



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO  
PRÓ-REITORIA DE GRADUAÇÃO  
DEPARTAMENTO DE ENGENHARIAS E TECNOLOGIA  
CURSO BACHARELADO INTERDISCIPLINAR EM TECNOLOGIA DE INFORMAÇÃO

FRANCISCO ADRIAN VINICIUS CHAVES SAMPAIO

**WAMG-UNET: UMA ARQUITETURA DE REDE NEURAL PARA RECONSTRUÇÃO DE  
IMAGENS DE SENSORIAMENTO REMOTO**

PAU DOS FERROS – RN

2026

FRANCISCO ADRIAN VINICIUS CHAVES SAMPAIO

**WAMG-UNET: UMA ARQUITETURA DE REDE NEURAL PARA RECONSTRUÇÃO DE  
IMAGENS DE SENSORIAMENTO REMOTO**

Monografia apresentada à Universidade Federal  
Rural do Semi-Árido como requisito para  
obtenção do título de Bacharel em Bacharelado  
Interdisciplinar em Tecnologia de Informação.

Orientador: Alex Pinheiro Feitosa, Prof.  
Dr.

PAU DOS FERROS – RN

2026

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

S192w Sampaio, Francisco Adrian Vinicius Chaves.  
WAMG-UNet: uma arquitetura de rede neural para  
reconstrução de imagens de sensoriamento remoto /  
Francisco Adrian Vinicius Chaves Sampaio. - 2026.  
76 f. : il.

Orientador: Alex Pinheiro Feitosa.  
Monografia (graduação) - Universidade Federal  
Rural do Semi-árido, Curso de Tecnologia da  
Informação, 2026.

1. Sensoriamento remoto. 2. Reconstrução de  
imagens. 3. Redes neurais. 4. Aprendizado  
profundo. 5. Remoção de nuvens. I. Feitosa, Alex  
Pinheiro, orient. II. Título.

Ficha catalográfica elaborada por sistema gerador automático em conformidade  
com AACR2 e os dados fornecidos pelo) autor(a).

Biblioteca Campus Pau dos Ferros  
Bibliotecária: Erlanda Maria Lopes da Silva  
CRB: 15/966

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

FRANCISCO ADRIAN VINICIUS CHAVES SAMPAIO

**WAMG-UNET: UMA ARQUITETURA DE REDE NEURAL PARA RECONSTRUÇÃO DE  
IMAGENS DE SENSORIAMENTO REMOTO**

Monografia apresentada à Universidade Federal  
Rural do Semi-Árido como requisito para  
obtenção do título de Bacharel em Bacharelado  
Interdisciplinar em Tecnologia de Informação.

Defendida em: 03/06/2026

**BANCA EXAMINADORA**

---

Alex Pinheiro Feitosa, Prof. Dr. (UFERSA)  
Presidente

---

Pedro Thiago Valério de Souza, Prof. Dr. (UFERSA)  
Membro Examinador

---

Paulo Cesar Moura da Silva, Prof. Dr. (UFERSA)  
Membro Examinador

## **AGRADECIMENTOS**

A Deus, princípio de todas as coisas, sem o qual nada seria.

Aos meus pais, Adriano e Núbia, pelo amor incondicional e por acreditarem em mim.

À minha irmã, Thaynan, pelo apoio constante ao longo de todos estes anos.

Aos demais familiares, que, cada um à sua maneira, me deram suporte para chegar até aqui.

Ao meu orientador, Prof. Dr. Alex Pinheiro, pela confiança depositada desde o início e pela paciência, dedicação e tempo investidos para que este trabalho tomasse forma.

Ao Prof. Dr. Pedro Thiago Valério de Souza, por aceitar compor esta banca examinadora, pelas valiosas contribuições a este trabalho e pelos inúmeros ensinamentos ao longo de toda a graduação.

Ao Prof. Dr. Paulo Cesar Moura da Silva, que também aceitou compor esta banca, pelas valiosas contribuições que engrandeceram este trabalho.

A todos os meus amigos, por dividirem comigo as angústias e as alegrias deste percurso.

A todos vocês, minha mais sincera gratidão; este trabalho também é de vocês.

## EPÍGRAFE

*Todos os homens sonham; mas não igualmente. Os que sonham à noite, nos recessos empoeirados de suas mentes, despertam de dia para ver que seus sonhos não passam de vaidade. Os que sonham de dia, porém, são homens perigosos — pois, de olhos abertos, podem agir sobre seus sonhos para torná-los possíveis. (T. E. Lawrence)*

## RESUMO

A presença de nuvens em imagens ópticas de sensoriamento remoto compromete a continuidade e a qualidade das observações da superfície terrestre, dificultando aplicações como monitoramento ambiental, análise de uso e cobertura do solo, agricultura e planejamento territorial. Este trabalho propõe a WAMG-UNet (*Wavelet-Attention Mask-Guided U-Net*), uma arquitetura neural baseada em U-Net para reconstrução de imagens de sensoriamento remoto degradadas por nuvens. A proposta combina três componentes principais: uso explícito de máscara de nuvem, processamento multiescala por meio da Transformada Wavelet Discreta e conexões de salto guiadas por máscara. Além disso, a configuração completa utiliza uma perda auxiliar no domínio wavelet, ponderada por uma estratégia de *Curriculum Learning*. A avaliação experimental foi conduzida nos subconjuntos RICE1 e RICE2, comparando quatro configurações: U-Net RGB, U-Net RGB+Mask, WAMG sem perda wavelet e WAMG-UNet Completa. Os resultados indicaram que a WAMG-UNet Completa obteve o melhor desempenho global, alcançando PSNR de 32,93 dB, SSIM de 0,919 e MAE de 0,0215 no conjunto de validação. A análise incremental mostrou que a maior contribuição quantitativa decorreu da arquitetura WAMG, enquanto a perda wavelet com modulação curricular atuou como componente de refinamento. As análises por subconjunto, cobertura de nuvens e custo computacional indicaram que o modelo proposto apresenta melhor qualidade de reconstrução que os baselines, embora com maior latência de inferência. Conclui-se que a WAMG-UNet constitui uma alternativa viável para reconstrução supervisionada de imagens ópticas afetadas por nuvens, especialmente em cenários de processamento offline ou em lote.

**Palavras-chave:** sensoriamento remoto; reconstrução de imagens; redes neurais; aprendizado profundo; remoção de nuvens

## ABSTRACT

The presence of clouds in optical remote sensing images compromises the continuity and quality of Earth surface observations, affecting applications such as environmental monitoring, land use and land cover analysis, agriculture, and territorial planning. This work proposes WAMG-UNet (*Wavelet-Attention Mask-Guided U-Net*), a neural architecture based on U-Net for reconstructing remote sensing images degraded by clouds. The proposed approach combines three main components: explicit use of cloud masks, multiscale processing through the Discrete Wavelet Transform, and mask-guided skip connections. In addition, the full configuration employs an auxiliary wavelet-domain loss weighted by a *Curriculum Learning* strategy. The experimental evaluation was conducted on the RICE1 and RICE2 subsets, comparing four configurations: U-Net RGB, U-Net RGB+Mask, WAMG without wavelet loss, and WAMG-UNet. The results showed that Full WAMG-UNet achieved the best overall performance, reaching a PSNR of 32.93 dB, an SSIM of 0.919, and an MAE of 0.0215 on the validation set. The incremental analysis indicated that the main quantitative contribution came from the WAMG architecture, while the wavelet loss with curriculum modulation acted as a refinement component. The analyses by subset, cloud coverage, and computational cost showed that the proposed model provides better reconstruction quality than the baselines, although with higher inference latency. Therefore, WAMG-UNet is a viable alternative for supervised reconstruction of optical images affected by clouds, especially in offline or batch-processing scenarios.

**Keywords:** remote sensing; image reconstruction; neural networks; deep learning; cloud removal

## LISTA DE TABELAS

|                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------|----|
| Tabela 1 – Modelos avaliados e componentes utilizados em cada configuração. . . . .                       | 52 |
| Tabela 2 – Modelos avaliados no protocolo experimental. . . . .                                           | 60 |
| Tabela 3 – Resumo do treinamento da WAMG-UNet Completa. . . . .                                           | 61 |
| Tabela 4 – Comparação quantitativa global entre os modelos avaliados no conjunto de<br>validação. . . . . | 62 |
| Tabela 5 – Desempenho em PSNR por subconjunto do dataset RICE. . . . .                                    | 63 |
| Tabela 6 – PSNR por intervalo de cobertura de nuvens. . . . .                                             | 64 |
| Tabela 7 – Comparação de eficiência computacional entre os modelos avaliados. . . . .                     | 68 |
| Tabela 8 – Análise incremental da contribuição dos componentes avaliados. . . . .                         | 68 |
| Tabela 9 – Comparação indicativa com a literatura no dataset RICE. . . . .                                | 70 |

## SUMÁRIO

|            |                                                               |    |
|------------|---------------------------------------------------------------|----|
| <b>1</b>   | <b>INTRODUÇÃO</b>                                             | 14 |
| <b>2</b>   | <b>REFERENCIAL TEÓRICO</b>                                    | 17 |
| <b>2.1</b> | <b>Sensoriamento Remoto Orbital</b>                           | 17 |
| 2.1.1      | Princípios Físicos                                            | 17 |
| 2.1.2      | Sensores Passivos e Modelagem Radiométrica                    | 19 |
| 2.1.2.1    | Conjunto de Dados                                             | 21 |
| 2.1.3      | Oclusão Atmosférica: Nuvens e Sombras                         | 22 |
| <b>2.2</b> | <b>Redes Neurais Artificiais</b>                              | 24 |
| 2.2.1      | Redes Neurais Convolucionais (CNN)                            | 28 |
| 2.2.1.1    | Arquiteturas de Codificação e Decodificação (Encoder-Decoder) | 30 |
| 2.2.2      | Mecanismos de Atenção                                         | 32 |
| 2.2.2.1    | Atenção de Canal ( <i>Channel Attention</i> )                 | 32 |
| 2.2.2.2    | Atenção Espacial ( <i>Spatial Attention</i> )                 | 32 |
| 2.2.2.3    | Atenção Guiada por Máscara ( <i>Mask Gated Attention</i> )    | 33 |
| 2.2.3      | Funções de Perda e Estratégias de Otimização                  | 34 |
| 2.2.3.1    | Perdas no Domínio Espacial                                    | 34 |
| 2.2.3.2    | Viés Espectral e Perdas no Domínio da Frequência              | 35 |
| 2.2.3.3    | Curriculum Learning e Agendamento de Perda                    | 36 |
| <b>2.3</b> | <b>Processamento de Sinais: A Transformada Wavelet</b>        | 36 |
| 2.3.1      | A Transformada Wavelet Discreta (DWT) em Imagens              | 37 |
| 2.3.2      | Wavelets em Redes Neurais                                     | 38 |
| <b>2.4</b> | <b>Estado da Arte em Detecção e Mascaramento</b>              | 39 |
| 2.4.1      | Limiares Físicos e Machine Learning Tradicional               | 39 |
| 2.4.2      | O Estado da Arte: Deep Learning e Segmentação Semântica       | 40 |
| <b>2.5</b> | <b>Estratégias para Reconstrução e Remoção de Nuvens</b>      | 41 |

|            |                                                           |           |
|------------|-----------------------------------------------------------|-----------|
| 2.5.1      | Fusão Multimodal: SAR-Óptico . . . . .                    | 41        |
| 2.5.2      | Modelos Generativos: Difusão e GANs . . . . .             | 42        |
| 2.5.3      | Aprendizado no Domínio da Frequência . . . . .            | 43        |
| <b>3</b>   | <b>METODOLOGIA . . . . .</b>                              | <b>46</b> |
| <b>3.1</b> | <b>Conjunto de Dados e Formulação da Tarefa . . . . .</b> | <b>46</b> |
| 3.1.1      | Conjunto de Dados RICE . . . . .                          | 46        |
| 3.1.2      | Particionamento dos Dados . . . . .                       | 47        |
| 3.1.3      | Geração de Máscaras no RICE1 . . . . .                    | 47        |
| 3.1.4      | Filtragem e Normalização . . . . .                        | 49        |
| <b>3.2</b> | <b>Arquitetura Proposta: WAMG-UNet . . . . .</b>          | <b>49</b> |
| 3.2.1      | Entrada Guiada por Máscara . . . . .                      | 49        |
| 3.2.2      | Subamostragem via Wavelet Pooling . . . . .               | 50        |
| 3.2.3      | Reconstrução via Transformada Wavelet Inversa . . . . .   | 50        |
| 3.2.4      | Conexões de Salto Guiadas por Máscara . . . . .           | 51        |
| <b>3.3</b> | <b>Modelos Comparados e Estudo de Ablação . . . . .</b>   | <b>52</b> |
| <b>3.4</b> | <b>Estratégia de Otimização e Treinamento . . . . .</b>   | <b>53</b> |
| 3.4.1      | Função de Perda . . . . .                                 | 53        |
| 3.4.2      | Curriculum Learning aplicado à perda wavelet . . . . .    | 53        |
| 3.4.3      | Otimização . . . . .                                      | 54        |
| 3.4.4      | Média Móvel Exponencial dos Pesos . . . . .               | 54        |
| 3.4.5      | Regularização . . . . .                                   | 55        |
| <b>3.5</b> | <b>Métricas de Avaliação . . . . .</b>                    | <b>55</b> |
| 3.5.1      | PSNR . . . . .                                            | 55        |
| 3.5.2      | SSIM . . . . .                                            | 56        |
| 3.5.3      | MAE . . . . .                                             | 57        |
| 3.5.4      | Indicadores de Custo Computacional . . . . .              | 58        |

|            |                                       |           |
|------------|---------------------------------------|-----------|
| <b>4</b>   | <b>RESULTADOS E DISCUSSÃO</b>         | <b>59</b> |
| <b>4.1</b> | <b>Configuração Experimental</b>      | <b>59</b> |
| 4.1.1      | Treinamento do Modelo Proposto        | 60        |
| <b>4.2</b> | <b>Resultados</b>                     | <b>61</b> |
| 4.2.1      | Comparação Global                     | 62        |
| 4.2.2      | Análise por Subconjuntos              | 63        |
| 4.2.3      | Análise por Cobertura de Nuvem        | 64        |
| 4.2.4      | Métricas Estruturais                  | 64        |
| 4.2.5      | Análise Qualitativa                   | 65        |
| <b>4.3</b> | <b>Análise de Custo Computacional</b> | <b>67</b> |
| <b>4.4</b> | <b>Discussão</b>                      | <b>68</b> |
| 4.4.1      | Comparação com a Literatura           | 70        |
| <b>5</b>   | <b>CONCLUSÃO</b>                      | <b>72</b> |
| <b>5.1</b> | <b>Considerações Finais</b>           | <b>72</b> |
|            | <b>REFERÊNCIAS</b>                    | <b>75</b> |

## 1 INTRODUÇÃO

O sensoriamento remoto orbital constitui uma das principais ferramentas para observação sistemática da superfície terrestre. Por meio de sensores embarcados em satélites, é possível adquirir imagens em diferentes escalas espaciais e temporais, permitindo aplicações em monitoramento ambiental, agricultura, planejamento urbano, análise de uso e cobertura do solo, gestão de recursos hídricos e acompanhamento de desastres naturais. Em especial, imagens ópticas apresentam grande relevância por registrarem informações radiométricas associadas à reflectância da superfície em faixas do espectro visível e infravermelho, possibilitando a identificação de padrões espaciais, texturas e alvos de interesse.

Apesar de sua ampla aplicabilidade, a aquisição de imagens ópticas orbitais é fortemente limitada por condições atmosféricas. A presença de nuvens e sombras pode degradar parcial ou totalmente a informação observada pelo sensor, comprometendo a continuidade espacial e temporal dos dados. Em nuvens finas ou semitransparentes, parte do sinal da superfície ainda pode estar presente, embora atenuado e contaminado por espalhamento atmosférico. Em nuvens espessas, por outro lado, a informação da superfície pode ser severamente bloqueada, tornando a reconstrução uma tarefa substancialmente mais difícil.

A remoção de nuvens em imagens de sensoriamento remoto pode ser compreendida como um problema de reconstrução de imagem. O objetivo consiste em estimar uma versão livre de nuvens a partir de uma imagem degradada, utilizando informações espaciais, radiométricas e, quando disponíveis, máscaras que indicam as regiões ocluídas. No entanto, trata-se de um problema mal-posto, pois em áreas completamente encobertas a informação real da superfície pode estar ausente na imagem de entrada. Nesses casos, o modelo precisa inferir o conteúdo com base no contexto espacial e nos padrões aprendidos durante o treinamento.

Nos últimos anos, métodos baseados em aprendizado profundo têm se destacado em tarefas de restauração e reconstrução de imagens. Arquiteturas convolucionais do tipo *encoder-decoder*, como a U-Net, são particularmente relevantes por combinarem extração hierárquica de características com recuperação progressiva da resolução espacial. Suas conexões de salto permitem transferir informações de alta resolução do codificador para o decodificador, favorecendo a preservação de detalhes locais. Entretanto, no contexto de imagens contaminadas por nuvens, essas mesmas conexões podem transportar artefatos atmosféricos para a etapa de reconstrução. Além disso, operações tradicionais de subamostragem, como *max pooling*, podem descartar componentes de alta frequência importantes para a preservação de bordas, texturas e transições

espaciais.

Diante dessas limitações, este trabalho propõe a WAMG-UNet (*Wavelet-Attention Mask-Guided U-Net*), uma arquitetura neural voltada à reconstrução de imagens ópticas de sensoriamento remoto afetadas por nuvens. A proposta parte da estrutura *encoder-decoder* da U-Net, mas incorpora mecanismos específicos para lidar com a degradação atmosférica. Em particular, a arquitetura utiliza a máscara de nuvem como informação auxiliar, emprega operações baseadas na Transformada Wavelet Discreta para preservar componentes multiescala e introduz conexões de salto guiadas por máscara, com o objetivo de reduzir a propagação direta de artefatos da região ocluída para a imagem reconstruída.

A hipótese central deste trabalho é que a combinação entre orientação explícita por máscara, processamento multiescala no domínio wavelet e modulação das conexões de salto pode melhorar a qualidade da reconstrução em comparação com baselines convolucionais convencionais. Além disso, investiga-se o impacto de uma perda auxiliar no domínio wavelet, ponderada por uma estratégia de *Curriculum Learning*, aplicada à configuração completa da arquitetura proposta. Essa estratégia busca priorizar inicialmente a reconstrução espacial global e, posteriormente, aumentar a importância da preservação de componentes de frequência.

Como principal contribuição deste trabalho, destaca-se a modelagem do mecanismo de Atenção Guiada por Máscara (*Mask Gated Attention*). Diferente das *skip connections* tradicionais, que transferem mapas de características integralmente e correm o risco de propagar o ruído atmosférico, este mecanismo atua como um filtro espacial adaptativo, utilizando a máscara de nuvem para modular diferencialmente a transferência de características entre regiões ocluídas e estruturais.

O objetivo geral deste trabalho é desenvolver e avaliar uma arquitetura baseada em aprendizado profundo para reconstrução de imagens de sensoriamento remoto degradadas por nuvens. Para alcançar esse objetivo, foram definidos os seguintes objetivos específicos:

- implementar a arquitetura WAMG-UNet, com ênfase na integração do processamento multiescala baseado em wavelets e no mecanismo de atenção *Mask Gated Attention*;
- estabelecer um protocolo experimental estruturado nos subconjuntos RICE1 e RICE2, garantindo a avaliação em diferentes regimes de degradação atmosférica;
- avaliar o desempenho do modelo proposto em relação a *baselines* baseados em U-Net, através de métricas de qualidade de imagem (PSNR, SSIM, MAE), eficiência computacional e análise visual;

- conduzir um estudo de ablação para isolar e validar a contribuição técnica de cada componente proposto, incluindo a estratégia de *Curriculum Learning* no domínio da frequência.

## 2 REFERENCIAL TEÓRICO

### 2.1 Sensoriamento Remoto Orbital

O sensoriamento remoto, em sua acepção física fundamental, consiste na obtenção de informações sobre um objeto ou fenômeno a partir de uma distância, sem contato físico direto, mediante o emprego de instrumentos capazes de mensurar a radiação eletromagnética (Elachi; Zyl, 2021). No contexto da observação da Terra, essa definição se especializa. O sensoriamento remoto orbital integra sensores embarcados em plataformas espaciais para registrar a radiação proveniente da superfície terrestre e da atmosfera. Esses instrumentos detectam e quantificam a interação da radiação eletromagnética com os constituintes físico-químicos dos alvos, gerando dados que, após processamento, subsidiam a extração de informações geofísicas (Novo, 2008).

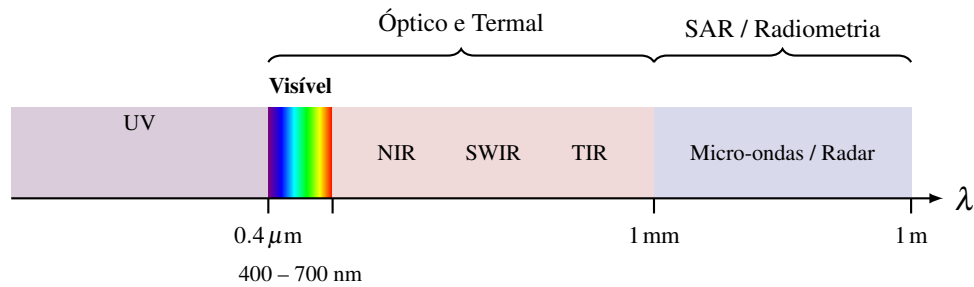
Entretanto, a aquisição de dados por sensoriamento remoto por meio óptico é intrinsecamente limitada por alguns fatores ambientais. A presença de nuvens, por exemplo, pode obstruir o campo de visada do sensor sobre a superfície, resultando em dados ausentes ou corrompidos. Além disso, as sombras projetadas por nuvens podem criar regiões de baixa radiação que são frequentemente confundidas com corpos d'água ou vegetação densa, introduzindo ambiguidades na interpretação dos dados.

