mesmo *encoder* e mecanismo de atenção, obteve leve melhora no *IoU* (0,6122), porém apresentou um desequilíbrio entre precisão (0,9127) e revocação (0,9023), resultando em um *F1-score* inferior (0,9069), com aumento considerável na latência (85,62 ms). Isso indica que, embora a adição de atenção possa contribuir marginalmente para a segmentação, o custo computacional torna sua aplicação menos vantajosa no contexto analisado.

Por outro lado, a Siamesa U-Net com *encoder* ResNet18, embora apresente menor latência (57,09 ms), obteve desempenho inferior em todas as métricas (*IoU* de 0,5597 e *F1-score* de 0,8860), o que demonstra que a redução na complexidade do *encoder* compromete significativamente a qualidade da segmentação. Esses resultados indicam que, para o problema em questão, a U-Net tradicional proporciona uma solução mais precisa e computacionalmente eficiente, enquanto as abordagens siamesas introduzem maior complexidade sem ganhos relevantes em desempenho.

O modelo U-Net indica uma precisa capacidade de delimitar as áreas de mudança. A máscara predita indica que essa região foi identificada, tanto em localização quanto em extensão. Isso pode ser visto na Fig 5, onde as áreas de mudança foram captadas, embora uma região de solo exposto tenha sido marcada incorretamente como área de mineração. Esse resultado é relevante no contexto deste trabalho, já que não se busca uma precisão milimétrica nos contornos, mas sim a identificação clara de onde as mudanças estão acontecendo. Apesar da perda de detalhes finos e de alguns erros de detecção, o importante é que o modelo consegue identificar as áreas principais de alteração.

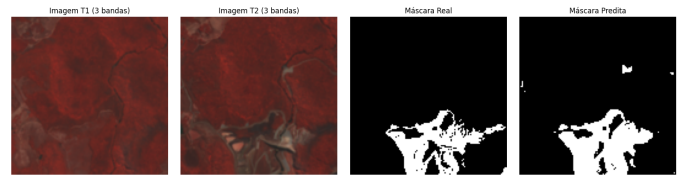


Figura 5. Resultado 1 de detecção de mudança: U-Net. Da esquerda para direita: (i) Imagem no instante T1; (ii) Imagem no instante T2; (iii) Máscara verdade (*ground truth*); e (iv) Máscara predita pelo modelo.

Na Fig 6, ao analisar a matriz de confusão gerada pelo modelo U-Net, é possível observar que o modelo identifica corretamente 95% dos pixels que não correspondem à mineração (verdadeiros negativos) e 63% dos pixels que correspondem à mineração (verdadeiros positivos). Entretanto, a taxa de falsos positivos é de 37%, ou seja, em 37% dos casos em que há mineração, o modelo previu incorretamente que não há mineração.

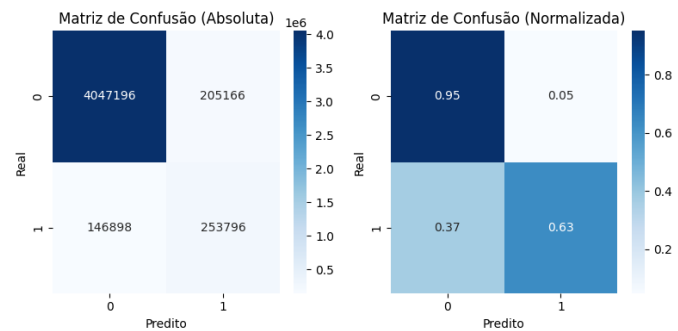


Figura 6. Matriz de confusão: rede U-Net. Da esquerda para direita: (i) Matriz de Confusão Absoluta T1; (ii) Matriz de Confusão Normalizada.

Por mais que a rede Siamesa U-Net ResNet34 *encoder* com o mecanismo de Atenção tenha apresentado um desempenho de 91% nas métricas, a análise visual dos resultados de segmentação revela uma identificação grosseira das mudanças. O modelo consegue indicar a localização geral das alterações, mas tem dificuldade em representar adequadamente a morfologia. A predição é imprecisa em termos de refinamento, tamanho e nível de detalhe, o que compromete a qualidade da segmentação, especialmente em regiões com contornos mais complexos. Isso pode ser observado na Fig 7.