Esses desafios motivam o desenvolvimento de técnicas avançadas de processamento de imagens para a reconstrução de informações perdidas devido à oclusão atmosférica.

#### 2.1.1 Princípios Físicos

A aquisição de dados em sensoriamento remoto baseia-se na detecção da Radiação Eletromagnética (REM) refletida, espalhada ou emitida pela superfície terrestre. O fluxo de informações entre o alvo e o sensor é codificado nas propriedades da onda, notadamente: frequência (ou comprimento de onda), intensidade, polarização e fase (Elachi; Zyl, 2021). O espectro eletromagnético de interesse para o sensoriamento remoto orbital concentra-se nas janelas atmosféricas — faixas onde a transmitância atmosférica é elevada —, abrangendo desde a região do visível e infravermelho até as micro-ondas. A Figura 1 apresenta as principais regiões espectrais utilizadas na observação da Terra.

Figura 1 – Divisão do espectro eletromagnético com ênfase nas faixas utilizadas em sensoria-  
mento remoto.

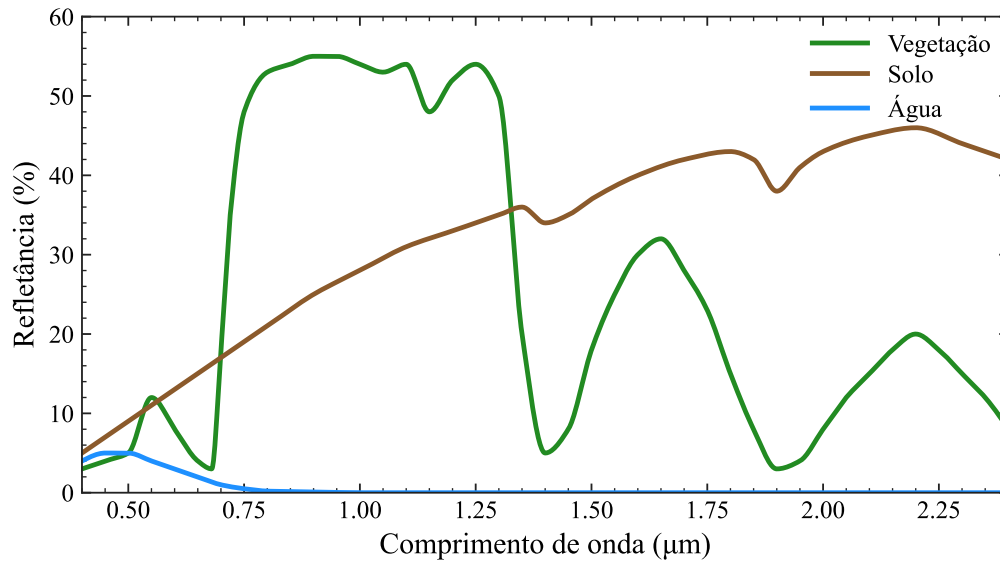


Fonte: Adaptado de Elachi (2021).

Para viabilizar a identificação dos alvos, deve-se considerar a sua interação com a radiação incidente — mediada por mecanismos de absorção, transmissão e espalhamento. A resposta radiométrica específica de um material ao longo do espectro eletromagnético constitui a sua assinatura espectral, ferramenta fundamental para a discriminação de coberturas na superfície terrestre. A Figura 2 ilustra o comportamento radiométrico de três alvos macroscópicos fundamentais: vegetação fotossinteticamente ativa, solo exposto e água clara, utilizando dados espectrométricos medidos em laboratório e catalogados na biblioteca espectral do Serviço Geológico dos Estados Unidos (USGS) (Kokaly *et al.*, 2017).

Pode-se observar que a vegetação apresenta forte absorção nas bandas do azul e do vermelho (devido à presença de clorofila), um leve pico de reflectância no verde e um aumento abrupto no infravermelho próximo, decorrente da estrutura celular interna das folhas. O solo exposto exibe uma curva de reflectância de crescimento monotônico do visível em direção ao infravermelho, enquanto a água atua como um forte absorvedor, com reflectância restrita à região do visível e tendendo a zero no infravermelho próximo e de ondas curtas (NIR e SWIR).

Figura 2 – Assinaturas espectrais de alvos típicos. A variação da reflectância ao longo do espectro é a base para a discriminação de objetos em imagens orbitais. Dados adaptados de (Kokaly *et al.*, 2017).



Fonte: Adaptado de Kokaly *et al.* (2017).

### 2.1.2 Sensores Passivos e Modelagem Radiométrica

Sensores passivos, como os embarcados nos satélites de observação da Terra, operam mensurando a radiância proveniente de fontes externas de iluminação — predominantemente o Sol — após a sua complexa interação com a superfície terrestre e a atmosfera. A magnitude do sinal registrado pelo sensor orbital, denominada Radiância no Topo da Atmosfera ( $L_{TOA}$ ), é governada pela Equação de Transferência Radiativa, que descreve os fenômenos físicos de absorção e espalhamento da radiação eletromagnética pelo meio óptico (Schott, 2007).

O fluxo radiante que incide inicialmente no sistema é definido pela Irradiância Solar Espectral Exoatmosférica ( $E_{\lambda}^0$ ). Este parâmetro físico quantifica a densidade de fluxo de energia eletromagnética não atenuada irradiada pelo Sol, por unidade de área e por unidade de comprimento de onda (tipicamente expressa em  $W \cdot m^{-2} \cdot \mu m^{-1}$ ), incidindo sobre uma superfície estritamente perpendicular aos raios solares no limite superior da atmosfera, a uma distância astronômica média exata de 1 Unidade Astronômica (UA).

No entanto, a irradiância efetiva que atinge o topo da atmosfera ( $E_{\lambda}^{TOA}$ ) no instante específico da passagem do satélite apresenta variações induzidas pela geometria orbital e pela posição solar, sendo modelada por (Schott, 2007):

$$E_{\lambda}^{\text{TOA}} = E_{\lambda}^0 \left( \frac{d_0}{d} \right)^2 \cos(\theta_s) \quad (2.1)$$

onde  $d$  representa a distância real Terra-Sol na data de aquisição,  $d_0$  é a distância média de referência (1 UA) e  $\theta_s$  é o ângulo zenital solar. O termo  $\left( \frac{d_0}{d} \right)^2$  realiza a correção da variação sazonal da distância Terra-Sol, enquanto o fator cosseno projeta o fluxo incidente sobre uma superfície plana e horizontal, ajustando a densidade de energia em função da elevação solar.

A radiância total espectral captada pelo sensor ( $L_{\text{TOA}}$ ), expressa em  $\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \mu\text{m}^{-1}$ , não é constituída unicamente pelos fótons refletidos de forma direta pelo alvo de interesse. O sinal sofre perturbações significativas decorrentes das interações atmosféricas e da refletância da vizinhança. O modelo radiométrico pode ser decomposto em três componentes aditivos fundamentais (Schott, 2007):

$$L_{\text{TOA}}(\lambda) = \underbrace{L_{\text{sup}}(\lambda) \cdot \tau_{\uparrow}(\lambda)}_{\text{superfície atenuada}} + \underbrace{L_{\text{adj}}(\lambda)}_{\text{adjacência}} + \underbrace{L_{\text{atm}}(\lambda)}_{\text{caminho}} \quad (2.2)$$

onde:

$L_{\text{sup}} \rightarrow$  radiância emergente da superfície

$\tau_{\uparrow} \rightarrow$  transmitância atmosférica ascendente (Superfície–Sensor)

$L_{\text{adj}} \rightarrow$  radiância de adjacência (espalhamento lateral)

$L_{\text{atm}} \rightarrow$  radiância de caminho (*path radiance*)

Para uma caracterização física rigorosa, a parcela do sinal proveniente da superfície ( $L_{\text{sup}}$ ) deve compilar a iluminação direta atenuada, a iluminação difusa (espalhamento descendente) e os efeitos não-lineares de espalhamento múltiplo confinados entre a superfície e a baixa atmosfera. O termo de superfície é expandido na seguinte formulação:

$$L_{\text{sup}}(\lambda) = \frac{\rho_{\lambda}}{\pi(1 - \rho_{\lambda}S)} \left[ E_{\lambda}^{\text{TOA}} \cdot \tau_{\downarrow}(\lambda) + E_{\lambda}^{\text{dif}} \right] \quad (2.3)$$

onde:

$\rho_\lambda$  reflectância de superfície (aproximação Lambertiana da BRDF)

Salbedo esférico da atmosfera (espalhamento múltiplo)

$\tau_\downarrow$  transmitância atmosférica descendente (Sol–Superfície)

$E_\lambda^{\text{dif}} \rightarrow$  irradiância difusa descendente (espalhamento atmosférico)

A presença de nuvens, sejam nuvens cúmulos espessos ou cirrus translúcidos, altera drasticamente este balanço de energia. Tais formações meteorológicas não apenas reduzem a transmitância ( $\tau_\downarrow$  e  $\tau_\uparrow$ ) a valores próximos de zero em certas espessuras ópticas, mas também maximizam a radiância de caminho ( $L_{\text{atm}}$ ). A alta não-linearidade das Equações 2.2 e 2.3 sob interferência de nuvens densas impõe limites severos às inversões analíticas tradicionais. Essa limitação física inerente fundamenta a necessidade da implementação de arquiteturas baseadas em inteligência artificial, capazes de mapear e isolar os padrões não-lineares de atenuação para estimar a reflectância de superfície ( $\rho_\lambda$ ) subjacente de maneira mais robusta.

#### 2.1.2.1 Conjunto de Dados

A validação experimental deste trabalho utiliza o *Remote Sensing Image Cloud rEmoving dataset* (RICE), proposto por (Lin *et al.*, 2019) com o objetivo explícito de suprir a escassez de dados anotados para o treinamento de redes neurais na tarefa de remoção de nuvens em imagens de sensoriamento remoto óptico. O RICE é estruturado em dois subconjuntos complementares, RICE1 e RICE2, que diferem tanto na fonte dos dados quanto na tipologia das oclusões representadas.

O RICE1 é composto por 500 pares de imagens adquiridas via Google Earth, com dimensão padronizada de  $512 \times 512$  pixels. Cada par contém uma imagem com cobertura de nuvens finas (*thin clouds*) e sua correspondente imagem de referência sem oclusão, permitindo ao modelo aprender o mapeamento direto entre a cena nublada e a superfície subjacente. A natureza semitransparente das nuvens presentes no RICE-I implica que o sinal de superfície não é completamente obliterado, mas sofre atenuação e distorção espectral consistente com os termos  $\tau_\downarrow$  e  $L_{\text{atm}}$  da Equação 2.2.

O RICE2 é composto por 736 conjuntos de imagens provenientes do sensor *Operational Land Imager* (OLI) e do *Thermal Infrared Sensor* (TIRS), a bordo do satélite Landsat 8. Cada conjunto contém três elementos: a imagem de referência livre de nuvens, a imagem com cobertura

de nuvens espessas (*thick clouds*) e a respectiva máscara binária de nuvens. Pares com e sem nuvens foram coletados para a mesma localização geográfica em um intervalo máximo de 15 dias, garantindo variação mínima nas condições de superfície. A cobertura média de nuvens nas imagens do RICE-II é de aproximadamente 24%. Neste subconjunto, a espessura óptica das nuvens é suficiente para reduzir a transmitância a valores próximos de zero, tornando a Equação 2.3 mal condicionada para inversão analítica e a informação de superfície praticamente irre recuperável por métodos clássicos.

A limitação crítica dos sensores passivos ópticos reside na opacidade atmosférica induzida pela cobertura de nuvens. Na presença de nuvens espessas, como as representadas no RICE-II, a transmissividade atmosférica tende a zero ( $\tau \rightarrow 0$ ) nas faixas do visível e infravermelho. Sob essas condições, a Equação 2.3 torna-se mal condicionada para inversão, pois a radiância registrada pelo sensor é dominada pela alta reflectância do topo da nuvem, obliterando o sinal proveniente da superfície ( $L_{\text{sup}}$ ). Este fenômeno de oclusão exige a aplicação de técnicas de reconstrução de dados (*inpainting*) baseadas em contextos espaciais ou temporais para a recuperação da informação perdida.

### 2.1.3 Oclusão Atmosférica: Nuvens e Sombras

A presença de nuvens representa a principal fonte de perda de informação em sensores ópticos passivos. A contaminação por nuvens é um fenômeno físico estruturado, governado pela interação da radiação com hidrometeoros e aerossóis. A degradação da imagem ocorre por processos de espalhamento, classificados conforme a relação entre o diâmetro da partícula espalhadora ( $D$ ) e o comprimento de onda ( $\lambda$ ) (Schott, 2007).

O espalhamento Rayleigh ( $D \ll \lambda$ ), predominante na interação com moléculas gasosas, apresenta intensidade inversamente proporcional a  $\lambda^4$ , causando um desvio ao azul, sendo responsável pela radiância de caminho base e a coloração azulada do céu. Em aerossóis e gotículas nas bordas de nuvens, predomina o espalhamento Mie ( $D \approx \lambda$ ), que atua como um filtro passa-baixas, reduzindo o contraste local e degradando as componentes de alta frequência da imagem. Por fim, o espalhamento não-seletivo ( $D \gg \lambda$ ), característico de nuvens espessas compostas por gotas e cristais de gelo, afeta uniformemente o espectro visível, resultando em saturação radiométrica e no bloqueio total do sinal proveniente da superfície.

A nuvem introduz uma dualidade de problemas: a adição de radiância espalhada (o brilho da nuvem) e a atenuação do sinal da superfície. Inspirando-se conceitualmente no Modelo de

Formação de Imagem com Neblina proposto por He et al. (He; Sun; Tang, 2009), adapta-se a fenomenologia para o domínio radiométrico. Neste trabalho, a mistura de sinais em pixels contaminados é aproximada pela seguinte equação linear:

$$L_{TOA}(\lambda) = L_{sup}(\lambda) \cdot \tau_{cloud}(\lambda) + L_{nuvem}(\lambda) \cdot (1 - \tau_{cloud}(\lambda)) \quad (2.4)$$

Nesta formulação, o modelo comporta-se como uma mistura linear governada pela opacidade da nuvem. O primeiro termo,  $L_{sup} \cdot \tau_{cloud}$ , representa a atenuação do sinal da superfície: quanto maior a espessura óptica da nuvem, menor o valor de  $\tau_{cloud}$  e, conseqüentemente, menor a contribuição espectral do alvo ao sensor. O segundo termo,  $L_{nuvem} \cdot (1 - \tau_{cloud})$ , representa a radiância aditiva intrínseca à nuvem, que passa a dominar o sinal registrado à medida que a transmissividade se aproxima de zero. Trata-se de uma modelagem simplificada para análise e interpretação do problema, não substituindo a formulação completa da transferência radiativa apresentada nas Equações 2.2 e 2.3.

A análise da Equação 2.4 permite categorizar o problema de reconstrução em dois regimes distintos, definidos pela magnitude da transmissividade da nuvem ( $\tau_{cloud}$ ):

**Nuvens Finas (Translucidez):** Ocorre quando  $0 < \tau_{cloud} < 1$ . Neste cenário, o primeiro termo da equação ( $L_{sup} \cdot \tau_{cloud}$ ) permanece significativo, indicando que o sinal da superfície atinge o sensor, embora atenuado e contaminado pela radiância aditiva da nuvem ( $L_{nuvem}$ ). A recuperação da informação é tratada como um problema de *dehazing* (remoção de neblina), onde o objetivo é estimar os parâmetros  $\tau_{cloud}$  e  $L_{nuvem}$  para inverter a equação e restaurar o contraste e a radiometria original de  $L_{sup}$ .

**Nuvens Espessas (Oclusão Total):** Ocorre quando  $\tau_{cloud} \rightarrow 0$ . Sob esta condição, o termo referente à superfície é anulado ( $L_{sup} \cdot \tau_{cloud} \approx 0$ ), e a radiância registrada  $L_{TOA}$  converge para a radiância intrínseca da nuvem ( $L_{nuvem}$ ). Como a informação espectral da superfície é fisicamente bloqueada, a inversão matemática direta torna-se impossível. A reconstrução transita para um problema de *inpainting* (restauração semântica), exigindo que o modelo infira as características perdidas (textura e espectro) baseando-se em correlações espaciais da vizinhança livre de nuvens ou em dados temporais históricos.

Concomitantemente à oclusão, deve-se considerar o efeito das sombras projetadas. Enquanto a nuvem atua no trajeto ascendente da radiação (bloqueando a visada do sensor), a sombra atua no trajeto descendente, bloqueando a irradiância solar direta ( $E_{\lambda}^{TOA}$ ). Nas regiões

sombreadas, o alvo é iluminado exclusivamente pela componente difusa ( $E_{\lambda}^{dif}$ ), resultando em uma redução drástica da radiância refletida  $L_{sup}$ . Isso gera "falsos alvos" escuros que, embora não estejam oclusos, apresentam radiometria severamente alterada, frequentemente confundida com corpos d'água por algoritmos de classificação (Zhu; Woodcock, 2012). Em imagens de alta resolução espacial como as do Sentinel-2, a geometria de aquisição introduz ainda um deslocamento espacial (paralaxe) entre a posição da nuvem e sua respectiva sombra, cuja magnitude varia em função da altura da nuvem e do ângulo zenital solar, adicionando uma camada de complexidade à detecção e remoção conjunta destes artefatos.

A persistência desses efeitos na faixa óptica impõe um desafio de continuidade de dados. Uma vez que o bloqueio físico impede a aquisição direta da radiância, a recuperação da informação transita de um problema de sensoriamento físico para um problema de estimativa estatística e aprendizado de representação.

## 2.2 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) representam uma classe de modelos estatísticos de aprendizado inspirados na arquitetura e funcionamento do sistema nervoso biológico. A unidade elementar de processamento é o neurônio artificial, cuja formalização matemática foi estabelecida por McCulloch e Pitts (McCulloch; Pitts, 1943). Neste modelo, o neurônio atua como uma porta lógica binária que integra sinais de entrada através de uma soma ponderada, disparando apenas se o potencial de ativação ultrapassar um limiar intrínseco.

A evolução para sistemas adaptativos ocorreu com o advento do Perceptron, proposto por (Rosenblatt, 1958). A inovação crucial residiu na introdução de pesos sinápticos ajustáveis e do termo de viés (*bias*,  $b$ ), que matematicamente desacopla o limiar de ativação da origem do sistema de coordenadas.

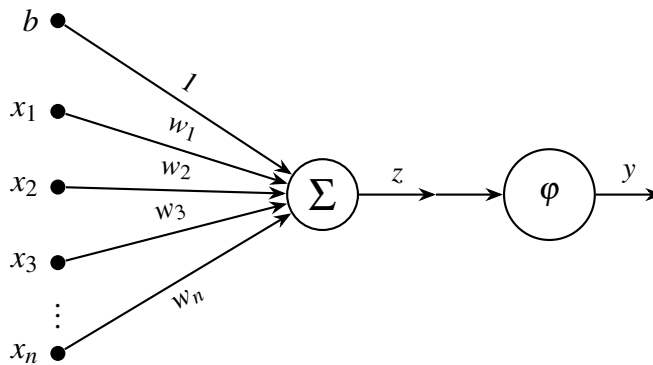
A operação do neurônio pode ser descrita como uma transformação afim seguida de uma não-linearidade pontual. Definindo o potencial de ativação  $z \in \mathbb{R}$  como o produto interno entre o vetor de pesos  $\mathbf{w} \in \mathbb{R}^n$  e o vetor de entrada  $\mathbf{x} \in \mathbb{R}^n$ , acrescido do viés  $b \in \mathbb{R}$ , a saída  $y \in \mathbb{R}$  é dada por  $y = \varphi(z)$ :

$$z = \mathbf{w}^T \mathbf{x} + b = \sum_{i=1}^n w_i x_i + b, \quad y = \varphi(z) \quad (2.5)$$

onde  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  é a função de ativação, aplicada elemento a elemento ao potencial  $z$ . Geometri-

camente, o conjunto  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{w}^T \mathbf{x} + b = 0\}$  define um hiperplano afim de dimensão  $n - 1$  no espaço de entrada. O aprendizado no Perceptron original consiste no ajuste iterativo de  $(\mathbf{w}, b)$  para posicionar  $\mathcal{H}$  de modo a separar linearmente as classes no espaço  $\mathbb{R}^n$ . A Figura 3 ilustra o esquema de funcionamento de um neurônio artificial.

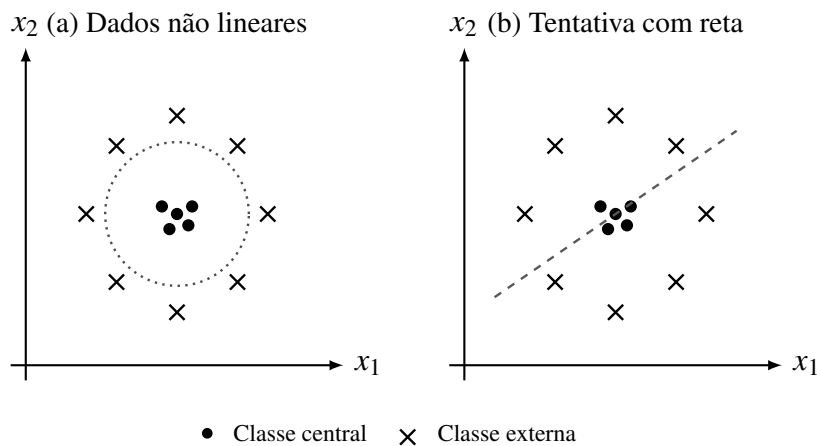
Figura 3 – Modelo esquemático de um neurônio artificial (Perceptron).



Fonte: Autoria própria (2026).

Contudo, Minsky e Papert (Minsky; Papert, 1969) demonstraram formalmente que essa arquitetura de camada única é restrita a conjuntos linearmente separáveis, falhando em representar funções lógicas não lineares. Por exemplo, se houver duas classes concêntricas no plano: uma classe compacta na região central e outra distribuída ao redor. Como ilustra a Figura 4, nesse cenário a fronteira ideal é curva (aproximadamente circular). Como o Perceptron de camada única implementa apenas fronteiras lineares, qualquer reta divide o espaço em dois semiplanos e inevitavelmente produz erros de classificação em uma das regiões.

Figura 4 – Limitação de separabilidade linear em duas classes concêntricas em  $\mathbb{R}^2$ .

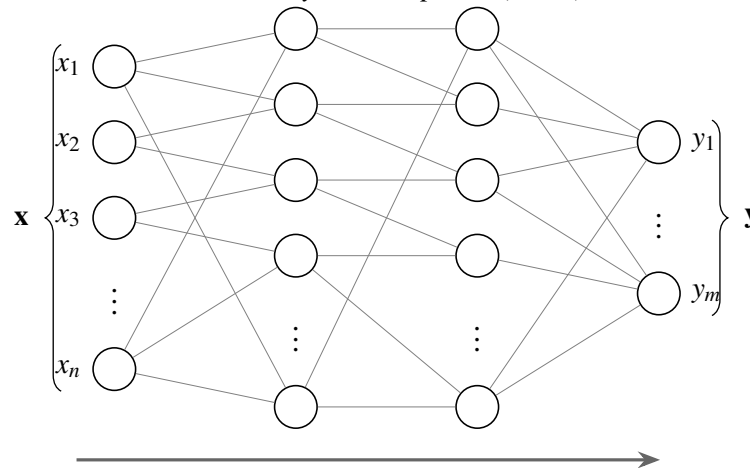


Fonte: Autoria própria (2026).

Para superar essa limitação topológica, consolidaram-se as arquiteturas *Multilayer Per-*

ceptron (MLP). Uma MLP é uma rede neural multicamadas, composta por uma camada de entrada, uma ou mais camadas ocultas (*hidden layers*) e uma camada de saída. A "profundidade" do modelo é definida justamente pelo número dessas camadas ocultas: quanto mais camadas, mais profundo é o modelo, caracterizando o campo do Aprendizado Profundo (*Deep Learning*) (Goodfellow; Bengio; Courville, 2016). A Figura 5 ilustra a arquitetura de uma MLP com duas camadas ocultas, como exemplo representativo.

Figura 5 – Estrutura de uma rede *Multilayer Perceptron* (MLP) com duas camadas ocultas.

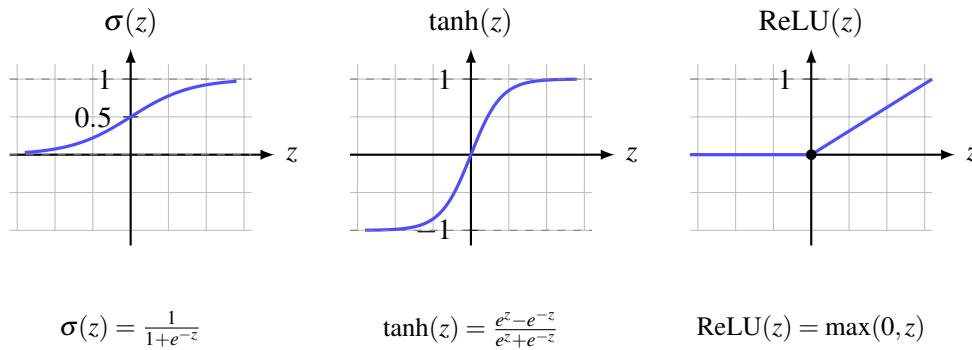


Fonte: Autoria própria (2026).

Do ponto de vista funcional, uma MLP pode ser entendida como uma sequência de camadas que transformam a entrada por meio de combinações lineares seguidas de funções de ativação (Hornik; Stinchcombe; White, 1989). Em cada etapa, a representação produzida pela camada anterior é reprocessada para gerar descrições mais abstratas dos dados. Esse encadeamento permite que a rede modele relações complexas que não seriam capturadas por uma transformação linear simples.

Nesse contexto, a não linearidade desempenha papel central, pois é ela que confere à rede capacidade de aprender fronteiras de decisão mais flexíveis e representações mais expressivas. Entre as funções de ativação mais empregadas estão a Sigmóide, a Tangente Hiperbólica e a ReLU (Goodfellow; Bengio; Courville, 2016). A Figura 6 representa graficamente estas funções.

Figura 6 – Comparativo das principais funções de ativação.



Fonte: Autoria própria (2026).

A rede neural opera baseada em um vasto conjunto de parâmetros, denotado por  $\theta$ , que engloba todos os pesos sinápticos e vieses de suas camadas. O treinamento da rede é um problema de otimização. Primeiro, define-se uma função de custo (ou perda),  $\mathcal{L}(\theta)$ , que atua como uma métrica de erro, mensurando a discrepância entre a previsão da rede e o dado real. O objetivo do aprendizado é encontrar a configuração ideal de parâmetros,  $\theta^*$ , que leve esse erro ao seu valor mínimo global.

Como é inviável testar todas as combinações possíveis de parâmetros, utiliza-se a otimização matemática via Gradiente Descendente. O princípio fundamental é calcular a inclinação da função de custo em relação a cada parâmetro, ou seja, o gradiente  $\nabla_{\theta} \mathcal{L}(\theta)$ . Este vetor aponta na direção de maior aumento da função de custo. Para minimizar o erro, a rede ajusta seus parâmetros na direção oposta ao gradiente.

Para calcular essa inclinação em redes profundas de forma computacionalmente viável, utiliza-se o algoritmo de retropropagação (*backpropagation*) (Rumelhart; Hinton; Williams, 1986). O método aplica a regra da cadeia do cálculo diferencial de forma reversa, propagando o "sinal de erro" da camada de saída de volta para as camadas de entrada para obter o gradiente exato de cada peso.

De posse da direção do erro, a rede atualiza seus pesos iterativamente. Na abordagem clássica do Gradiente Descendente Estocástico (SGD), a atualização segue a regra:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta_t) \quad (2.6)$$

em que  $\eta$  é a taxa de aprendizado (*learning rate*), que define o "tamanho do passo" da otimização. O SGD é robusto, mas como utiliza um passo rígido para todos os parâmetros, pode ficar preso

em vales estreitos (mínimos locais) ou demorar excessivamente para convergir em arquiteturas complexas.

Para solucionar essas ineficiências, arquiteturas modernas empregam o otimizador Adam (*Adaptive Moment Estimation*) (Kingma; Ba, 2017). O Adam age como um otimizador com "inércia" e memória: ele calcula estimativas em movimento tanto da média quanto da variância dos gradientes anteriores. Isso permite que ele ajuste o tamanho do passo para cada parâmetro de forma independente, freando em descidas abruptas e acelerando em platôs:

$$\theta_t = \theta_{t-1} - \eta \cdot \frac{\mathbf{m}_t}{\sqrt{\mathbf{v}_t + \epsilon}} \quad (2.7)$$

Devido a essa adaptação dinâmica, o Adam tende a convergir muito mais rapidamente que o SGD, sendo amplamente adotado como o padrão no treinamento de redes neurais profundas.

### 2.2.1 Redes Neurais Convolucionais (CNN)

Enquanto redes densas (MLP) convertem a entrada em vetor e perdem a estrutura espacial original, as Redes Neurais Convolucionais (CNN) foram concebidas para explorar explicitamente a geometria de dados em grade, como imagens. A consolidação desse paradigma com a LeNet-5 (Lecun *et al.*, 1998) introduziu dois princípios centrais: localidade (cada unidade observa uma vizinhança limitada) e compartilhamento de pesos (o mesmo filtro é aplicado ao longo de toda a imagem), o que reduz parâmetros e melhora a generalização.

A operação básica é a convolução discreta<sup>1</sup>. Intuitivamente, um filtro  $\mathbf{K} \in \mathbb{R}^{m \times n}$  percorre a imagem  $\mathbf{I} \in \mathbb{R}^{H \times W}$  e calcula uma resposta local em cada posição:

$$S(i, j) = (\mathbf{I} * \mathbf{K})(i, j) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} I(i+u, j+v) K(u, v) + b \quad (2.8)$$

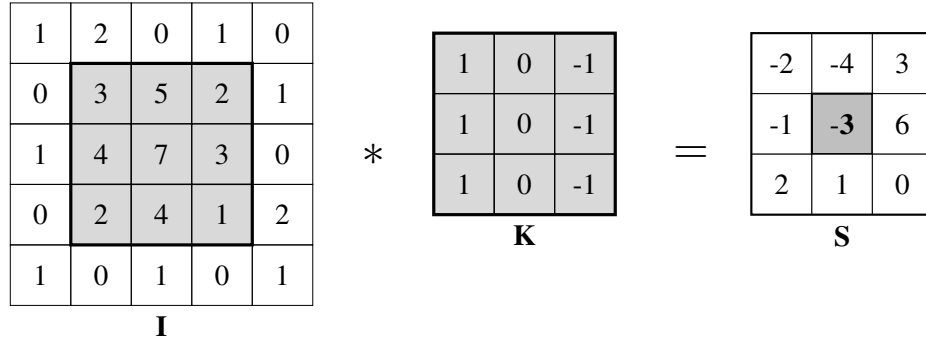
onde  $b \in \mathbb{R}$  é o viés do filtro. Em prática, uma camada usa vários filtros em paralelo, cada um especializado em padrões como bordas, texturas e transições espectrais.

Esse mecanismo torna a convolução *equivariante à translação*: se um padrão desloca na entrada, sua ativação também desloca no mapa de saída. Para reconstrução de imagem, essa propriedade é mais útil que invariância completa, pois preserva informação de onde cada

<sup>1</sup> Em implementações de aprendizado profundo, a operação realizada é tecnicamente uma correlação cruzada, pois não há espelhamento do kernel; mantém-se o termo “convolução” por convenção.

estrutura aparece. A Figura 7 ilustra a operação de convolução com um filtro  $3 \times 3$  aplicado a uma imagem  $5 \times 5$ .

Figura 7 – Diagrama ilustrando a convolução 2D.



Fonte: Autoria própria (2026).

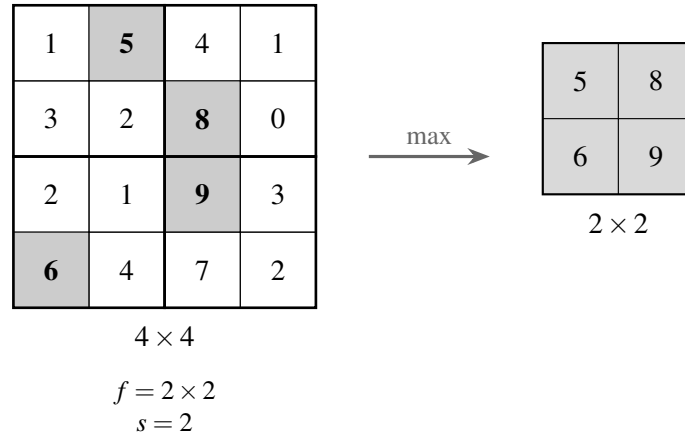
A dimensionalidade espacial do mapa de saída é determinística e depende do tamanho do filtro ( $f$ ), do preenchimento das bordas (*padding*,  $\mathcal{P}$ ) e do passo (*stride*,  $s$ ). Para uma entrada de largura  $W_{in}$ :

$$W_{out} = \left\lfloor \frac{W_{in} - f + 2\mathcal{P}}{s} \right\rfloor + 1 \quad (2.9)$$

onde  $\lfloor \cdot \rfloor$  denota a função piso (*floor*), que garante que o resultado seja um inteiro. A mesma fórmula aplica-se à dimensão de altura  $H_{out}$ , de forma análoga. Note que para  $\mathcal{P} = \lfloor f/2 \rfloor$  e  $s = 1$  (preenchimento equivalente, *same padding*), tem-se  $W_{out} = W_{in}$ , preservando a resolução espacial.

O controle desses hiperparâmetros é decisivo em segmentação e reconstrução, pois define o compromisso entre contexto e resolução. Camadas de *pooling* reduzem dimensionalidade e aumentam campo receptivo efetivo, porém também removem detalhe espacial fino. A Figura 8 ilustra a operação de *Max Pooling*, com uma janela de  $2 \times 2$  e passo  $s = 2$ .

Figura 8 – Diagrama ilustrando a operação de agrupamento máximo (*Max Pooling*) com janela  $2 \times 2$  e passo  $s = 2$ .



Fonte: Autoria própria (2026).

O uso extensivo de subamostragem degrada a fidelidade espacial, aspecto crítico em tarefas pixel a pixel como remoção de nuvens. Embora arquiteturas como VGG (Simonyan; Zisserman, 2014) e ResNet (He *et al.*, 2016) tenham sido fundamentais para classificação, elas comprimem fortemente o mapa espacial e, isoladamente, não são ideais para reconstrução de alta precisão.

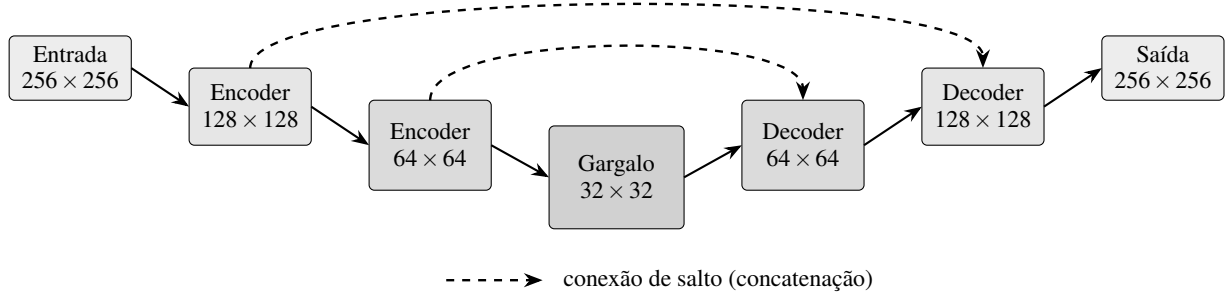
#### 2.2.1.1 Arquiteturas de Codificação e Decodificação (Encoder-Decoder)

Para contornar a perda de resolução espacial inerente às redes convolucionais profundas, consolidaram-se as arquiteturas *Encoder-Decoder*. Nesse cenário, a U-Net (Ronneberger; Fischer; Brox, 2015) estabeleceu-se como o padrão-ouro para tarefas de reconstrução e segmentação de imagens. Sua inovação central reside em separar o processamento em dois caminhos complementares e simétricos:

1. **Codificador (*Encoder*):** Atua como um funil de compressão. Através de convoluções e operações sucessivas de subamostragem (*pooling*), ele reduz progressivamente a dimensão espacial da imagem, ao mesmo tempo em que aumenta a profundidade dos canais. Seu objetivo é extrair "o que" está na cena (o contexto semântico abstrato), sacrificando intencionalmente a informação exata de "onde" as estruturas estão.
2. **Decodificador (*Decoder*):** Funciona como um mecanismo de expansão. Ele recebe a representação latente comprimida e utiliza convoluções transpostas para recuperar gradativamente a resolução original. Seu papel é pegar a compreensão profunda da cena e projetá-la de volta na grade de pixels.

A Figura 9 ilustra esse fluxo. Em termos operacionais, o *encoder* identifica o contexto semântico global, o gargalo (*bottleneck*) consolida essa essência matemática, e o *decoder* inicia a reconstrução espacial.

Figura 9 – Diagrama do fluxo de uma arquitetura Encoder-Decoder no estilo U-Net.



Fonte: Autoria própria.

No entanto, se a arquitetura dependesse exclusivamente desse fluxo profundo, o modelo reconstruiria formas corretas, porém inevitavelmente borradas, devido à perda de precisão espacial durante a compressão. É nesse ponto que atua a principal característica da U-Net: as conexões de salto (*skip connections*). Elas atuam como atalhos de informação, transferindo mapas de características de alta resolução das camadas rasas do *encoder* diretamente para as camadas correspondentes do *decoder*.

Matematicamente, é fundamental ressaltar que essa fusão espacial não ocorre por adição residual, mas por concatenação:

$$\mathbf{x}_{dec}^{(i)} = \text{Concat}\left(\mathcal{U}\left(\mathbf{x}_{dec}^{(i+1)}\right), \mathbf{x}_{enc}^{(i)}\right) \quad (2.10)$$

Nesta formulação,  $\mathcal{U}(\cdot)$  representa a operação de expansão espacial (*upsampling*) aplicada à camada mais profunda do decodificador  $\mathbf{x}_{dec}^{(i+1)}$ . A função  $\text{Concat}(\cdot, \cdot)$  agrupa esses canais expandidos lado a lado com os canais espacialmente ricos do codificador  $\mathbf{x}_{enc}^{(i)}$ .

A vantagem crítica da concatenação sobre a adição matemática é a preservação isolada da informação. O bloco seguinte do *decoder* recebe, de forma independente, o contexto semântico abstrato (vindo das profundezas da rede) e os detalhes finos de textura e contorno (vindos das conexões de salto). Essa arquitetura reduz a difusão de bordas e garante alta fidelidade estrutural, justificando a eficácia da família U-Net em tarefas críticas como o *inpainting* e a restauração de pixels oclusos em sensoriamento remoto (Isola *et al.*, 2018).

### 2.2.2 Mecanismos de Atenção

As Redes Neurais Convolucionais clássicas operam sob o princípio da invariância à translação e aplicam os mesmos pesos a todas as regiões locais de um mapa de características. Essa uniformidade, contudo, torna-se uma desvantagem em tarefas de restauração de imagens severamente degradadas. Em cenários de oclusão por nuvens, o ruído não está distribuído uniformemente; portanto, o modelo não deve investir a mesma capacidade computacional em uma região completamente opaca e em uma região com alta fidelidade de bordas.

Para solucionar essa assimetria, introduzem-se os Mecanismos de Atenção, conforme ilustrado na Figura 10. Eles atuam como filtros dinâmicos e adaptativos que recalibram o fluxo de informação, ensinando a rede a priorizar *o que* (filtragem semântica) e *onde* (filtragem espacial) processar com maior intensidade.

#### 2.2.2.1 Atenção de Canal (*Channel Attention*)

Na extração profunda de características, cada canal atua como um detector especializado (por exemplo, detectores de texturas de vegetação ou detectores de ruído atmosférico). A Atenção de Canal, popularizada pela arquitetura SE-Net (Hu *et al.*, 2019), funciona como um equalizador de sinais. O mecanismo comprime o contexto espacial de cada canal em um único valor numérico através de um mecanismo de *Global Average Pooling* (Lin; Chen; Yan, 2014), e utiliza uma rede auxiliar para estimar um vetor de pesos  $\alpha \in (0, 1)^C$ . A saída do canal  $c$ , denotada por  $\hat{\mathbf{X}}_c$ , é o resultado da recalibração matricial da entrada  $\mathbf{X}_c$ :

$$\hat{\mathbf{X}}_c = \alpha_c \mathbf{X}_c, \quad \alpha = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \quad (2.11)$$

Desta forma, o escalar  $\alpha_c$  pondera toda a matriz de características do canal  $c$ . Canais que carregam assinaturas relevantes para a reconstrução são amplificados, enquanto canais dominados por ruído ou características redundantes são severamente atenuados.

#### 2.2.2.2 Atenção Espacial (*Spatial Attention*)

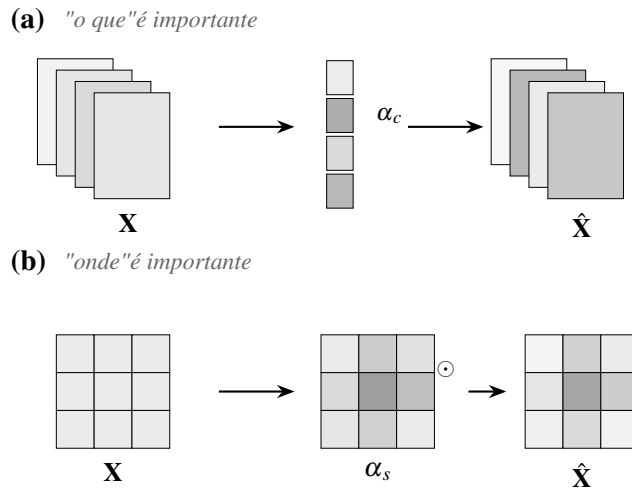
Enquanto a atenção de canal decide *o que* é importante, a Atenção Espacial (como a proposta no módulo CBAM (Woo *et al.*, 2018)) foca em *onde* a informação crítica reside. Este mecanismo colapsa a dimensão dos canais, computando estatísticas ao longo do eixo espectral

(média e valor máximo), e aplica uma convolução com campo receptivo amplo ( $7 \times 7$ ) para gerar um mapa de probabilidade bidimensional  $\alpha_s$ :

$$\alpha_s = \sigma(f^{7 \times 7}([\text{AvgPool}_C(\mathbf{X}); \text{MaxPool}_C(\mathbf{X})])), \quad \hat{\mathbf{X}} = \alpha_s \odot \mathbf{X} \quad (2.12)$$

O produto de Hadamard ( $\odot$ ) denota a multiplicação elemento a elemento entre duas matrizes ou tensores de dimensões idênticas. Nesse contexto, o mapa de atenção bidimensional  $\alpha_s$  (cujos valores variam entre 0 e 1) atua como um filtro espacial sobre o mapa de características original  $\mathbf{X}$ . A operação matemática age multiplicando cada pixel de  $\mathbf{X}$  pelo seu respectivo peso de atenção em  $\alpha_s$ , suprimindo ativações em áreas irrelevantes (onde  $\alpha \approx 0$ ) e destacando estruturas geométricas cruciais no plano da imagem (onde  $\alpha \approx 1$ ).

Figura 10 – Comparativo conceitual da (a) Atenção de Canal e (b) Atenção Espacial.



Fonte: Autoria própria (2026).