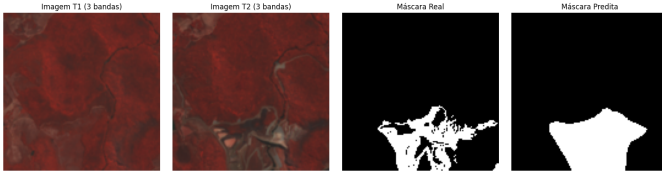


Figura 7. Resultado 1 de detecção de mudança: rede Siamesa U-Net com *encoder* ResNet-34. Da esquerda para direita: (i) Imagem no instante T1; (ii) Imagem no instante T2; (iii) Máscara verdade (*ground truth*); e (iv) Máscara predita pelo modelo.

Observando a Fig 8, verifica-se que o modelo identifica corretamente 94% dos pixels que não representam áreas de mineração e 55% dos pixels que representam áreas de mineração. Considerando os falsos positivos, o modelo identificou incorretamente 6% dos pixels como sendo áreas de mineração.

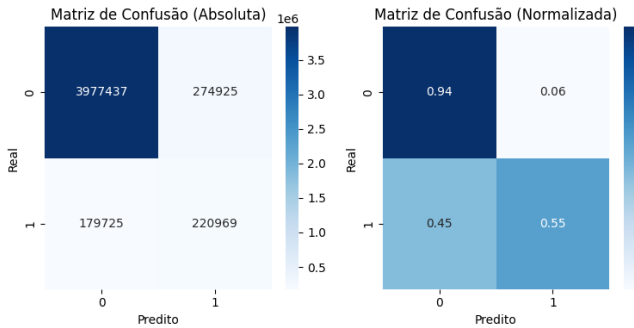


Figura 8. Matriz de confusão: rede U-Net com *encoder* ResNet-34. Da esquerda para direita: (i) Matriz de Confusão Absoluta T1; (ii) Matriz de Confusão Normalizada.

Por mais que a identificação obtida pela rede siamesa com *encoder* ResNet-34 seja incorreta, o modelo demonstrou estar aprendendo, indicando que seria interessante treinar o modelo além de 50 épocas. Podemos visualizar isso na Fig 7, a rede consegue identificar onde está ocorrendo a mudança, mas demonstrou ter problemas para identificar sua morfologia.

Os resultados indicaram uma capacidade superior da U-Net se comparada à rede Siamesa, indicando que é uma potencial candidata para a obtenção de bons resultados aplicados a próximos estudos. Considerando que os resultados do modelo Siamesa com *encoder* ResNet-34 sejam próximos aos da U-Net, o que o faz ser considerado para uma nova investigação aplicada a um contexto multiclasse.

A. Análise via XAI

Visto que o modelo U-Net obteve o melhor desempenho na identificação de mudanças, torna-se relevante compreender como ele chegou a essas decisões. Para isso, aplicamos uma técnica de XAI, utilizando o Grad-CAM para gerar mapas de ativação sobre imagens do conjunto de teste. A Fig9, relativa ao Resultado 1 anterior, apresenta as imagens T1 e T2, os respectivos mapas de ativação obtidos pela técnica Grad-CAM.

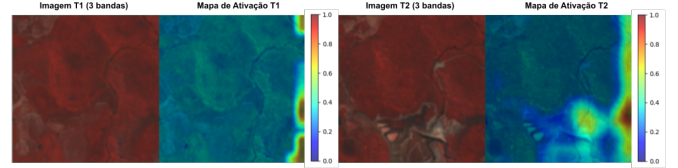


Figura 9. Mapa de Ativação 1: U-Net com Grad-CAM. Da esquerda para direita: (i) Imagem no instante T1; (ii) Mapa de Ativação no instante T1; (iii) Imagem no instante T2; e (iv) Mapa de Ativação no instante T2.

Correlacionando os resultados obtidos pelos mapas de ativação, é possível notar que, no tempo T1, o mapa apresenta uma predominância das cores azul e verde, indicando baixa e média ativação. Ainda assim, observa-se uma região de ativação mais intensa, em amarelo e vermelho, concentrada na borda direita da imagem. Isso sugere que, para a imagem T1, o modelo apresenta maior sensibilidade nas bordas e, possivelmente, a algumas características sutis presentes nessa região. Essa análise pode ser observada na Fig 9.

O mapa de ativação do tempo T2 mostra que o modelo foca intensamente na região inferior direita, onde há características de mudança em uma área de mineração. O modelo identifica que essa característica está presente nessa região inferior e, por isso, concentra ali sua atenção. No entanto, a visão do modelo, como apresentado no Grad-CAM, é mais voltada para uma área geral e menos para contornos finos. Isso se traduz em uma máscara predita que captura a localização geral da mudança, mas suaviza e arredonda os detalhes, resultando na forma de “gota” ou mancha observada.