### 2.2.2.3 Atenção Guiada por Máscara (*Mask Gated Attention*)

A aplicação mais crítica da modulação de características no contexto de arquiteturas *Encoder-Decoder* ocorre no controle das conexões de salto (*skip connections*). Na U-Net clássica, a concatenação irrestrita de mapas rasos pode transportar artefatos indesejados e assinaturas espectrais da própria nuvem diretamente para a fase final de reconstrução. Como contribuição metodológica para mitigar este problema, propõe-se um mecanismo de atenção guiada.

Em vez de empregar *Attention Gates* condicionados ao decodificador, adotamos um mecanismo de *gating* guiado por máscara, no qual a separação entre regiões ocluídas e não ocluídas é explicitamente modelada. Dado o mapa de características do codificador  $\mathbf{x}^l$  e uma

máscara binária de nuvens  $\mathbf{m}$ , define-se um mapa de ativação  $\alpha(\mathbf{x}^l, \mathbf{m}) \in (0, 1)^{H \times W}$  aprendido por uma projeção convolucional:

$$\mathbf{f}_{\text{fg}} = \alpha \odot \mathbf{x}^l, \quad \mathbf{f}_{\text{bg}} = (1 - \alpha) \odot \mathbf{x}^l \quad (2.13)$$

$$\hat{\mathbf{x}}^l = \phi([\mathbf{f}_{\text{fg}}, \mathbf{f}_{\text{bg}}]) \quad (2.14)$$

onde  $\phi(\cdot)$  representa uma projeção linear que funde as componentes de primeiro plano (*foreground*) e fundo (*background*). Esse mecanismo não atua como um filtro rígido, mas como uma decomposição adaptativa das características, permitindo que a rede module diferencialmente regiões potencialmente contaminadas e regiões estruturais relevantes durante a reconstrução.

### 2.2.3 Funções de Perda e Estratégias de Otimização

A função de perda atua como o mecanismo de orientação do modelo, mensurando a divergência entre a reconstrução gerada e a imagem de referência. Em problemas de reconstrução sob oclusão atmosférica, a escolha desta métrica é decisiva: diferentes formulações matemáticas induzem propriedades visuais distintas, controlando o compromisso entre a suavização do ruído e a preservação de detalhes de alta frequência. Otimizar a rede consiste em seleccionar a penalidade adequada para a topologia do erro esperado.

#### 2.2.3.1 Perdas no Domínio Espacial

As perdas espaciais operam no domínio dos pixels, mensurando o desvio absoluto entre a imagem (com  $N$  pixels) reconstruída  $\hat{I}$  e o alvo  $I$ . As formulações tradicionais divergem em sua sensibilidade a erros extremos: o Erro Quadrático Médio (MSE ou  $\ell_2$ ) e o Erro Absoluto Médio (MAE ou  $\ell_1$ ).

O MSE (*Mean Squared Error*) é definido como:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N \|I_i - \hat{I}_i\|_2^2 \quad (2.15)$$

Embora apresente boas propriedades de convergência, essa amplificação quadrática do erro torna o modelo hipersensível a *outliers*. Em cenários de remoção de nuvens, onde a oclusão

gera artefatos de alto contraste, o MSE tende a forçar a rede a suavizar severamente a imagem para minimizar grandes erros locais, resultando em texturas borradas (Zhao *et al.*, 2018).

Em contrapartida, o MAE (*Mean Absolute Error*) acumula o erro de forma linear:

$$\mathcal{L}_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N \|I_i - \hat{I}_i\|_1 \quad (2.16)$$

O crescimento linear torna o MAE estatisticamente robusto a artefatos e contaminações extremas. Ao não amplificar desproporcionalmente o erro nas regiões oclusas, o modelo consegue focar na preservação de bordas e na coerência estrutural global da superfície terrestre.

### 2.2.3.2 Viés Espectral e Perdas no Domínio da Frequência

A despeito de sua robustez, as perdas espaciais não impedem o fenômeno conhecido como Viés Espectral (*Spectral Bias*) (Rahaman *et al.*, 2019). Redes neurais profundas possuem a propensão inerente a aprender componentes de baixa frequência (estruturas globais de cor e iluminação) nas etapas iniciais do treinamento, negligenciando componentes de alta frequência (texturas finas, arruamentos, bordas hídras).

Para mitigar essa assimetria, a literatura propõe perdas de supervisão espectral, como a *Focal Frequency Loss* (FFL) (Jiang *et al.*, 2021), que modula o foco do treinamento diretamente no espectro:

$$\mathcal{L}_{\text{FFL}} = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} w(u, v) |\mathcal{F}(I)[u, v] - \mathcal{F}(\hat{I})[u, v]|^2 \quad (2.17)$$

Nesta formulação,  $\mathcal{F}(\cdot)$  representa a Transformada de Fourier bidimensional, e  $w(u, v)$  é um peso que enfatiza as frequências onde a divergência é maior. Ao penalizar diretamente as discrepâncias no domínio da frequência, o modelo é forçado a aprender não apenas a estrutura global, mas também os detalhes finos que são cruciais para a percepção visual e a análise quantitativa de imagens de sensoriamento remoto.

É imperativo notar, todavia, que formulações baseadas na Transformada de Fourier possuem suporte global: uma imperfeição local na reconstrução de uma nuvem reverbera por múltiplos coeficientes do espectro. Isso justifica a exploração arquitetural para transformadas de suporte compacto, que isolam o erro simultaneamente no espaço e na frequência.

### 2.2.3.3 Curriculum Learning e Agendamento de Perda

A otimização simultânea da coerência global (baixas frequências) e do refinamento textural (altas frequências) frequentemente desestabiliza a convergência nos estágios iniciais. A estratégia de *Curriculum Learning* (Bengio *et al.*, 2009) resolve esse conflito introduzindo complexidade de forma progressiva.

Em funções de perda compostas, esse processo é modulado por um coeficiente temporal (*loss scheduling*):

$$\mathcal{L}_{\text{total}}(t) = \mathcal{L}_{\text{base}} + \gamma(t) \mathcal{L}_{\text{aux}} \quad (2.18)$$

A componente  $\mathcal{L}_{\text{base}}$  estabiliza a macroestrutura da cena, enquanto  $\mathcal{L}_{\text{aux}}$  foca na penalidade espectral. O termo escalar  $\gamma(t)$  é inicialmente mantido baixo, priorizando a estabilidade estrutural. Conforme a curva de aprendizado atinge platôs,  $\gamma(t)$  é incrementado, forçando a rede a focar no detalhamento microestrutural. Esse agendamento isola as fases de aprendizado, resultando em imagens não apenas radiométricamente precisas, mas perceptualmente fidedignas (Wang; Chen; Zhu, 2021).

## 2.3 Processamento de Sinais: A Transformada Wavelet

Em tarefas de reconstrução de imagens, identificar "o que" é um padrão semântico é tão vital quanto saber "em que escala espacial" ele ocorre. As convoluções convencionais capturam dependências locais com maestria, mas falham em separar explicitamente a estrutura global da imagem (baixas frequências) dos seus contornos e texturas finas (altas frequências). A Transformada Wavelet supre essa lacuna ao fornecer uma decomposição multiescala que preserva a localização espacial das estruturas (Mallat, 1989).

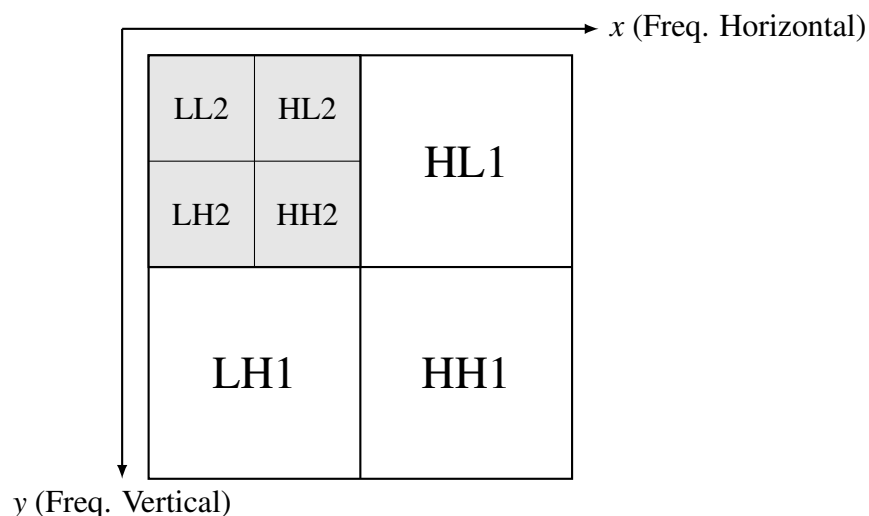
Intuitivamente, a Wavelet atua como um prisma matemático. Ela separa o conteúdo visual suave e redundante das variações rápidas de contraste. Essa propriedade é excepcionalmente valiosa na remoção de nuvens, visto que os detalhes finos da superfície terrestre são as primeiras características obliteradas pelo espalhamento atmosférico.

### 2.3.1 A Transformada Wavelet Discreta (DWT) em Imagens

Na prática computacional, a DWT é implementada através de bancos de filtros bidimensionais ortogonais. Ao aplicar filtros passa-baixa e passa-alta sucessivos, a decomposição de nível 1 divide a imagem original em quatro sub-bandas espaciais de dimensões reduzidas ( $\frac{H}{2} \times \frac{W}{2}$ ), conforme ilustrado na Figura 11:

- **LL (Low-Low):** Contém os coeficientes de aproximação, representando uma versão suavizada do conteúdo global da cena.
- **LH (Low-High):** Isola as componentes de alta frequência vertical, destacando bordas e transições puramente horizontais.
- **HL (High-Low):** Isola as componentes de alta frequência horizontal, evidenciando as bordas verticais.
- **HH (High-High):** Captura as transições diagonais extremas e concentra a maior parte do ruído de alta frequência da imagem.

Figura 11 – Decomposição Wavelet 2D em dois níveis.



Fonte: Autoria própria (2026).

Dentre as diversas famílias matemáticas existentes, este trabalho adota a Wavelet de Haar. Sua formulação geométrica em degrau é computacionalmente leve e exibe uma resposta ótima na localização de descontinuidades abruptas (Kelly, 1996).

Matematicamente, a Transformada Wavelet de Haar em uma dimensão é definida por um filtro passa-baixa **l** e um filtro passa-alta **h**, dados por:

$$\mathbf{l} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{h} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (2.19)$$

Como imagens são sinais bidimensionais, a DWT de Haar 2D é obtida pela aplicação separável desses filtros ao longo das linhas e colunas, o que equivale ao produto externo entre os vetores de filtro uni-dimensionais. Isso resulta em quatro núcleos (*kernels*) de convolução de dimensão  $2 \times 2$ , frequentemente normalizados na implementação de redes neurais com um fator de  $\frac{1}{2}$ :

$$\mathbf{K}_{LL} = \mathbf{l}\mathbf{l}^T = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{K}_{HL} = \mathbf{h}\mathbf{l}^T = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \quad (2.20)$$

$$\mathbf{K}_{LH} = \mathbf{h}\mathbf{l}^T = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad \mathbf{K}_{HH} = \mathbf{h}\mathbf{h}^T = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (2.21)$$

Em termos práticos, ao convoluir a imagem original (com *stride* de 2 para realizar a subamostragem) com estes quatro filtros, a rede isola comportamentos espaciais distintos. O núcleo  $\mathbf{K}_{LL}$  calcula a média local (aproximação), enquanto os núcleos  $\mathbf{K}_{HL}$ ,  $\mathbf{K}_{LH}$  e  $\mathbf{K}_{HH}$  calculam as diferenças direcionais. Esta propriedade torna o modelo de Haar ideal para capturar componentes de alta frequência que são essenciais frequentemente perdidas em operações comuns de *max pooling*.

### 2.3.2 Wavelets em Redes Neurais

A integração da DWT em arquiteturas profundas representa uma evolução crítica em relação às operações clássicas de subamostragem (*downsampling*).

Em redes convolucionais padrão, operadores como o *Max Pooling* reduzem a dimensionalidade descartando ativamente informações (preservando unicamente o valor máximo dentro de uma janela local). Embora essa agressividade seja útil em tarefas de classificação por gerar invariância, ela é altamente deletéria em tarefas de restauração fina (*inpainting*), onde cada pixel descartado representa uma perda de fidelidade geométrica.

A DWT, por outro lado, é um operador matemático estritamente inversível. Ela não descarta a energia do sinal, mas a redistribui de forma organizada em sub-bandas. Ao substituir

o *pooling* clássico pela DWT no estágio de codificação, a rede neural passa a processar canais limpos com a estrutura global (LL) separadamente dos mapas de bordas puras (LH, HL, HH). Essa entrega pré-processada de altas frequências contorna o Viés Espectral do aprendizado profundo, otimizando de maneira robusta a recuperação de texturas e contornos finos na imagem final (Liu *et al.*, 2018).

## 2.4 Estado da Arte em Detecção e Mascaramento

A discriminação precisa entre pixels contaminados por nuvens e a superfície terrestre, denominada mascaramento de nuvens (*cloud masking*), é o pré-requisito fundamental para a exploração operacional de imagens ópticas. Sem uma estimativa confiável da máscara de oclusão, etapas posteriores de análise — como classificação de cobertura do solo, detecção de corpos hídricos e reconstrução de regiões degradadas — ficam sujeitas a erros sistemáticos. A evolução histórica destas estratégias pode ser categorizada taxonomicamente em três gerações: limiares físicos, aprendizado de máquina clássico e aprendizado profundo (Singh; Pal; Biswas, 2025).

### 2.4.1 Limiares Físicos e Machine Learning Tradicional

A primeira geração fundamenta-se nas propriedades radiométricas intrínsecas das nuvens: alta reflectância no espectro visível e baixa temperatura de brilho no canal termal. O expoente desta classe é o algoritmo Fmask (Qiu; Zhu; He, 2019), que combina regras de decisão baseadas em física com operações morfológicas para separar nuvens espessas, nuvens cirrus, sombras e neve em imagens Landsat e Sentinel-2. Derivações como o *s2cloudless* adaptaram esse paradigma para produção em larga escala via redes convolucionais leves treinadas sobre limites definidos fisicamente, representando uma ponte entre a primeira e a segunda geração.

Embora apresentem alta eficiência computacional e interpretabilidade, métodos de limiar encontram desafios naturais em cenários de ambiguidade espectral — gelo permanente, areia branca e nuvens cirrus finas compartilham faixas de reflectância sobrepostas com as nuvens, gerando taxas de falso-positivo que comprometem a qualidade das máscaras em biomas como o semiárido brasileiro (Singh; Pal; Biswas, 2025).

A segunda geração buscou flexibilizar essa rigidez através da classificação supervisionada, em particular *Support Vector Machines* (SVM) e *Random Forests* (RF). Estes modelos permitiram contornos de decisão mais suaves e, quando treinados com atributos espectrais,

texturais e multitemporais, apresentaram desempenho superior aos limiares físicos em regiões de ambiguidade (Anzalone; Pagliaro; Tutone, 2024). Contudo, por dependerem de engenharia manual de atributos (*feature engineering*), sua capacidade de generalização direta entre diferentes sensores e biomas costuma requerer calibrações extensas, limitando a escalabilidade operacional. Experimentos reportados na literatura demonstram que, mesmo com conjuntos de atributos cuidadosamente projetados, modelos RF treinados em Landsat apresentam degradação expressiva de desempenho quando aplicados diretamente a imagens Sentinel-2, evidenciando a dependência de domínio desta geração (Anzalone; Pagliaro; Tutone, 2024).

#### 2.4.2 O Estado da Arte: Deep Learning e Segmentação Semântica

A transição para o Aprendizado Profundo representou um marco metodológico ao substituir a engenharia manual pela extração hierárquica e orientada a dados de características (Dong; Wang; Abbas, 2021). Arquiteturas *Encoder-Decoder*, como a U-Net e suas variantes, consolidaram-se por sua capacidade de modelar o contexto espacial em múltiplas escalas. Essa percepção de campo amplo permite que a rede distinga entre nuvens e superfícies artificiais brilhantes analisando padrões texturais — capacidade ausente em abordagens de limiar —, resultando em detecção mais robusta, especialmente em bordas difusas e sob cobertura parcial.

Revisões recentes sistematizaram o campo, catalogando mais de 80 métodos publicados entre 2016 e 2024 e identificando que arquiteturas baseadas em U-Net respondem por cerca de 40% das propostas de aprendizado profundo para detecção de nuvens, enquanto abordagens híbridas CNN-Transformer cresceram de forma acelerada após 2021 (Singh; Pal; Biswas, 2025). Esse crescimento reflete a capacidade dos Transformers de capturar dependências globais de longo alcance — relevantes para identificar padrões de cobertura extensa — que convoluções de campo receptivo limitado não modelam com eficiência.

Um aspecto crítico desta geração é a necessidade de conjuntos de dados anotados em larga escala. A qualidade das máscaras de treinamento condiciona diretamente o teto de desempenho dos modelos: ruído de anotação, inconsistência entre anotadores e viés geográfico dos dados de treinamento são fontes recorrentes de degradação (Anzalone; Pagliaro; Tutone, 2024). No contexto deste trabalho, essa limitação é particularmente relevante: o subconjunto RICE1 não dispõe de máscaras oficiais, exigindo o procedimento de geração adaptativa descrito na Seção 2.4, cujos artefatos de estimativa propagam-se diretamente para o desempenho dos modelos que dependem desta informação.

## 2.5 Estratégias para Reconstrução e Remoção de Nuvens

A remoção de nuvens (*cloud removal*) transcende a detecção, configurando-se como um problema inverso mal-posto de restauração semântica. O espaço de soluções é subdeterminado: para uma mesma imagem degradada, múltiplas superfícies subjacentes são radiometricamente consistentes com a observação disponível, especialmente sob nuvens espessas onde  $\tau_{\text{cloud}} \rightarrow 0$  e o sinal de superfície é fisicamente anulado, conforme discutido na Seção 2.4. Para preencher o vazio de informação deixado pela oclusão, a literatura contemporânea explora três vertentes principais: a fusão multimodal, os modelos generativos e o processamento no domínio da frequência.

### 2.5.1 Fusão Multimodal: SAR-Óptico

A complementaridade física entre imagens ópticas e dados de Radar de Abertura Sintética (SAR) constitui uma das vertentes mais consolidadas para remoção de nuvens (Chan; Koo, 2008). Como o SAR opera na faixa de micro-ondas — comprimentos de onda entre 1 cm e 1 m —, suas ondas atravessam a cobertura de nuvens sem atenuação significativa, registrando a geometria e as propriedades dielétricas da superfície independentemente das condições atmosféricas. Essa propriedade torna os dados SAR um complemento estrutural natural para a reconstrução do domínio óptico ocluído.

Arquiteturas recentes exploram essa complementaridade com crescente sofisticação. O SCT-CR propõe a fusão sinérgica entre módulos convolucionais — eficientes na extração de texturas locais — e camadas de atenção Transformer — capazes de modelar dependências de longo alcance entre regiões da cena — para harmonizar as representações SAR e óptica (Ma *et al.*, 2024). Abordagens puramente baseadas em Transformers exploram o mecanismo de autoatenção para capturar relações globais entre regiões livres de nuvem e regiões ocluídas, orientando a síntese pelas estruturas coerentes detectadas no canal radar (Han; Wang; Zhang, 2023; Vaswani *et al.*, 2017). Modelos adversariais multimodais combinam Transformers e GANs, em que o gerador utiliza a estrutura SAR como condicionamento espacial e o discriminador penaliza inconsistências locais de textura, obtendo resultados visualmente mais nítidos do que abordagens exclusivamente convolucionais (Li; Liu; Li, 2023).

Apesar do aporte estrutural que os dados SAR oferecem, esta abordagem enfrenta conflitos intrínsecos de domínio que limitam sua generalização:

- O SAR codifica rugosidade superficial e constante dielétrica, enquanto o sensor óptico registra reflectância bidirecional. O mapeamento entre esses domínios é altamente não-linear e dependente do tipo de cobertura: estruturas claramente visíveis no SAR — como arruamentos e telhados metálicos — podem não ter correspondência direta em cor ou textura na imagem óptica (Xu *et al.*, 2022). Modelos treinados em uma região geográfica específica frequentemente não generalizam para biomas com diferentes padrões de retroespalhamento.
- Diferenças na geometria de visada — incidência lateral no SAR contra visada nadir no sensor óptico — introduzem deslocamentos espaciais que exigem rotinas rigorosas de co-registro subpixel. Adicionalmente, o ruído multiplicativo (*speckle*) característico do SAR pode ser inadvertidamente propagado para a imagem óptica reconstruída caso o modelo não incorpore mecanismos explícitos de supressão (Zhu *et al.*, 2021).
- A disponibilidade simultânea de pares SAR-óptico co-registrados e livres de nuvens para treinamento é geograficamente restrita, introduzindo viés de cobertura nos modelos treinados nessa modalidade.

Essa dependência de dados pareados multimodais representa uma barreira operacional relevante: em aplicações como o monitoramento de reservatórios no semiárido brasileiro — onde imagens Sentinel-1 (SAR) e Sentinel-2 (óptico) devem ser adquiridas com cobertura temporal compatível — a probabilidade de se obter pares co-registrados de alta qualidade decresce com a resolução temporal exigida.

### 2.5.2 Modelos Generativos: Difusão e GANs

Para atenuar a dependência de dados pareados de diferentes sensores, os modelos generativos focam em aprender a distribuição de probabilidade das imagens limpas condicionada à observação degradada. Esta abordagem reformula a remoção de nuvens como um problema de amostragem condicional: em vez de regredir diretamente para um único valor de pixel reconstruído, o modelo aprende a distribuição de superfícies plausíveis e gera realizações a partir dela.

Recentemente, os Modelos de Difusão (Sohl-Dickstein *et al.*, 2015) — que aprendem a reverter iterativamente um processo de perturbação gaussiana — têm demonstrado capacidade notável de síntese de texturas realistas e espacialmente coerentes. Formulações como os *Denoi-sing Diffusion Probabilistic Models* (DDPM) estendem-se ao domínio do sensoriamento remoto

por meio de condicionamentos espaciais explícitos. O DiffCR agiliza o processo de reversão via destilação do número de passos de difusão, mantendo qualidade visual competitiva com fração do custo de inferência dos modelos DDPM padrão (Zou *et al.*, 2024). O IDF-CR decompõe a reconstrução em sub-regiões e aplica atenção iterativa para refinar a precisão geográfica de cada segmento, sendo particularmente eficaz em cenas com cobertura mista de nuvens finas e espessas (Wang *et al.*, 2024). Revisões recentes indicam que modelos de difusão superam consistentemente baselines convolucionais em métricas perceptuais como FID e LPIPS, embora a vantagem seja menos pronunciada em métricas de fidelidade pixel-a-pixel como PSNR e SSIM (Ning *et al.*, 2025; Edirisinghe; Schoen-Phelan; Hensman, 2026).

As Redes Generativas Adversariais (GANs) (Liu *et al.*, 2022) permanecem relevantes por oferecer alta nitidez estrutural com custo de inferência determinístico — uma vantagem operacional sobre modelos iterativos. A formulação adversarial induz o gerador a produzir imagens que enganam o discriminador, favorecendo texturas de alta frequência que perdas de reconstrução padrão tendem a suavizar. Contudo, o jogo adversarial introduz instabilidade no treinamento: colapso de modo (*mode collapse*), divergência do discriminador e sensibilidade extrema à escolha de hiperparâmetros são limitações documentadas que dificultam a reprodutibilidade e a adaptação a novos conjuntos de dados (Edirisinghe; Schoen-Phelan; Hensman, 2026).

Do ponto de vista prático, modelos generativos impõem desafios operacionais relevantes para aplicações de sensoriamento remoto em produção. Os modelos de difusão requerem dezenas a centenas de passos de reversão por imagem, tornando o processamento de grandes acervos — como os produzidos pelo programa Copernicus — computacionalmente custoso sem infraestrutura de GPU dedicada. As GANs, embora mais rápidas em inferência, demandam controle rigoroso do processo de treinamento para garantir estabilidade. Esses fatores motivam a busca por arquiteturas determinísticas de menor custo que preservem detalhes de alta frequência sem o custo iterativo ou a instabilidade adversarial.

### 2.5.3 Aprendizado no Domínio da Frequência

O processamento no domínio da frequência surge como uma alternativa pragmática que busca equilibrar custo computacional, estabilidade de treinamento e preservação de detalhes. Esta vertente aborda diretamente o fenômeno do Viés Espectral (*Spectral Bias*) (Rahaman *et al.*, 2019): redes neurais profundas exibem predisposição inerente a aprender componentes de

baixa frequência nas fases iniciais do treinamento, relegando a recuperação de texturas finas — bordas hídricas, arruamentos, padrões agrícolas — para etapas avançadas ou, em casos extremos, negligenciando-as completamente. Em imagens de sensoriamento remoto, essa assimetria de aprendizado é particularmente prejudicial: são exatamente os componentes de alta frequência que codificam as feições de interesse para análises de cobertura do solo e recursos hídricos.

A *Focal Frequency Loss* (FFL) (Jiang *et al.*, 2021) introduziu um mecanismo de atenção no domínio de Fourier que pondera dinamicamente as frequências pelo erro residual: componentes nos quais o modelo ainda diverge da referência recebem penalização amplificada, forçando o aprendizado a focar progressivamente nas bandas subespecificadas. Essa formulação demonstrou ganhos consistentes em tarefas de reconstrução e síntese de imagens (Jiang *et al.*, 2021), mas herda uma limitação estrutural das transformadas de Fourier: seu suporte global implica que uma imperfeição localizada — como um artefato residual em região de borda de nuvem — distribui seu erro por todo o espectro de frequências, dificultando a localização e correção precisa.

A análise por Wavelets supera essa limitação ao fornecer uma decomposição que é simultaneamente localizada no espaço e na frequência. O FDT-Net adota uma estratégia curricular explícita: nas épocas iniciais, a rede é guiada exclusivamente pela perda espacial para estabilizar a macroestrutura da reconstrução; à medida que a convergência avança, uma perda wavelet é progressivamente ativada, direcionando o treinamento para a preservação de bordas e texturas nas sub-bandas de alta frequência, como LH, HL e HH da DWT (Jiang *et al.*, 2024). Esse agendamento evita o conflito entre objetivos de baixa e alta frequência nas fases iniciais, quando os gradientes de reconstrução global ainda dominam a otimização.

Além da supervisão em frequência, a DWT tem sido explorada como operador de subamostragem dentro das próprias arquiteturas. Trabalhos em redes de restauração demonstram que substituir operações de *Max Pooling* por decomposição wavelet permite reduzir a resolução espacial preservando a energia das sub-bandas de detalhe — informação que o pooling convencional descarta irreversivelmente (Liu *et al.*, 2018). No domínio da detecção de nuvens, a família Haar foi empregada para extração de atributos texturais multirresolução, com vantagem sobre filtros convolucionais fixos por não introduzir parâmetros adicionais no processo de extração (Surya; Rahiman, 2017).

A DWT possui ainda a propriedade matemática de inversibilidade exata: ao contrário do *Max Pooling*, a Transformada Wavelet Discreta Inversa (IDWT) permite reconstruir o sinal original sem perda de informação, desde que todas as sub-bandas sejam preservadas. Essa carac-

terística é explorada arquiteturalmente na WAMG-UNet: a substituição de *pooling* e *upsampling* convencionais por pares DWT/IDWT cria um fluxo encoder-decoder onde a energia de alta frequência nunca é descartada, apenas redistribuída entre sub-bandas processadas separadamente pelas camadas convolucionais subsequentes.

Diante deste panorama metodológico, observa-se que as abordagens atuais impõem compromissos distintos: a fusão SAR-óptico reduz a incerteza geométrica ao custo de dependência de dados multimodais pareados; os modelos generativos produzem texturas perceptualmente realistas ao custo de instabilidade de treinamento ou inferência iterativa custosa; o domínio da frequência oferece estabilidade e custo controlado ao custo de limitar-se ao contexto de uma única aquisição óptica.

Motivado por esse balanço, este trabalho apoia-se no processamento no domínio da frequência. A inovação estrutural proposta reside na integração da DWT (*Wavelet Pooling*) com mecanismos de Atenção Espacial Guiada por Máscara. Diferentemente de usos prévios de wavelets (Liu *et al.*, 2018), a arquitetura WAMG-UNet, descrita na Seção 3.2, não apenas extrai sub-bandas de alta frequência, mas utiliza a máscara de nuvem como sinal de controle para modular o fluxo dessas sub-bandas nas conexões de salto — mitigando a propagação acidental de texturas da nuvem para a imagem restaurada. Esta escolha busca resultados de alta fidelidade estrutural sem o custo iterativo dos modelos de difusão e sem a dependência de dados SAR.

### 3 METODOLOGIA

Este capítulo descreve os procedimentos metodológicos adotados para o desenvolvimento e a avaliação da arquitetura WAMG-UNet. A metodologia foi organizada em cinco etapas principais: preparação do conjunto de dados, formulação da tarefa, descrição da arquitetura proposta, definição dos modelos comparativos e configuração do treinamento e da avaliação experimental.

A proposta central consiste em avaliar se a combinação entre orientação explícita por máscara, processamento multiescala baseado em wavelets e uma perda auxiliar no domínio da frequência contribui para a reconstrução de imagens ópticas de sensoriamento remoto degradadas por nuvens. Para isso, o modelo proposto foi comparado com modelos base (*baselines*) baseados em U-Net e com uma variante ablativa da própria WAMG-UNet.

#### 3.1 Conjunto de Dados e Formulação da Tarefa

##### 3.1.1 Conjunto de Dados RICE

A validação experimental foi conduzida com o conjunto de dados *Remote Sensing Image Cloud Removal* (RICE), proposto por (Lin *et al.*, 2019). O RICE é amplamente utilizado em estudos de remoção de nuvens por aprendizado profundo e é composto por dois subconjuntos complementares, denominados RICE1 e RICE2.

O RICE1 contém pares de imagens com nuvens finas ou semitransparentes e suas respectivas imagens de referência sem nuvens. Nesse cenário, a informação da superfície ainda está parcialmente presente na imagem degradada, embora atenuada por espalhamento atmosférico, redução de contraste e alterações radiométricas. Assim, a tarefa aproxima-se de um problema de restauração, no qual o modelo deve recuperar contraste, textura e fidelidade radiométrica.

O RICE2, por sua vez, apresenta nuvens mais espessas e regiões de oclusão mais severa, além de máscaras binárias associadas às regiões contaminadas. Nesse caso, parte da informação da superfície pode estar fortemente degradada ou ausente, aproximando o problema de uma tarefa de reconstrução ou *inpainting*. A utilização conjunta dos dois subconjuntos permite avaliar os modelos em regimes distintos de degradação, desde nuvens finas até cenários de oclusão mais intensa.

Neste trabalho, cada amostra supervisionada é representada por uma tripla:

$$(\mathbf{I}, \mathbf{M}, \mathbf{Y}), \quad (3.1)$$

em que  $\mathbf{I}$  representa a imagem de entrada com nuvens,  $\mathbf{M}$  representa a máscara binária de nuvens e  $\mathbf{Y}$  representa a imagem de referência livre de nuvens. O objetivo do modelo é aprender uma função  $f_\theta$  capaz de estimar uma imagem reconstruída  $\hat{\mathbf{Y}}$ :

$$\hat{\mathbf{Y}} = f_\theta(\mathbf{I}, \mathbf{M}), \quad (3.2)$$

minimizando a discrepância entre  $\hat{\mathbf{Y}}$  e  $\mathbf{Y}$ .

### 3.1.2 Particionamento dos Dados

Para garantir consistência experimental e reduzir o risco de vazamento de dados, o particionamento foi realizado de forma controlada, utilizando semente aleatória fixa  $s = 42$ . A separação entre treinamento e validação foi mantida constante para todos os modelos avaliados, assegurando que as comparações fossem realizadas sob o mesmo protocolo.

As imagens foram processadas em resolução de  $256 \times 256$  pixels e normalizadas para o intervalo  $[0, 1]$ . Essa normalização foi adotada para compatibilizar as entradas com as funções de perda, métricas quantitativas e arquiteturas avaliadas. O mesmo conjunto de validação foi utilizado para todos os modelos, permitindo comparação direta entre baselines, ablações e a arquitetura completa.

### 3.1.3 Geração de Máscaras no RICE1

Como o subconjunto RICE1 não disponibiliza máscaras oficiais de nuvem, foi utilizado um procedimento de geração adaptativa de máscaras a partir da diferença entre a imagem degradada e sua correspondente imagem de referência. Para estimar as regiões contaminadas por nuvens no RICE1, calculou-se inicialmente a diferença absoluta entre a imagem com nuvens e a imagem de referência em cada canal RGB. Em seguida, essas diferenças foram combinadas por uma soma ponderada, atribuindo maior peso ao canal azul, seguido dos canais verde e vermelho. Essa escolha foi adotada empiricamente, considerando que nuvens finas e efeitos de

espalhamento atmosférico tendem a produzir alterações mais perceptíveis nas faixas de menor comprimento de onda do espectro visível.

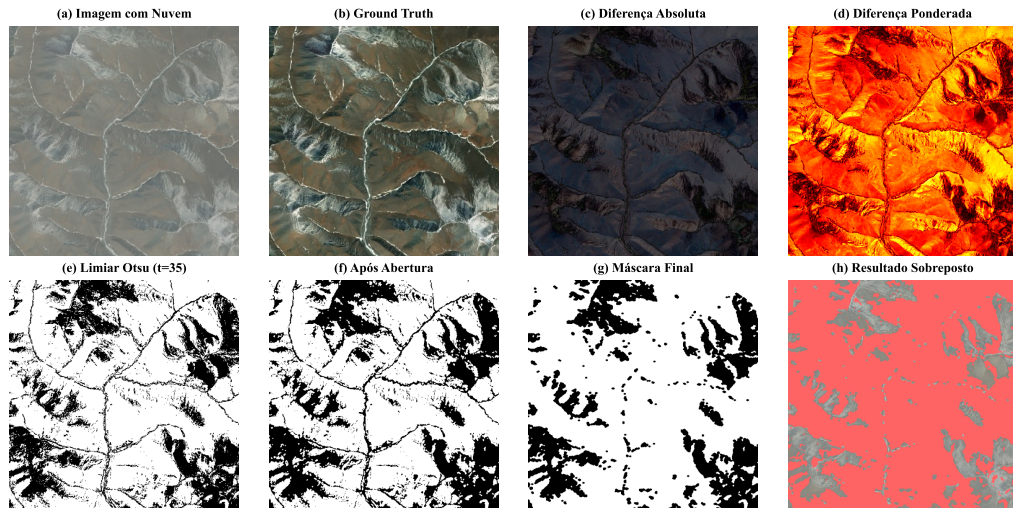
Assim, a diferença espectral ponderada foi definida por:

$$\Delta_w = 0,40 |I_{\text{cloud}}^B - I_{\text{clean}}^B| + 0,35 |I_{\text{cloud}}^G - I_{\text{clean}}^G| + 0,25 |I_{\text{cloud}}^R - I_{\text{clean}}^R|. \quad (3.3)$$

Nessa formulação,  $I_{\text{cloud}}^B$ ,  $I_{\text{cloud}}^G$  e  $I_{\text{cloud}}^R$  representam os canais azul, verde e vermelho da imagem com nuvens, enquanto  $I_{\text{clean}}^B$ ,  $I_{\text{clean}}^G$  e  $I_{\text{clean}}^R$  representam os canais correspondentes da imagem de referência. Os pesos 0,40, 0,35 e 0,25 foram definidos de forma empírica a partir da inspeção visual das máscaras geradas, buscando realçar regiões com maior probabilidade de contaminação por nuvens sem amplificar excessivamente diferenças radiométricas não associadas à oclusão.

Após o cálculo de  $\Delta_w$ , a imagem de diferença foi submetida à limiarização automática pelo método de Otsu (Otsu, 1979). Em seguida, operações morfológicas foram aplicadas para remover ruídos isolados e preencher pequenas discontinuidades, produzindo uma máscara binária mais consistente para uso no treinamento.

Figura 12 – Pipeline de geração adaptativa de máscaras para o subconjunto RICE1.



Fonte: Autoria própria (2026).

No RICE2, foram utilizadas as máscaras disponibilizadas pelo próprio conjunto de dados. Dessa forma, todos os modelos que utilizam informação de máscara recebem uma representação binária das regiões contaminadas em ambos os subconjuntos.

### 3.1.4 Filtragem e Normalização

Após a preparação das imagens e máscaras, as amostras foram submetidas a um processo de filtragem de qualidade. Foram descartados exemplos com baixa variância, regiões excessivamente homogêneas, baixa presença de nuvens ou proporção elevada de pixels inválidos. Essa etapa reduz a presença de amostras pouco informativas no treinamento e favorece o aprendizado de padrões relevantes de degradação e reconstrução.

Todas as imagens foram convertidas para ponto flutuante e normalizadas para o intervalo  $[0, 1]$ . Essa normalização torna as métricas de erro diretamente comparáveis e estabiliza o processo de otimização.

## 3.2 Arquitetura Proposta: WAMG-UNet

A arquitetura proposta, denominada *Wavelet-Attention Mask-Guided U-Net* (WAMG-UNet), é baseada na estrutura *encoder-decoder* da U-Net clássica (Ronneberger; Fischer; Brox, 2015). A U-Net é uma escolha natural para tarefas de reconstrução de imagens por combinar extração hierárquica de características com conexões de salto entre codificador e decodificador, como discutido na Seção 2.2.1.1.

Apesar de sua eficácia, a U-Net convencional apresenta limitações relevantes para remoção de nuvens. Primeiro, operações tradicionais de subamostragem, como *max pooling*, podem descartar detalhes espaciais e componentes de alta frequência. Segundo, as conexões de salto podem transportar para o decodificador informações contaminadas por nuvens, reintroduzindo artefatos atmosféricos na reconstrução final.

A WAMG-UNet busca mitigar esses problemas por meio de três modificações principais: uso explícito da máscara de nuvem, subamostragem baseada na Transformada Wavelet Discreta e conexões de salto guiadas por máscara.

### 3.2.1 Entrada Guiada por Máscara

A WAMG-UNet recebe como entrada a imagem RGB degradada  $\mathbf{I}$  e a máscara de nuvem  $\mathbf{M}$ . A imagem fornece a informação radiométrica e espacial disponível, enquanto a máscara informa explicitamente quais regiões apresentam contaminação por nuvens.

Diferentemente da U-Net RGB, que precisa inferir implicitamente a localização das regiões degradadas, a WAMG-UNet utiliza a máscara como sinal auxiliar durante o processamento.

Essa informação permite diferenciar regiões confiáveis de regiões que exigem reconstrução, reduzindo a ambiguidade da tarefa.

### 3.2.2 Subamostragem via Wavelet Pooling

Na U-Net tradicional, a redução de resolução espacial é usualmente realizada por operações de *pooling*. Embora eficientes, essas operações tendem a descartar parte da informação local, o que pode prejudicar a reconstrução de bordas, texturas e transições espaciais.

Para preservar informação multiescala, a WAMG-UNet utiliza subamostragem baseada na Transformada Wavelet Discreta (DWT) de Haar. Em cada estágio do codificador, um mapa de características  $\mathbf{F}_l$  é decomposto em quatro sub-bandas:

$$[\mathbf{F}_{LL}, \mathbf{F}_{LH}, \mathbf{F}_{HL}, \mathbf{F}_{HH}] = \mathcal{W}(\mathbf{F}_l), \quad (3.4)$$

em que  $\mathcal{W}(\cdot)$  representa a decomposição wavelet. A sub-banda LL contém a aproximação de baixa frequência, enquanto LH, HL e HH representam detalhes horizontais, verticais e diagonais.

Essa decomposição permite que o modelo reduza a resolução espacial preservando componentes de alta frequência que seriam parcialmente perdidos em operações de pooling convencionais. As sub-bandas são então processadas por camadas convolucionais, permitindo que a rede aprenda representações multiescala adequadas à reconstrução.

### 3.2.3 Reconstrução via Transformada Wavelet Inversa

No decodificador, a recuperação da resolução espacial é realizada por operações associadas à Transformada Wavelet Discreta Inversa (IDWT). A IDWT recompõe a representação espacial a partir das sub-bandas de baixa e alta frequência, favorecendo a preservação de estruturas locais durante o processo de reconstrução.

A operação de reconstrução pode ser representada por:

$$\mathbf{F}_l^{rec} = \mathcal{W}^{-1}(\mathbf{F}_{LL}, \mathbf{F}_{LH}, \mathbf{F}_{HL}, \mathbf{F}_{HH}), \quad (3.5)$$

em que  $\mathcal{W}^{-1}(\cdot)$  representa a transformada wavelet inversa.

O uso combinado de DWT no codificador e IDWT no decodificador permite estruturar a rede como um modelo multirresolução, preservando informações de baixa e alta frequência ao longo do fluxo *encoder-decoder*.

### 3.2.4 Conexões de Salto Guiadas por Máscara

As conexões de salto da U-Net transferem mapas de características de alta resolução do codificador para o decodificador. Em tarefas de reconstrução, esse mecanismo é útil para preservar detalhes espaciais; entretanto, em imagens com nuvens, essas conexões também podem transportar artefatos atmosféricos.

Para reduzir esse problema, a WAMG-UNet introduz conexões de salto guiadas pela máscara de nuvem. Seja  $\mathbf{x}_{enc}$  o mapa de características produzido pelo codificador em determinado nível, e seja  $\mathbf{m}$  a máscara redimensionada para a mesma escala espacial. O mecanismo de modulação aprende um mapa de controle  $\mathbf{g}$  (equivalente ao mapa de atenção espacial  $\alpha$  detalhado na Seção 2.2.2.3):

$$\mathbf{g} = \sigma(\mathbf{W}_g * \text{Concat}(\mathbf{x}_{enc}, \mathbf{m}) + \mathbf{b}_g), \quad (3.6)$$

em que  $\sigma(\cdot)$  é a função sigmoide,  $*$  representa convolução,  $\mathbf{W}_g$  são pesos aprendidos de uma camada convolucional  $1 \times 1$  e  $\mathbf{b}_g$  é o termo de viés.

A saída da conexão de salto é então projetada e normalizada:

$$\mathbf{x}_{skip} = \text{BatchNorm}(\mathbf{W}_p * \text{Concat}(\mathbf{x}_{enc} \odot \mathbf{g}, \mathbf{x}_{enc} \odot (1 - \mathbf{g}))), \quad (3.7)$$

em que  $\text{BatchNorm}(\cdot)$  denota a operação de normalização de lote (*Batch Normalization*), responsável por estabilizar a distribuição das características projetadas;  $\odot$  representa o produto elemento a elemento; e  $\mathbf{W}_p$  representa os pesos da convolução de projeção ( $1 \times 1$ ). Essa formulação permite que a rede trate de forma distinta regiões consideradas confiáveis (fundo) e regiões potencialmente degradadas pela presença de nuvens (primeiro plano), mitigando a propagação de ruídos na etapa de decodificação.

### 3.3 Modelos Comparados e Estudo de Ablação

Para avaliar a contribuição dos componentes propostos, foram treinadas quatro configurações principais. A comparação entre essas configurações permite isolar progressivamente o efeito da máscara de nuvem, da arquitetura WAMG e da perda wavelet com Curriculum Learning. A Tabela 1 resume os modelos avaliados e os componentes utilizados em cada configuração.