O falso positivo localizado na máscara predita pode ser resultado de uma ativação sutil em outra parte da imagem T2 que o Grad-CAM não destacou de forma tão proeminente, ou ainda pode representar um ruído na decisão final do modelo.

V. CONCLUSÃO

Este trabalho apresentou uma abordagem, baseada em AP, para a detecção de mudanças em áreas de mineração a céu aberto, por meio de imagens bi-temporais do satélite Sentinel-2. Foram comparados quatro modelos: U-Net padrão, redes Siamesas com *encoder* ResNet-18 e ResNet-34, e uma U-Net com *encoder* ResNet-34 junto a um mecanismo de atenção. Os resultados dos experimentos indicaram que o modelo U-Net apresentou um desempenho superior nas métricas de avaliação, com destaque para o F1-Score de 92,66% e IoU de 66,94%, indicando sua eficiência na delimitação de áreas alteradas. A análise visual da detecção de mudança corrobora

esses resultados, evidenciando que o modelo identifica com precisão a localização e a morfologia das alterações. As redes siamesas, por mais que tenham resultados próximos e inferiores, demonstraram potencial para aplicações futuras. A abordagem proposta indica bons resultados aplicados ao monitoramento automatizado de áreas mineradas.

Para estudos futuros, é proposto o desenvolvimento de uma nova base para o contexto multiclasse, e o cálculo de mais índices específicos, como índice de vegetação de diferença normalizada (NDVI) [25] e índice de água de diferença normalizada (NDWI) [26]. Também é proposto o uso de modelos de aprendizado profundo U-Net ++ e Unet (codificador ResNet-50) adicionado de mecanismo de atenção, visando uma possível melhora na identificação das mudanças.

Além dos resultados quantitativos e qualitativos, a aplicação da técnica de inteligência artificial explicável Grad-CAM permitiu investigar como o modelo U-Net toma suas decisões. Os mapas de ativação evidenciaram que o modelo é capaz de focar em regiões relevantes das imagens multitemporais, especialmente nas áreas alteradas por atividade mineradora. Observou-se que, embora a rede capture a localização geral das mudanças com precisão significativa, ela tende a suavizar os contornos, resultando em máscaras com formas mais arredondadas. A análise revelou, ainda, que o modelo aprende a diferenciar regiões com e sem mudança ao focar em bordas e características sutis das imagens. Essa interpretação dos mapas de ativação reforça a confiança nos resultados obtidos e abre caminho para ajustes e refinamento em modelos futuros.

AGRADECIMENTOS

Esta pesquisa foi desenvolvida no Laboratório de Inteligência Artificial para Aplicações AeroEspaciais e Ambientais (LIAREA) no âmbito do projeto *Classificação de imagens e dados via redes neurais profundas para múltiplos domínios* (IDeepS). O projeto IDeepS (disponível online: <https://github.com/vsantjr/IDeepS> (acessado em 28 de Julho de 2025)) é apoiado pelo Laboratório Nacional de Computação Científica (LNCC, MCTI, Brasil) por meio de recursos do supercomputador SDumont. Este estudo foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