Tabela 1 – Modelos avaliados e componentes utilizados em cada configuração.

| Modelo                 | Entrada       | Wavelet Pooling | Mecanismo de máscara | Perda wavelet | Curriculum |
|------------------------|---------------|-----------------|----------------------|---------------|------------|
| U-Net RGB              | RGB           | Não             | Não                  | Não           | Não        |
| U-Net RGB+Mask         | RGB + máscara | Não             | Não                  | Não           | Não        |
| WAMG sem perda wavelet | RGB + máscara | Sim             | Sim                  | Não           | Não        |
| WAMG-UNet Completa     | RGB + máscara | Sim             | Sim                  | Sim           | Sim        |

A U-Net RGB representa o baseline mais simples, pois utiliza apenas a imagem degradada como entrada. A U-Net RGB+Mask mantém a estrutura geral da U-Net, mas adiciona a máscara de nuvem como canal auxiliar, permitindo avaliar o impacto direto da informação espacial de oclusão.

A variante WAMG sem perda wavelet preserva os principais componentes arquiteturais da WAMG-UNet, incluindo o processamento baseado em wavelets e as conexões de salto guiadas por máscara, mas remove a penalização no domínio wavelet durante o treinamento. Essa configuração permite avaliar a contribuição da arquitetura independentemente da perda auxiliar em frequência.

A WAMG-UNet Completa corresponde à configuração completa proposta neste trabalho, combinando arquitetura WAMG, perda wavelet e modulação curricular do peso associado à componente de frequência.

É importante destacar que este estudo não isola separadamente o efeito da perda wavelet e do Curriculum Learning, pois não foi treinada uma variante WAMG com perda wavelet fixa sem modulação curricular. Assim, a comparação entre WAMG sem perda wavelet e WAMG-UNet Completa avalia o efeito conjunto da inclusão da perda wavelet e de sua modulação durante o treinamento.

### 3.4 Estratégia de Otimização e Treinamento

#### 3.4.1 Função de Perda

A configuração completa da WAMG-UNet foi treinada com uma função de perda composta por um termo espacial e um termo no domínio wavelet:

$$\mathcal{L}_{total} = \mathcal{L}_{pixel} + \lambda_w(t)\mathcal{L}_{wavelet}. \quad (3.8)$$

O termo  $\mathcal{L}_{pixel}$  corresponde à perda L1 entre a imagem reconstruída  $\hat{\mathbf{Y}}$  e a imagem de referência  $\mathbf{Y}$ :

$$\mathcal{L}_{pixel} = \frac{1}{N} \sum_{i=1}^N |\hat{\mathbf{Y}}_i - \mathbf{Y}_i|. \quad (3.9)$$

O termo  $\mathcal{L}_{wavelet}$  aplica uma penalização equivalente no domínio da Transformada Wavelet Discreta, comparando as sub-bandas da reconstrução e da referência:

$$\mathcal{L}_{wavelet} = \frac{1}{N} \sum_{i=1}^N |\mathcal{W}(\hat{\mathbf{Y}}_i) - \mathcal{W}(\mathbf{Y}_i)|, \quad (3.10)$$

em que  $\mathcal{W}(\cdot)$  representa a decomposição wavelet de Haar.

Nos modelos U-Net RGB, U-Net RGB+Mask e WAMG sem perda wavelet, a função de perda utilizada no treinamento corresponde apenas ao termo espacial  $\mathcal{L}_{pixel}$ . A perda wavelet foi empregada exclusivamente na configuração WAMG-UNet Completa.

#### 3.4.2 Curriculum Learning aplicado à perda wavelet

Na WAMG-UNet Completa, o peso  $\lambda_w(t)$  da componente wavelet foi modulado por uma estratégia de *Curriculum Learning*. A motivação dessa estratégia é permitir que o modelo priorize inicialmente a reconstrução global da cena, guiado pela perda espacial, e aumente posteriormente a importância da penalização em frequência, favorecendo o refinamento de detalhes locais.

O peso da perda wavelet foi inicializado em:

$$\lambda_w(0) = 0,5. \quad (3.11)$$

Durante o treinamento,  $\lambda_w(t)$  foi ajustado de acordo com o comportamento da métrica de validação, aumentando em 0.25 quando o processo de otimização indicava estabilização.

Dessa forma, a componente wavelet atuou como um termo progressivo de refinamento, sem dominar a otimização principal. Esse ajuste foi aplicado exclusivamente à WAMG-UNet Completa, pois os demais modelos não possuem termo wavelet ativo na função de perda.

### 3.4.3 Otimização

Todos os modelos foram treinados com o otimizador Adam, utilizando semente aleatória fixa igual a 42. A taxa de aprendizado foi controlada por um scheduler do tipo *ReduceLROnPlateau*, que reduz o valor da taxa quando a métrica de validação deixa de apresentar melhora significativa.

O treinamento foi conduzido por até 70 épocas. Para cada configuração, o melhor checkpoint foi selecionado com base no maior valor de PSNR obtido no conjunto de validação. Essa escolha prioriza a fidelidade da reconstrução e evita selecionar um checkpoint final que eventualmente apresente desempenho inferior ao melhor ponto de generalização.

### 3.4.4 Média Móvel Exponencial dos Pesos

Nas configurações WAMG, utilizou-se Média Móvel Exponencial dos pesos (*Exponential Moving Average – EMA*) para aumentar a estabilidade da validação e reduzir oscilações associadas ao processo de otimização. A EMA mantém uma versão suavizada dos parâmetros da rede ao longo do treinamento:

$$\bar{\theta}_t = \mu \bar{\theta}_{t-1} + (1 - \mu) \theta_t, \quad (3.12)$$

em que  $\theta_t$  representa os pesos do modelo no instante  $t$ ,  $\bar{\theta}_t$  representa os pesos suavizados e  $\mu$  é o fator de decaimento. Neste trabalho, adotou-se  $\mu = 0,999$ .

Durante a validação, os pesos suavizados pela EMA foram utilizados para reduzir a sensibilidade a variações pontuais do treinamento, favorecendo maior estabilidade nas métricas de desempenho (Morales-Brotons; Vogels; Hendriks, 2024).

### 3.4.5 Regularização

Para mitigar *overfitting*, empregou-se *Dropout* nas camadas intermediárias do codificador e no gargalo da rede. Durante o treinamento, ativações são desativadas aleatoriamente com probabilidade  $p = 0,1$ , forçando o modelo a aprender representações menos dependentes de unidades específicas. Durante a inferência, o *Dropout* é desativado, conforme a formulação original do conceito (Srivastava *et al.*, 2014).

## 3.5 Métricas de Avaliação

A avaliação dos modelos foi conduzida por meio de métricas quantitativas de reconstrução, análise qualitativa das imagens geradas e indicadores de custo computacional. Para a avaliação quantitativa da qualidade das reconstruções, foram utilizadas três métricas complementares: PSNR (*Peak Signal-to-Noise Ratio*), SSIM (*Structural Similarity Index Measure*) e MAE (*Mean Absolute Error*).

A escolha dessas métricas busca contemplar diferentes aspectos da qualidade da reconstrução. O PSNR mede a fidelidade global da imagem em termos de erro quadrático, sendo sensível a grandes discrepâncias pixel a pixel. O SSIM avalia a preservação estrutural da imagem, considerando características locais de luminância, contraste e estrutura. O MAE, por sua vez, mede diretamente o erro absoluto médio, oferecendo uma interpretação mais simples da magnitude média das diferenças entre a reconstrução e a imagem de referência.

Além das métricas de qualidade visual, também foram mensurados indicadores de eficiência computacional, incluindo número de parâmetros treináveis, latência média de inferência, throughput em imagens por segundo e memória máxima alocada na GPU. Esses indicadores permitem avaliar o compromisso entre desempenho de reconstrução e custo prático de execução.

### 3.5.1 PSNR

O *Peak Signal-to-Noise Ratio* (PSNR) é uma métrica amplamente utilizada em tarefas de restauração, compressão e reconstrução de imagens. Ele mede a relação entre o valor máximo possível do sinal e o erro quadrático médio entre a imagem reconstruída e a imagem de referência. Em termos intuitivos, quanto menor o erro entre a reconstrução e a imagem alvo, maior será o valor de PSNR.

Inicialmente, calcula-se o erro quadrático médio, ou *Mean Squared Error* (MSE), entre a

imagem reconstruída  $\hat{\mathbf{Y}}$  e a imagem de referência  $\mathbf{Y}$ :

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{Y}}_i - \mathbf{Y}_i)^2, \quad (3.13)$$

em que  $N$  representa o número total de pixels avaliados, considerando todos os canais da imagem. Como as imagens utilizadas neste trabalho foram normalizadas para o intervalo  $[0, 1]$ , o valor máximo possível do sinal é igual a 1. Dessa forma, o PSNR é definido como:

$$\text{PSNR} = 10 \log_{10} \left( \frac{1}{\text{MSE}} \right). \quad (3.14)$$

O PSNR é expresso em decibéis (dB), e valores maiores indicam maior fidelidade da reconstrução em relação à imagem de referência. No contexto deste trabalho, essa métrica é útil para comparar de forma objetiva o erro global produzido por diferentes arquiteturas.

Entretanto, o PSNR apresenta limitações. Por ser baseado no erro quadrático médio, ele pode penalizar fortemente diferenças pontuais de intensidade, mas não necessariamente reflete a qualidade perceptual ou a preservação de estruturas visuais. Assim, duas imagens com valores semelhantes de PSNR podem apresentar diferenças perceptuais relevantes. Por esse motivo, o PSNR foi utilizado em conjunto com SSIM e MAE.

### 3.5.2 SSIM

O *Structural Similarity Index Measure* (SSIM) é uma métrica voltada à avaliação da similaridade estrutural entre duas imagens. Diferentemente de métricas baseadas apenas em erro pixel a pixel, como MSE e MAE, o SSIM considera características locais relacionadas à luminância, ao contraste e à estrutura. Essa abordagem torna a métrica mais próxima da percepção visual humana em diversas tarefas de reconstrução de imagens.

O SSIM entre a imagem reconstruída  $\hat{\mathbf{Y}}$  e a imagem de referência  $\mathbf{Y}$  é definido por:

$$\text{SSIM}(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{(2\mu_{\hat{\mathbf{Y}}}\mu_{\mathbf{Y}} + C_1)(2\sigma_{\hat{\mathbf{Y}}\mathbf{Y}} + C_2)}{(\mu_{\hat{\mathbf{Y}}}^2 + \mu_{\mathbf{Y}}^2 + C_1)(\sigma_{\hat{\mathbf{Y}}}^2 + \sigma_{\mathbf{Y}}^2 + C_2)}, \quad (3.15)$$

em que  $\mu_{\hat{\mathbf{Y}}}$  e  $\mu_{\mathbf{Y}}$  representam as médias locais da imagem reconstruída e da imagem de referência, respectivamente;  $\sigma_{\hat{\mathbf{Y}}}^2$  e  $\sigma_{\mathbf{Y}}^2$  representam suas variâncias;  $\sigma_{\hat{\mathbf{Y}}\mathbf{Y}}$  representa a covariância entre as duas

imagens; e  $C_1$  e  $C_2$  são constantes de estabilização numérica, utilizadas para evitar instabilidades quando os denominadores assumem valores próximos de zero.

Valores de SSIM variam tipicamente entre 0 e 1, sendo que valores mais próximos de 1 indicam maior similaridade estrutural. No contexto da remoção de nuvens, essa métrica é particularmente relevante porque permite avaliar se a reconstrução preserva estruturas espaciais importantes, como bordas, texturas, transições entre classes de cobertura do solo e padrões geométricos da cena.

Apesar disso, o SSIM também possui limitações. A métrica pode ser menos sensível a pequenas variações radiométricas globais e não avalia diretamente a coerência semântica da reconstrução. Portanto, ela foi utilizada como complemento ao PSNR e ao MAE, e não como critério isolado de avaliação.

### 3.5.3 MAE

O *Mean Absolute Error* (MAE) mede o erro absoluto médio entre a imagem reconstruída e a imagem de referência. Diferentemente do MSE, que eleva os erros ao quadrado e, portanto, penaliza mais fortemente discrepâncias grandes, o MAE calcula a diferença média em magnitude absoluta:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{\mathbf{Y}}_i - \mathbf{Y}_i|. \quad (3.16)$$

em que  $N$  representa o número total de pixels avaliados, incluindo os canais da imagem.

Valores menores de MAE indicam menor discrepância média entre a reconstrução e a imagem alvo. Como as imagens foram normalizadas para o intervalo  $[0, 1]$ , o MAE pode ser interpretado diretamente como a diferença absoluta média normalizada entre os pixels reconstruídos e os pixels de referência.

No contexto deste trabalho, o MAE é útil por fornecer uma medida simples e interpretável do erro médio. Além disso, por não elevar os erros ao quadrado, ele é menos sensível a valores extremos do que o MSE e o PSNR. Dessa forma, o MAE complementa o PSNR, permitindo verificar se ganhos em relação sinal-ruído também correspondem a reduções consistentes do erro absoluto médio.

### 3.5.4 Indicadores de Custo Computacional

Além da qualidade das reconstruções, avaliou-se também o custo computacional dos modelos. Essa análise é importante porque arquiteturas mais complexas podem apresentar melhor desempenho quantitativo, mas exigir maior tempo de inferência, mais memória ou maior capacidade de processamento.

Foram considerados quatro indicadores principais:

- Número de parâmetros treináveis: indica a complexidade do modelo e a quantidade de pesos ajustados durante o treinamento.
- Latência média de inferência: corresponde ao tempo médio necessário para processar uma imagem.
- Throughput: mede a quantidade de imagens processadas por segundo.
- Memória máxima alocada na GPU: indica o consumo de memória durante a execução do modelo.

Esses indicadores permitem avaliar o compromisso entre qualidade de reconstrução e viabilidade prática. No caso de aplicações *offline*, como pré-processamento em lote de imagens de sensoriamento remoto, modelos com maior latência podem ser aceitáveis caso produzam reconstruções de melhor qualidade. Por outro lado, em aplicações com restrições severas de tempo ou recursos computacionais, modelos mais leves podem ser preferíveis.

## 4 RESULTADOS E DISCUSSÃO

Este capítulo apresenta a avaliação experimental da arquitetura WAMG-UNet para a tarefa de reconstrução de imagens de sensoriamento remoto degradadas por nuvens. Para facilitar a leitura, os resultados foram organizados em quatro blocos: configuração experimental, análise quantitativa, análise qualitativa e custo computacional, seguidos pela discussão geral. Além da configuração completa do modelo proposto, foram avaliados modelos base (*baseline*) e variantes ablativas, com o intuito de isolar o impacto da máscara de nuvem, dos componentes baseados em wavelets e da perda auxiliar no domínio da frequência.

O desenvolvimento experimental foi conduzido em ambiente Python, utilizando a biblioteca PyTorch para a implementação das redes neurais, otimização e compilação do modelo. O carregamento e as transformações das imagens foram gerenciados pelas bibliotecas Pillow e Torchvision, enquanto as métricas de avaliação (PSNR e SSIM) foram calculadas por meio do pacote TorchMetrics. A manipulação de dados estruturados e operações matriciais foi realizada com o auxílio das bibliotecas Pandas e NumPy. Por fim, o rastreamento do experimento e a geração de gráficos de evolução e reconstrução utilizaram as bibliotecas TensorBoard e Matplotlib.

Os experimentos foram executados com semente aleatória fixa igual a 42, utilizando imagens redimensionadas para  $256 \times 256$  pixels. A avaliação final foi realizada sobre 4,250 amostras, sendo 1,881 provenientes do subconjunto RICE1 e 2,369 provenientes do subconjunto RICE2. O processamento foi conduzido em ambiente com 32 GB de RAM, 8 vCPUs Intel Xeon @ 2,3 GHz e uma GPU NVIDIA L4 com 24 GB de VRAM.

### 4.1 Configuração Experimental

O protocolo experimental foi estruturado para comparar a arquitetura proposta com baselines progressivamente mais informativos. Essa organização permite avaliar não apenas o desempenho final da WAMG-UNet, mas também a contribuição relativa de seus principais componentes.

A Tabela 2 apresenta os modelos avaliados. A U-Net RGB corresponde ao baseline mais simples, utilizando apenas a imagem nublada como entrada. A U-Net RGB+Mask adiciona a máscara de nuvem como canal auxiliar, permitindo avaliar o impacto direto da informação espacial de oclusão. A variante WAMG sem perda wavelet mantém a arquitetura proposta, mas

remove a penalização no domínio da frequência. Por fim, a WAMG-UNet Completa corresponde à configuração completa do modelo proposto.

Tabela 2 – Modelos avaliados no protocolo experimental.

| Modelo                 | Entrada       | Parâmetros | Perda wavelet | Descrição                                  |
|------------------------|---------------|------------|---------------|--------------------------------------------|
| U-Net RGB              | RGB           | 1,91M      | Não           | Baseline sem máscara explícita             |
| U-Net RGB+Mask         | RGB + máscara | 1,93M      | Não           | Baseline com máscara concatenada à entrada |
| WAMG sem perda wavelet | RGB + máscara | 3,58M      | Não           | Variante ablativa da arquitetura proposta  |
| WAMG-UNet Completa     | RGB + máscara | 3,58M      | Sim           | Modelo proposto completo                   |

Fonte: Autoria própria (2026).

Todos os modelos foram treinados sob o mesmo protocolo geral, com semente aleatória fixa, mesma resolução de entrada, mesma partição de validação e mesmo critério de seleção do melhor checkpoint. A escolha do melhor modelo em cada configuração foi realizada com base no maior valor de PSNR observado no conjunto de validação.

Os principais hiperparâmetros utilizados foram:

- semente aleatória: 42;
- resolução de entrada:  $256 \times 256$  pixels;
- otimizador: Adam;
- taxa de aprendizado inicial:  $4 \times 10^{-4}$ ;
- tamanho do processamento em lote: 128;
- agendamento da taxa de aprendizado: *ReduceLROnPlateau*;
- critério de parada antecipada: paciência de 15 épocas;
- número máximo de épocas: 70.

Na configuração completa da WAMG-UNet, a função de perda foi composta por um termo espacial e um termo no domínio wavelet:

$$\mathcal{L}_{total} = \mathcal{L}_{pixel} + \lambda_w \mathcal{L}_{wavelet}. \quad (4.1)$$

Na configuração da WAMG-UNet completa, esta perda era modulada por um algoritmo de Curriculum Learning, no entanto, em outras configurações sem perda de frequência,  $\lambda_w$  assumia o valor de 0.

#### 4.1.1 Treinamento do Modelo Proposto

A WAMG-UNet Completa foi treinada por até 70 épocas. Durante o treinamento, o modulador (*scheduler*) de taxa de aprendizado reduziu progressivamente o valor inicial,

encerrando a execução com taxa de aprendizado de  $4,0 \times 10^{-6}$ . Esse comportamento indica que o modelo atingiu uma região de menor variação no conjunto de validação, com refinamentos incrementais nas últimas épocas.

A Tabela 3 resume o treinamento da configuração completa.

Tabela 3 – Resumo do treinamento da WAMG-UNet Completa.

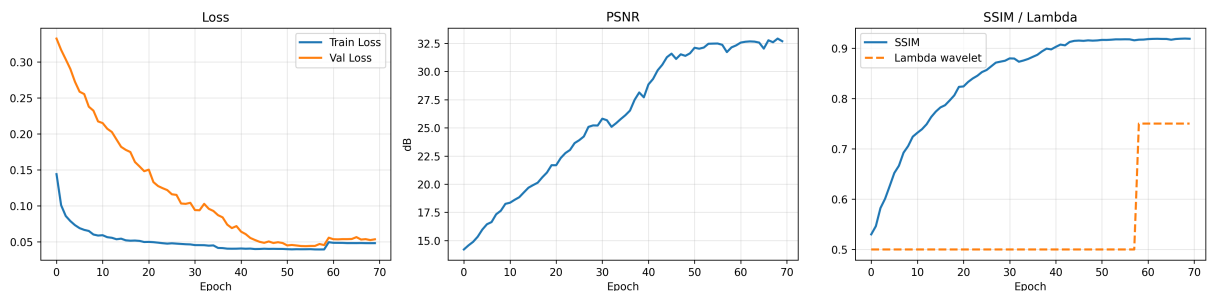
| Métrica                     | Valor                |
|-----------------------------|----------------------|
| Épocas realizadas           | 70                   |
| Melhor PSNR de validação    | 32,93 dB             |
| Loss de treino              | 0,0479               |
| Loss de validação           | 0,0533               |
| Componente espacial da loss | 0,0220               |
| Componente wavelet da loss  | 0,0417               |
| Learning rate final         | $4,0 \times 10^{-6}$ |
| Tempo médio por época       | 285 s                |

Fonte: Autoria própria (2026).

A decomposição da perda indica que o termo espacial permaneceu como componente central da otimização, enquanto a perda wavelet atuou como regularização auxiliar voltada à preservação de detalhes em sub-bandas de frequência.

As curvas de treinamento, apresentadas na Figura 13, mostram a evolução da perda total, bem como as métricas de PSNR e SSIM comparado ao valor da perda de frequência.

Figura 13 – Curva de treinamento da WAMG-UNet Completa.



Fonte: Autoria própria (2026).

## 4.2 Resultados

A Tabela 4 apresenta a comparação global entre os quatro modelos avaliados. Foram consideradas três métricas de qualidade: PSNR, SSIM e MAE. O PSNR mede a relação sinal-ruído da reconstrução, de modo que valores mais elevados indicam menor distorção em relação à imagem de referência. De forma geral, valores acima de 30 dB sugerem reconstruções

visualmente aceitáveis, enquanto valores acima de 40 dB indicam alta fidelidade; contudo, esses intervalos devem ser interpretados como referências aproximadas, pois a qualidade perceptual também depende das características da imagem e dos artefatos gerados. O SSIM avalia a similaridade estrutural entre as imagens e varia de -1 a 1, sendo valores próximos de 1 associados a maior preservação de luminância, contraste e estrutura. Valores em torno de 0,90 ou superiores são frequentemente considerados indicativos de alta similaridade estrutural (Wang *et al.*, 2004). Por fim, o MAE quantifica o erro absoluto médio entre a imagem reconstruída e a referência, de modo que valores menores indicam menor discrepância média entre os pixels.