REFERÊNCIAS

- [1] B. Milanez, *Mineração, ambiente e sociedade*. Juiz de Fora: Universidade Federal de Juiz de Fora, 2017. Acesso em: 02 maio 2025.
- [2] Mineração Sustentável, “Mineração a céu aberto,” 2024. Acesso em: 25 maio 2025.
- [3] M. B. Soares, L. R. F. Alleoni, and C. E. P. Cerri, “Garimpos ilegais de ouro podem emitir 3,5 toneladas de carbono por hectare e concentrar mercúrio no solo,” 2024. Disponível em: <https://agencia.fapesp.br/garimpos-ilegais-de-ouro-podem-emitir-35-toneladas-de-carbono-por-hectare->
- [4] L. Cortinhas Ferreira Neto, C. G. Diniz, R. V. Maretto, C. Persello, M. L. Silva Pinheiro, M. C. Castro, L. W. R. Sadeck, A. F. Filho, J. Cansado, A. A. A. Souza, J. P. Feitosa, D. C. Santos, M. Adami, P. W. M. Souza-Filho, A. Stein, A. Biehl, and A. Klautau, “Mineração ilegal descontrolada e garimpo na Amazônia brasileira,” *Nature Communications*, vol. 15, p. 9847, nov 2024.
- [5] MapBiomass, “Garimpo e mineração industrial no Brasil (1985–2022),” 2023. Acesso em: 30 abr. 2025.
- [6] A. E. C. Oliveira, M. C. Domingos, V. S. Júnior, T. S. Körting, C. E. Dias, H. C. d. S. Pereira, and D. C. Mendes, “Mining sites recognition via scene classification using 3d convolutional neural networks,” in *Anais do XXI Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, (Salvador, Brasil), INPE, 2025.
- [7] K. Jabłońska, M. Maksymowicz, D. Tanajewski, W. Kaczan, M. Zięba, and M. Wilgucki, “Minecam: Application of combined remote sensing and machine learning for segmentation and change detection of mining areas enabling multi-purpose monitoring,” *Remote Sensing*, vol. 16, no. 6, 2024.
- [8] A. Huete, “A soil-adjusted vegetation index (savi),” *Remote Sensing of Environment*, vol. 25, no. 3, pp. 295–309, 1988.
- [9] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. Wells, and A. Frangi, eds.), vol. 9351, pp. xx–xx, Springer, Cham, 2015.
- [10] R. Caye Daudt, B. Le Saux, and A. Boulch, “Fully convolutional siamese networks for change detection,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 4063–4067, 2018.
- [11] P. Iakubovskii, “Segmentation models pytorch,” 2019. Disponível em: https://github.com/qubvel/segmentation_models.pytorch. Acesso em: 30 jul. 2025.
- [12] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, “Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization,” *CoRR*, vol. abs/1610.02391, 2016.
- [13] P. Kozińska and J. Górniak-Zimroz, “A review of methods in the field of detecting illegal open-pit mining activities,” *IOP Conference Series: Earth and Environmental Science*, vol. 942, p. 012027, nov 2021.
- [14] Y. Xia, F. Xia, Z. Hui, H. Li, R. Wan, and J. Ai, “Combined ps-insar technology and high-resolution optical remote sensing for identifying illegal underground mining in the suburb of yangquan city, shanxi province, china,” *Remote Sensing*, vol. 15, no. 14, p. 3565, 2023.
- [15] M. Guo, X. Zhang, et al., “Minenetcd: A benchmark dataset and deep learning-based methods for mine change detection from satellite imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 203, pp. 29–46, 2024.
- [16] C. C. Corrigan and S. A. Ikonnikova, “A review of the use of ai in the mining industry: Insights and ethical considerations for multi-objective optimization,” *The Extractive Industries and Society*, vol. 17, p. 101440, 2024.
- [17] European Space Agency, “Copernicus Open Access Hub.” <https://scihub.copernicus.eu/>, 2023. Accessed: 2025-05-25.
- [18] H. C. S. Pereira, V. A. Santiago Júnior, and E. H. Shiguemori, “Change-mining2025: Dataset para detecção de mudanças com sentinel-2,” <https://github.com/herculesc/CHANGE-MINING2025>, 2025. Desenvolvido pelos autores do presente trabalho. Acessado em: Jul. 2025.
- [19] QGIS Development Team, “QGIS Geographic Information System,” 2024. Open Source Geospatial Foundation Project.
- [20] Y. Zhan, J. Yang, M. Gong, Z. Liu, Q. Luo, W. Wang, and G.-S. Xia, “Change detection methods for remote sensing in the last decade: A comprehensive review,” *arXiv preprint arXiv:2305.05813*, 2023.
- [21] D. M. W. Powers, “Evaluation: From precision, recall and f-factor to roc, informedness, markedness correlation,” Tech. Rep. SIE-07-001, School of Informatics and Engineering, Flinders University, 2007. Accessed via Internet Archive on June 8, 2025.
- [22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (voc) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [23] S. Han, H. Mao, and W. J. Dally, “Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding,” in *International Conference on Learning Representations (ICLR)*, 2016.
- [24] I. Loshchilov and F. Hutter, “Fixing weight decay regularization in adam,” *CoRR*, vol. abs/1711.05101, 2017.
- [25] J. Rouse, R. Haas, J. Schell, and D. Deering, “Monitoring vegetation systems in the great plains with erts,” in *Third Earth Resources Technology Satellite-1 Symposium*, vol. 1, pp. 309–317, NASA SP-351, 1974.
- [26] S. McFeeters, “The use of the normalized difference water index (ndwi) in the delineation of open water features,” *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425–1432, 1996.