Tabela 4 – Comparação quantitativa global entre os modelos avaliados no conjunto de validação.

| Modelo                    | PSNR (dB) ↑  | SSIM ↑       | MAE ↓         | Latência (ms) ↓ | Mem. GPU (MB) ↓ |
|---------------------------|--------------|--------------|---------------|-----------------|-----------------|
| U-Net RGB                 | 31,40        | 0,916        | 0,0251        | 1,90            | 306,7           |
| U-Net RGB+Mask            | 32,22        | 0,913        | 0,0227        | 1,92            | 310,2           |
| WAMG sem perda wavelet    | 32,91        | 0,916        | 0,0215        | 4,58            | 441,8           |
| <b>WAMG-UNet Completa</b> | <b>32,93</b> | <b>0,919</b> | <b>0,0215</b> | 4,60            | 441,8           |

Fonte: Autoria própria (2026).

#### 4.2.1 Comparação Global

A WAMG-UNet Completa obteve o melhor desempenho global, alcançando 32,93 dB de PSNR, SSIM de 0,919 e MAE de 0,0215. Em relação ao baseline U-Net RGB, o ganho total foi de 1,53 dB em PSNR, acompanhado de redução do erro absoluto médio. Esse resultado indica que a combinação entre máscara explícita, arquitetura baseada em wavelets e perda auxiliar no domínio da frequência contribuiu para melhorar a fidelidade da reconstrução.

A comparação entre U-Net RGB e U-Net RGB+Mask mostra que a simples inclusão da máscara como canal de entrada elevou o PSNR de 31,40 dB para 32,22 dB, representando ganho de 0,82 dB. Apesar disso, o SSIM apresentou pequena redução, de 0,916 para 0,913. Esse comportamento sugere que a máscara favoreceu a redução do erro absoluto e o aumento da relação sinal-ruído, mas não necessariamente melhorou todos os aspectos estruturais capturados pela métrica SSIM.

A comparação entre U-Net RGB+Mask e WAMG sem perda wavelet permite avaliar o impacto da arquitetura proposta sem considerar a perda no domínio da frequência. Nesse caso, o PSNR aumentou de 32,22 dB para 32,91 dB, indicando ganho de 0,69 dB associado à estrutura WAMG, que combina componentes wavelet e conexões de salto guiadas por máscara.

Por fim, a diferença entre WAMG sem perda wavelet e WAMG-UNet Completa foi

pequena em termos de PSNR agregado, passando de 32,91 dB para 32,93 dB. Esse resultado indica que a perda wavelet teve contribuição quantitativa marginal nas métricas globais, embora possa exercer efeito qualitativo na preservação de detalhes finos e transições locais.

#### 4.2.2 Análise por Subconjuntos

Os subconjuntos RICE1 e RICE2 representam cenários distintos de degradação por nuvens. Enquanto o RICE1 contém predominantemente nuvens finas e semitransparentes, o RICE2 apresenta nuvens mais espessas e regiões de oclusão mais severa. A avaliação separada por subconjunto permite verificar se o desempenho global dos modelos se mantém em diferentes regimes de dificuldade.

A Tabela 5 apresenta os valores de PSNR obtidos em cada subconjunto.

Tabela 5 – Desempenho em PSNR por subconjunto do dataset RICE.

| Subconjunto | U-Net RGB | U-Net RGB+Mask | WAMG s/ perda wavelet | WAMG Completa | Amostras |
|-------------|-----------|----------------|-----------------------|---------------|----------|
| RICE1       | 30,25     | 32,15          | 32,91                 | <b>32,95</b>  | 1,881    |
| RICE2       | 32,31     | 32,28          | 32,90                 | <b>32,92</b>  | 2,369    |

Fonte: Autoria própria (2026).

No RICE1, a diferença entre os modelos é mais evidente. A adição da máscara à U-Net elevou o PSNR de 30,25 dB para 32,15 dB, indicando ganho de 1,90 dB. Esse resultado sugere que, em cenários de nuvens mais finas, a informação explícita sobre a localização da oclusão é particularmente útil para orientar a reconstrução. A WAMG-UNet Completa atingiu 32,95 dB, superando o baseline RGB em 2,70 dB.

No RICE2, as diferenças entre U-Net RGB e U-Net RGB+Mask foram menores. A U-Net RGB obteve 32,31 dB, enquanto a U-Net RGB+Mask obteve 32,28 dB. Esse comportamento sugere que, em cenários de nuvens mais espessas, a simples indicação espacial da máscara pode não ser suficiente para recuperar informação de superfície ausente, pois a tarefa passa a depender mais fortemente de inferência contextual. Ainda assim, a WAMG-UNet Completa manteve o melhor desempenho, com 32,92 dB.

Esses resultados indicam que a WAMG-UNet apresentou desempenho superior nos dois subconjuntos, mas o ganho relativo foi mais expressivo no RICE1. Isso é coerente com a natureza da tarefa: em nuvens finas, parte da estrutura visual permanece disponível e pode ser explorada pelo modelo; em nuvens espessas, a reconstrução torna-se mais incerta, pois a informação da superfície pode estar severamente degradada ou ausente.

#### 4.2.3 Análise por Cobertura de Nuvem

Além da avaliação por subconjunto, os resultados foram estratificados em intervalos de cobertura de nuvens. Essa análise permite investigar como os modelos se comportam à medida que a proporção de pixels degradados aumenta.

A Tabela 6 apresenta os valores de PSNR para cinco faixas de cobertura.

Tabela 6 – PSNR por intervalo de cobertura de nuvens.

| Cobertura | U-Net RGB | U-Net RGB+Mask | WAMG s/ perda wavelet | WAMG Completa | Amostras |
|-----------|-----------|----------------|-----------------------|---------------|----------|
| 0–20%     | 33,10     | 33,99          | 34,81                 | <b>34,95</b>  | 1,155    |
| 20–40%    | 31,89     | 32,86          | <b>33,57</b>          | 33,56         | 1,023    |
| 40–60%    | 31,04     | 31,93          | <b>32,61</b>          | 32,49         | 878      |
| 60–80%    | 30,57     | 31,43          | 32,01                 | <b>32,06</b>  | 648      |
| 80–100%   | 28,43     | 28,67          | 29,19                 | <b>29,25</b>  | 546      |

Fonte: Autoria própria (2026).

Observa-se uma tendência geral de queda no PSNR à medida que a cobertura de nuvens aumenta. Esse comportamento é esperado, pois altos níveis de oclusão reduzem a quantidade de informação visual disponível para guiar a reconstrução. Em coberturas entre 0% e 20%, a WAMG-UNet Completa alcançou 34,95 dB, enquanto no intervalo de 80% a 100% o desempenho caiu para 29,25 dB.

A WAMG-UNet Completa apresentou o melhor ou um dos melhores desempenhos em todas as faixas de cobertura. Entretanto, a diferença entre a configuração completa e a variante sem perda wavelet permaneceu pequena, geralmente inferior a 0,15 dB. Esse resultado reforça a interpretação de que a maior contribuição quantitativa está associada à arquitetura WAMG, enquanto a perda wavelet exerce papel complementar.

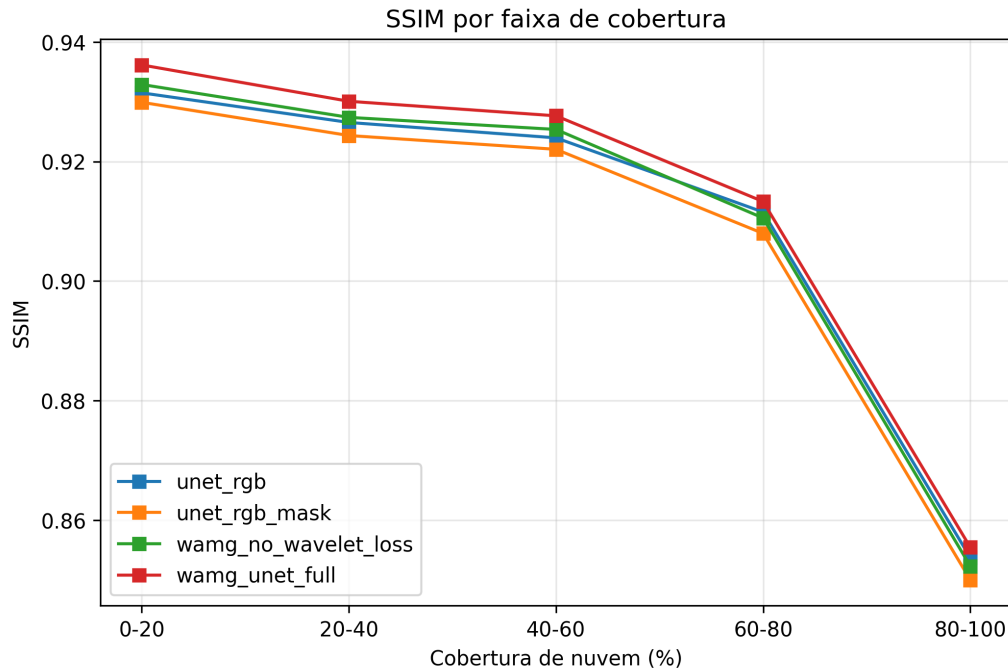
Nos cenários de cobertura severa, todos os modelos apresentaram redução significativa de desempenho. Isso evidencia uma limitação inerente à reconstrução a partir de imagem única: quando grande parte da cena está ocluída, o modelo passa a depender predominantemente de padrões aprendidos e contexto espacial, não de informação observada diretamente.

#### 4.2.4 Métricas Estruturais

Além do PSNR, o SSIM foi utilizado para avaliar a similaridade estrutural entre as reconstruções e as imagens de referência. Essa métrica é particularmente relevante em tarefas de restauração, pois considera variações de luminância, contraste e estrutura local.

A Figura 14 apresenta o comportamento do SSIM em função da cobertura de nuvens.

Figura 14 – SSIM por intervalo de cobertura de nuvens para os modelos avaliados.



Fonte: Autoria própria (2026).

O comportamento do SSIM acompanha a tendência observada no PSNR: as maiores similaridades estruturais ocorrem em faixas de menor cobertura, enquanto coberturas severas resultam em degradação perceptual mais evidente. Esse padrão confirma que o aumento da oclusão compromete não apenas o erro médio, mas também a preservação da estrutura visual da cena.

A Figura 15 apresenta a variação do PSNR por faixa de cobertura, reforçando a mesma tendência quantitativa.

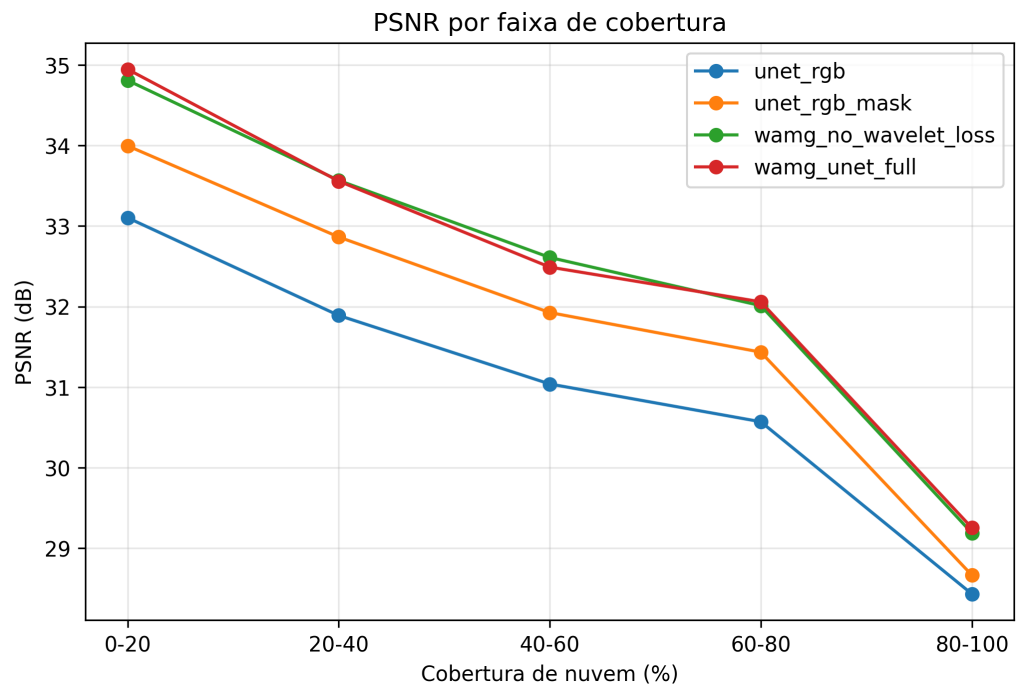
A análise conjunta de PSNR e SSIM indica que a WAMG-UNet mantém vantagem relativa nos cenários avaliados, mas também evidencia que o desempenho é fortemente condicionado pela quantidade de informação observável na imagem de entrada.

#### 4.2.5 Análise Qualitativa

A análise qualitativa complementa as métricas numéricas ao permitir observar diferenças locais nas reconstruções. A Figura 16 apresenta exemplos representativos das saídas produzidas pelos modelos avaliados.

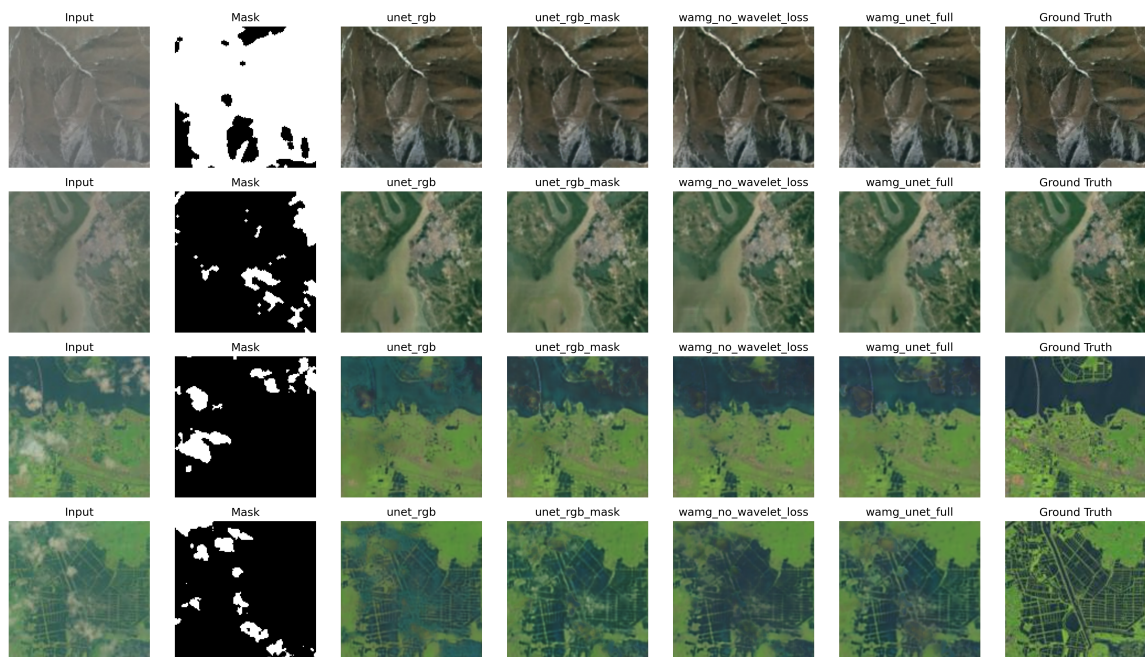
Em cenários de cobertura leve, a WAMG-UNet tende a preservar melhor estruturas locais, bordas e texturas finas. Os baselines U-Net também conseguem reconstruir a estrutura global da cena, mas apresentam maior tendência à suavização, especialmente em regiões onde a nuvem

Figura 15 – PSNR por intervalo de cobertura de nuvens para os modelos avaliados.



Fonte: Autoria própria (2026).

Figura 16 – Comparação qualitativa entre os modelos avaliados em diferentes cenários de cobertura de nuvens.



Fonte: Autoria própria (2026).

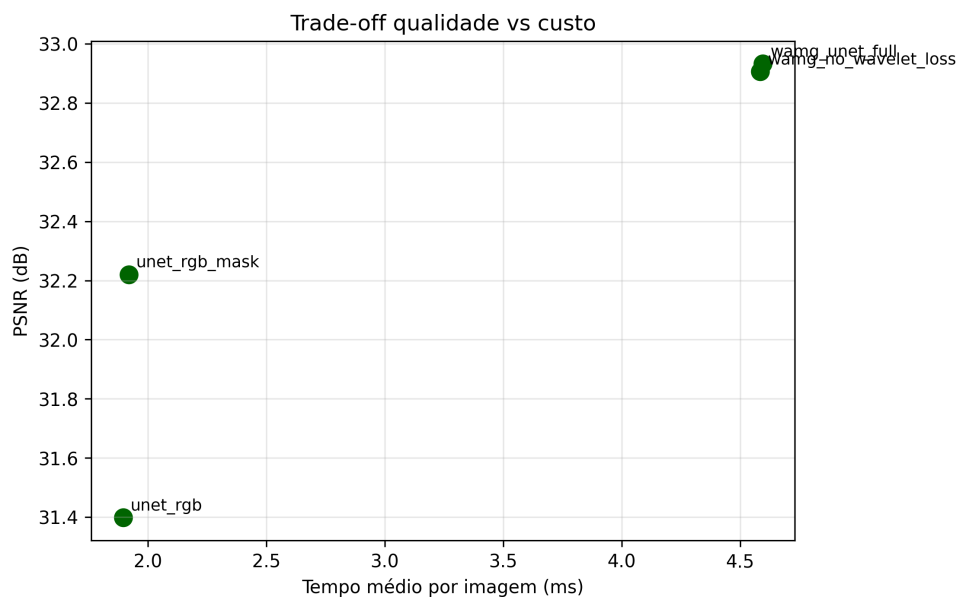
altera o contraste local. Esse comportamento é coerente com o uso de componentes wavelet, que favorecem a preservação de informação em diferentes escalas espaciais.

Em coberturas moderadas, as diferenças visuais entre os modelos tornam-se mais sutis. Ainda assim, a WAMG-UNet apresenta maior consistência em regiões de transição entre áreas nubladas e livres de nuvem. Já em cenários de oclusão severa, todos os modelos exibem maior incerteza e podem produzir artefatos visuais, refletindo a dificuldade de reconstruir informação que não está diretamente presente na entrada.

### 4.3 Análise de Custo Computacional

A avaliação de custo computacional é importante para verificar a viabilidade prática dos modelos em cenários reais de processamento de imagens. A Figura 17 apresenta a relação entre qualidade de reconstrução, medida por PSNR, e latência média de inferência.

Figura 17 – Relação entre qualidade de reconstrução, medida por PSNR, e latência de inferência dos modelos avaliados.



Fonte: Autoria própria (2026).

A WAMG-UNet ocupa a região de maior qualidade, mas também apresenta maior latência em comparação às U-Nets convencionais. Esse aumento de custo está associado aos componentes adicionais da arquitetura, especialmente às operações de decomposição/reconstrução wavelet e ao mecanismo de controle por máscara. Ainda assim, a latência média de 4,60 ms indica viabilidade para aplicações de processamento em lote ou pré-processamento *offline* de imagens.

Por outro lado, os baselines U-Net apresentam menor latência e maior throughput, sendo

mais adequados para cenários em que a velocidade é o principal requisito. A U-Net RGB+Mask representa um compromisso interessante: apresenta ganho expressivo em PSNR em relação à U-Net RGB, mantendo praticamente o mesmo custo computacional.

A Tabela 7 sintetiza as métricas de eficiência computacional dos modelos avaliados.

Tabela 7 – Comparação de eficiência computacional entre os modelos avaliados.

| Modelo                 | Parâmetros | Latência (ms) ↓ | Throughput (img/s) ↑ | GPU Mem. (MB) ↓ |
|------------------------|------------|-----------------|----------------------|-----------------|
| U-Net RGB              | 1,91M      | 1,90            | 527,7                | 306,7           |
| U-Net RGB+Mask         | 1,93M      | 1,92            | 520,9                | 310,2           |
| WAMG sem perda wavelet | 3,58M      | 4,58            | 218,2                | 441,8           |
| WAMG-UNet Completa     | 3,58M      | 4,60            | 217,6                | 441,8           |

Fonte: Autoria própria (2026).

Os modelos WAMG possuem aproximadamente 1,87 vezes mais parâmetros que os baselines U-Net e apresentam latência cerca de 2,4 vezes maior. Esse custo adicional, entretanto, resulta em melhor qualidade de reconstrução. A decisão entre utilizar WAMG-UNet ou uma U-Net convencional depende, portanto, do contexto de aplicação. Em cenários *offline*, nos quais a qualidade visual e a fidelidade estrutural são prioritárias, o custo adicional da WAMG-UNet é justificável. Em cenários de processamento massivo com restrições severas de tempo, a U-Net RGB+Mask pode representar uma alternativa eficiente.

#### 4.4 Discussão

Os resultados obtidos sustentam a hipótese de que a integração entre máscara explícita, processamento multiescala por wavelets e arquitetura *encoder-decoder* pode melhorar a reconstrução de imagens afetadas por nuvens. O ganho total da WAMG-UNet Completa em relação à U-Net RGB foi de 1,53 dB em PSNR, acompanhado de redução no MAE e melhor SSIM global.

A análise incremental dos componentes é apresentada na Tabela 8. Essa tabela organiza os modelos em ordem crescente de complexidade, permitindo observar a contribuição relativa de cada modificação arquitetural ou função de perda.

Tabela 8 – Análise incremental da contribuição dos componentes avaliados.

| Modelo                 | PSNR (dB) | Ganho incremental | Componente avaliado           |
|------------------------|-----------|-------------------|-------------------------------|
| U-Net RGB              | 31,40     | —                 | Baseline                      |
| U-Net RGB+Mask         | 32,22     | +0,82             | Máscara como canal de entrada |
| WAMG sem perda wavelet | 32,91     | +0,69             | Arquitetura WAMG              |
| WAMG-UNet Completa     | 32,93     | +0,02             | Perda wavelet                 |

Fonte: Autoria própria (2026).

A primeira contribuição observada é a utilidade da máscara de nuvem. A adição da máscara como canal de entrada produziu ganho de 0,82 dB em relação à U-Net RGB, com custo computacional praticamente inalterado. Esse resultado indica que informar explicitamente a localização da degradação ajuda a rede a diferenciar regiões confiáveis de regiões que exigem reconstrução.

A segunda contribuição está associada à arquitetura WAMG. Ao comparar U-Net RGB+Mask com WAMG sem perda wavelet, observa-se ganho de 0,69 dB. Esse resultado sugere que o uso de operações wavelet e conexões de salto guiadas por máscara contribui para melhorar a reconstrução, sobretudo por oferecer uma representação multiescala mais adequada para preservar estruturas locais.

A terceira contribuição, relacionada à perda wavelet, apresentou impacto pequeno nas métricas globais. O ganho de 0,02 dB em PSNR indica que a perda no domínio da frequência não foi o principal fator de melhoria quantitativa agregada. Ainda assim, sua inclusão pode ser útil para favorecer detalhes finos em casos específicos, sendo recomendável que estudos futuros investiguem métricas complementares de textura e frequência para avaliar melhor esse efeito.

Apesar dos resultados favoráveis, algumas limitações foram identificadas. A primeira está relacionada aos cenários de cobertura severa. Na faixa de 80% a 100% de cobertura, todos os modelos apresentaram queda expressiva de desempenho, com PSNR próximo de 29 dB. Isso indica que, quando grande parte da imagem está ocluída, a reconstrução a partir de uma única aquisição óptica torna-se altamente incerta.

A segunda limitação refere-se ao comportamento no RICE2. Nesse subconjunto, a diferença entre U-Net RGB e U-Net RGB+Mask foi praticamente nula, sugerindo que a máscara nem sempre agrega informação suficiente em casos de nuvens espessas. Nesses cenários, estratégias multitemporais ou multimodais, como o uso combinado de dados ópticos e SAR, podem ser mais adequadas.

A terceira limitação é o custo computacional. Embora a WAMG-UNet apresente latência baixa em termos absolutos, ela é mais custosa que os baselines U-Net. Assim, sua adoção deve considerar o compromisso entre qualidade de reconstrução e eficiência de processamento.

Por fim, a contribuição da perda wavelet foi marginal nas métricas globais utilizadas. Isso não invalida sua utilidade, mas indica que PSNR, SSIM e MAE podem não capturar completamente diferenças finas de textura e frequência. Avaliações futuras podem incluir métricas perceptuais, espectrais ou específicas para análise de alta frequência.

#### 4.4.1 Comparação com a Literatura

Os resultados da WAMG-UNet foram contrastados com valores reportados na literatura que utilizam o dataset RICE e métricas como PSNR e SSIM. A Tabela 9 reúne esses trabalhos, organizados por subconjunto e em ordem cronológica. Ressalta-se que esta comparação é apenas indicativa e não constitui uma comparação direta, uma vez que os protocolos experimentais diferem entre si em aspectos como particionamento dos dados, recorte em patches, normalização das imagens, definição dos conjuntos de validação, escolha das métricas e implementação. Os valores aqui reportados foram extraídos diretamente dos artigos originais, sem reexecução dos modelos no mesmo ambiente computacional.

Tabela 9 – Comparação indicativa com a literatura no dataset RICE.

| <b>Trabalho</b>              | <b>Método</b> | <b>PSNR RICE1</b> | <b>PSNR RICE2</b> |
|------------------------------|---------------|-------------------|-------------------|
| (Lin <i>et al.</i> , 2019)   | pix2pix       | 31,03             | 30,04             |
| (Chen; Chen; Li, 2021)       | AGGAN         | 31,54             | —                 |
| (Liu <i>et al.</i> , 2022)   | SI-SA GAN     | 31,39             | —                 |
| (Zi <i>et al.</i> , 2023)    | WaveCNN-CR    | 35,74             | —                 |
| (Shen <i>et al.</i> , 2024)  | GAN proposta  | —                 | 35,11             |
| (Huang <i>et al.</i> , 2024) | ACA-CRNet     | —                 | 33,70             |
| (Yu; Idris; Wang, 2025)      | DC4CR         | 35,54             | 33,68             |
| <b>WAMG-UNet</b>             | WAMG-UNet     | 32,95             | 32,92             |

Fonte: Autoria própria, com valores extraídos dos trabalhos citados.

(Lin *et al.*, 2019) propuseram o conjunto de dados RICE e reportaram a pix2pix como baseline em ambos os subconjuntos, alcançando PSNR de 31,03 dB no RICE1 e 30,04 dB no RICE2. (Chen; Chen; Li, 2021) empregam uma GAN com atenção espacial (AGGAN) no RICE-I, recortando 49 patches de  $224 \times 224$  por par de imagem, com PSNR de 31,54 dB e SSIM de 0,9508. (Liu *et al.*, 2022) propõem a SI-SA GAN, que combina informação espacial e auto-atenção, com divisão 8:2 entre treinamento e validação, reportando PSNR de 31,39 dB e mSSIM de 0,9732. (Zi *et al.*, 2023) propuseram a WaveCNN-CR, que integra wavelets à CNN para remoção de nuvens finas no RICE1 (divisão 400/100), alcançando PSNR de 35,74 dB e SSIM de 0,9650.

No RICE2, (Shen *et al.*, 2024) propuseram uma GAN com extração de textura, espaço de cor HSV/RGB e módulo de alta frequência, formulada como *dehazing*(nuvens finas), reportando PSNR de 35,11 dB e SSIM de 0,8764. (Huang *et al.*, 2024) propuseram o ACA-CRNet, que

integra o mecanismo AC-Attention ao DSen2-CR, avaliado com patches  $128 \times 128$  e particionamento próprio, alcançando PSNR de 33,70 dB e SSIM de 0,891. (Yu; Idris; Wang, 2025) propuseram o DC4CR, *framework* baseado em difusão que utiliza Stable Diffusion pré-treinado (divisão 80/20), reportando PSNR de 35,54 dB (RICE1) e 33,68 dB (RICE2), com SSIM de 0,9887 e 0,8859, respectivamente.

A WAMG-UNet obteve PSNR de 32,95 dB no RICE1 e 32,92 dB no RICE2. No RICE1, esse valor supera os baselines AGGAN e SI-SA GAN e situa-se abaixo dos métodos mais recentes (WaveCNN-CR e DC4CR), que se beneficiam de arquiteturas mais profundas, pré-treinamento em grandes bases ou divisões de dados distintas. No RICE2, o PSNR da WAMG-UNet é competitivo com o ACA-CRNet (33,70 dB) e o DC4CR (33,68 dB), embora inferior ao método de (Shen *et al.*, 2024) (35,11 dB), cujo protocolo de *dehazing* difere substancialmente.

Do ponto de vista arquitetural, a WAMG-UNet utiliza uma abordagem distinta: integra máscara explícita de nuvem, processamento wavelet multiescala e perda auxiliar no domínio da frequência. As demais abordagens aqui discutidas baseiam-se em GANs (Chen; Chen; Li, 2021); (Liu *et al.*, 2022); (Shen *et al.*, 2024), atenção contextual (ACA-CRNet), wavelets integradas à CNN (WaveCNN-CR) ou modelos de difusão (DC4CR). Essas diferenças metodológicas tornam cada abordagem particularmente adequada a contextos específicos.

Em paradigma distinto, (Zou *et al.*, 2024) apresentam o DiffCR, um modelo de difusão condicional que opera sobre datasets multitemporais Sentinel-2 (Sen2MTC). A comparação numérica direta não se aplica, dado que o protocolo difere substancialmente da abordagem de imagem única adotada neste trabalho. O DiffCR é citado como referência de tendência da área, com custo computacional elevado (22,91 M de parâmetros e 45,86 G de MACs), contrastando com os 3,58 M de parâmetros e 4,60 ms de latência da WAMG-UNet.

Em síntese, os resultados obtidos situam a WAMG-UNet em uma faixa competitiva em relação aos valores reportados na literatura para o dataset RICE, especialmente quando se considera a consistência entre os dois subconjuntos. A comparação, entretanto, é apenas contextual e não conclusiva, servindo como referência para o leitor situar a ordem de grandeza do desempenho obtido.

Em conjunto, os resultados indicam que a WAMG-UNet representa uma alternativa competitiva para reconstrução de imagens de sensoriamento remoto degradadas por nuvens, especialmente quando há disponibilidade de máscara de nuvem para orientar o processo.

## 5 CONCLUSÃO

### 5.1 Considerações Finais

A presença de nuvens em imagens ópticas de sensoriamento remoto representa uma limitação recorrente para o monitoramento contínuo da superfície terrestre. Em sensores passivos, a oclusão atmosférica pode degradar parcial ou totalmente a informação radiométrica da cena, comprometendo análises posteriores de uso e cobertura do solo, vegetação, recursos hídricos e dinâmica territorial. Nesse contexto, técnicas de reconstrução baseadas em aprendizado profundo surgem como alternativas promissoras para restaurar regiões afetadas por nuvens, especialmente quando há disponibilidade de imagens de referência e máscaras de oclusão para treinamento supervisionado.

Este trabalho propôs e avaliou a WAMG-UNet, uma arquitetura neural baseada na estrutura encoder-decoder da U-Net, adaptada para a tarefa de reconstrução de imagens de sensoriamento remoto degradadas por nuvens. A proposta combinou três elementos principais: uso explícito da máscara de nuvem, processamento multiescala por meio da Transformada Wavelet Discreta e conexões de salto guiadas por máscara. Além disso, a configuração completa incorporou uma perda auxiliar no domínio wavelet, com modulação curricular do seu peso durante o treinamento.

A motivação da arquitetura proposta esteve associada a duas limitações recorrentes em redes encoder-decoder convencionais. A primeira refere-se à perda de detalhes finos durante as etapas de subamostragem, especialmente quando operações como *Max Pooling* são utilizadas. A segunda está relacionada à possibilidade de propagação de artefatos atmosféricos pelas conexões de salto, uma vez que essas conexões transferem diretamente mapas de características do codificador para o decodificador. A WAMG-UNet buscou mitigar esses problemas substituindo a subamostragem convencional por operações baseadas em wavelets e modulando as conexões de salto com auxílio da máscara de nuvem.

Os experimentos foram conduzidos sobre os subconjuntos RICE1 e RICE2, permitindo avaliar o comportamento dos modelos em cenários distintos de degradação. Foram comparadas quatro configurações principais: U-Net RGB, U-Net RGB+Mask, WAMG sem perda wavelet e WAMG-UNet Completa. Essa organização possibilitou analisar de forma incremental a contribuição da máscara, da arquitetura WAMG e da perda wavelet com Curriculum Learning.

Os resultados quantitativos indicaram que a WAMG-UNet Completa obteve o melhor

desempenho global entre os modelos avaliados, alcançando PSNR de 32,93 dB, SSIM de 0,919 e MAE de 0,0215 no conjunto de validação. Em relação ao baseline U-Net RGB, o modelo proposto apresentou ganho de 1,53 dB em PSNR. A comparação entre U-Net RGB e U-Net RGB+Mask evidenciou a relevância da máscara de nuvem como informação auxiliar, enquanto a comparação entre U-Net RGB+Mask e WAMG sem perda wavelet mostrou o impacto positivo da arquitetura proposta.

A análise ablativa indicou que a maior parte do ganho quantitativo decorreu da arquitetura WAMG, especialmente da combinação entre processamento wavelet e conexões de salto guiadas por máscara. A inclusão da perda wavelet com Curriculum Learning na WAMG-UNet Completa produziu ganho incremental pequeno nas métricas globais, elevando o PSNR de 32,91 dB para 32,93 dB em relação à variante sem essa perda. Assim, conclui-se que a perda wavelet modulada curricularmente atuou principalmente como componente de refinamento, enquanto a contribuição mais expressiva esteve associada à estrutura arquitetural do modelo.

A comparação contextual com a literatura situou a WAMG-UNet em faixa competitiva no dataset RICE. No RICE1, o PSNR de 32,95 dB demonstra competitividade com os modelos de estado da arte AGGAN(Chen; Chen; Li, 2021) e SI-SA GAN(Liu *et al.*, 2022), WaveCNN-CR(Zi *et al.*, 2023) e DC4CR(Yu; Idris; Wang, 2025), que empregam arquiteturas mais profundas ou pré-treinamento em grandes bases. No RICE2, os 32,92 dB são próximos ao ACA-CRNet (33,70 dB) e ao DC4CR (33,68 dB). Esses valores, contrastados com protocolos distintos entre os trabalhos, servem para balizar o desempenho da WAMG-UNet, indicando a eficácia do modelo.

A avaliação por cobertura de nuvens mostrou que o desempenho dos modelos diminui à medida que a proporção de regiões ocluídas aumenta. Esse comportamento reforça a natureza mal-posta da reconstrução a partir de uma única imagem óptica: em casos de oclusão severa, a informação da superfície pode estar fisicamente ausente, exigindo que o modelo infira padrões a partir do contexto espacial disponível. Ainda assim, a WAMG-UNet manteve desempenho competitivo em todas as faixas avaliadas, apresentando maior robustez relativa em relação aos baselines.

Do ponto de vista qualitativo, as reconstruções produzidas pela WAMG-UNet apresentaram maior preservação de estruturas locais, bordas e texturas em cenários de cobertura leve e moderada. Nos casos de oclusão mais intensa, observou-se aumento da incerteza visual e maior ocorrência de artefatos, comportamento esperado diante da ausência de informação observável na entrada. Esses resultados reforçam que a abordagem proposta é mais adequada para cenários

em que ainda existe contexto espacial suficiente para orientar a reconstrução.

Em relação ao custo computacional, a WAMG-UNet apresentou maior número de parâmetros e maior latência de inferência que os baselines U-Net. Esse aumento de custo é consequência direta dos componentes adicionais da arquitetura, como operações wavelet e mecanismos de modulação por máscara. Ainda assim, a latência observada permanece compatível com aplicações de processamento offline ou em lote, nas quais a qualidade da reconstrução pode ser priorizada em relação à velocidade de inferência.

Dessa forma, conclui-se que a WAMG-UNet constitui uma alternativa viável para a reconstrução de imagens ópticas de sensoriamento remoto afetadas por nuvens em um protocolo supervisionado com o dataset RICE. A arquitetura demonstrou ganhos consistentes em relação a baselines U-Net, especialmente pela integração entre máscara de nuvem, processamento multiescala e controle das conexões de salto. Embora a perda wavelet com Curriculum Learning tenha apresentado impacto quantitativo reduzido nas métricas globais, sua inclusão contribui para a formulação de uma abordagem orientada à preservação de detalhes, cuja avaliação pode ser aprofundada com métricas perceptuais e espectrais em trabalhos futuros.

## REFERÊNCIAS

- ANZALONE, A. *et al.* An Introduction to Machine and Deep Learning Methods for Cloud Masking Applications. Multidisciplinary Digital Publishing Institute, v. 14, n. 7, p. 2887, 2024. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/14/7/2887>.
- BENGIO, Y. *et al.* Curriculum learning. In: **Proceedings of the 26th Annual International Conference on Machine Learning**. Association for Computing Machinery, 2009. (ICML '09), p. 41–48. ISBN 978-1-60558-516-1. Disponível em: <https://doi.org/10.1145/1553374.1553380>.
- CHAN, Y. K.; KOO, V. C. AN INTRODUCTION TO SYNTHETIC APERTURE RADAR (SAR). **Progress In Electromagnetics Research B**, v. 2, p. 27–60, 2008. ISSN 1937-6472. Disponível em: <http://www.jpier.org/PIERB/pier.php?paper=07110101>.
- CHEN, H. *et al.* Attentive generative adversarial network for removing thin cloud from a single remote sensing image. v. 18, n. 11, p. 1901–1905, 2021. ISSN 1558-0571. Disponível em: <https://ieeexplore.ieee.org/document/9144461>.
- DONG, S. *et al.* A survey on deep learning and its applications. v. 40, p. 100379, 2021. ISSN 1574-0137. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574013721000198>.
- EDIRISINGHE, S. *et al.* Generative deep learning models for cloud removal in satellite imagery: A comparative review of GANs and diffusion methods. v. 19, p. 100110, 2026. ISSN 2667-3932. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2667393225000298>.
- ELACHI, C.; ZYL, J. van. **Physics and Techniques of Remote Sensing**. 3. ed. Hoboken, NJ: Wiley, 2021. (Wiley Series in Remote Sensing). Also available as e-book, ISBN: 9781119523123 (adobe pdf), 9781119523086 (epub). ISBN 9781119523017. Disponível em: <https://lccn.loc.gov/2020051941>.
- GOODFELLOW, I. *et al.* **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.
- HAN, S. *et al.* Former-cr: A transformer-based thick cloud removal method with optical and sar imagery. **Remote Sensing**, v. 15, n. 5, 2023. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/15/5/1196>.
- HE, K. *et al.* Single Image Haze Removal Using Dark Channel Prior. 2009.
- HE, K. *et al.* Deep Residual Learning for Image Recognition. In: . [s.n.], 2016. p. 770–778. Disponível em: [https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/He\\_Deep\\_Residual\\_Learning\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html).
- HORNIK, K. *et al.* Multilayer feedforward networks are universal approximators. v. 2, n. 5, p. 359–366, 1989. ISSN 0893-6080. Disponível em: <https://www.sciencedirect.com/science/article/pii/0893608089900208>.
- HU, J. *et al.* **Squeeze-and-Excitation Networks**. 2019. Disponível em: <http://arxiv.org/abs/1709.01507>.
- HUANG, W. *et al.* **Attentive Contextual Attention for Cloud Removal**. 2024. Disponível em: <https://arxiv.org/abs/2411.13042>.

ISOLA, P. *et al.* **Image-to-Image Translation with Conditional Adversarial Networks**. 2018. Disponível em: <http://arxiv.org/abs/1611.07004>.

JIANG, B. *et al.* FDT-Net: Deep-Learning Network for Thin-Cloud Removal in Remote Sensing Image Using Frequency-Domain Training Strategy. v. 21, p. 1–5, 2024. ISSN 1558-0571. Disponível em: <https://ieeexplore.ieee.org/document/10502157>.

JIANG, L. *et al.* **Focal Frequency Loss for Image Reconstruction and Synthesis**. 2021. Disponível em: <http://arxiv.org/abs/2012.12821>.

KELLY, S. E. Gibbs phenomenon for wavelets. **Applied and Computational Harmonic Analysis**, v. 3, n. 1, p. 72–81, 1996. ISSN 1063-5203. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1063520396900068>.

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. 2017. Disponível em: <https://arxiv.org/abs/1412.6980>.

KOKALY, R. F. *et al.* **USGS Spectral Library Version 7**. [S.l.], 2017. 61 p. Disponível em: <https://pubs.er.usgs.gov/publication/ds1035>.

LECUN, Y. *et al.* Gradient-Based Learning Applied to Document Recognition. v. 86, p. 2278–2324, 1998.

LI, C. *et al.* Transformer Meets GAN: Cloud-Free Multispectral Image Reconstruction via Multisensor Data Fusion in Satellite Images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 61, p. 1–13, 2023. ISSN 1558-0644.

LIN, D. *et al.* **A Remote Sensing Image Dataset for Cloud Removal**. 2019. Disponível em: <http://arxiv.org/abs/1901.00600>.

LIN, M. *et al.* **Network In Network**. 2014. Disponível em: <https://arxiv.org/abs/1312.4400>.

LIU, J. *et al.* SI-SA GAN: A Generative Adversarial Network Combined With Spatial Information and Self-Attention for Removing Thin Cloud in Optical Remote Sensing Images. v. 10, p. 114318–114330, 2022. ISSN 2169-3536. Disponível em: <https://ieeexplore.ieee.org/document/9915386>.

LIU, P. *et al.* **Multi-level Wavelet-CNN for Image Restoration**. 2018. Disponível em: <https://arxiv.org/abs/1805.07071>.

MA, J. *et al.* SCT-CR: A synergistic convolution-transformer modeling method using SAR-optical data fusion for cloud removal. v. 130, p. 103909, 2024. ISSN 1569-8432. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1569843224002632>.

MALLAT, S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. v. 11, n. 7, p. 674–693, 1989. ISSN 0162-8828. Disponível em: <https://doi.org/10.1109/34.192463>.

MCCULLOCH, W. S.; PITTS, W. A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY. 1943.

MINSKY, M.; PAPERT, S. **Perceptrons; an Introduction to Computational Geometry**. MIT Press, 1969. ISBN 978-0-262-63022-1. Disponível em: <https://books.google.com.br/books?id=Ow1OAQAIAAJ>.

MORALES-BROTONS, D. *et al.* **Exponential Moving Average of Weights in Deep Learning: Dynamics and Benefits**. 2024. Disponível em: <https://arxiv.org/abs/2411.18704>.

NING, J. *et al.* Cloud Removal Advances: A Comprehensive Review and Analysis for Optical Remote Sensing Images. v. 18, p. 15914–15930, 2025. ISSN 2151-1535. Disponível em: <https://ieeexplore.ieee.org/abstract/document/11039671>.

NOVO, E. M. L. d. M. **Sensoriamento remoto: princípios e aplicações**. 3. ed. São Paulo: Blucher, 2008. CDD-621.3678. ISBN 978-85-212-0441-1.

OTSU, N. A Threshold Selection Method from Gray-Level Histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62–66, 1979. ISSN 0018-9472, 2168-2909. Disponível em: <http://ieeexplore.ieee.org/document/4310076/>.

QIU, S. *et al.* Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. v. 231, p. 111205, 2019. ISSN 00344257. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0034425719302172>.

RAHAMAN, N. *et al.* **On the Spectral Bias of Neural Networks**. 2019. Disponível em: <http://arxiv.org/abs/1806.08734>.

RONNEBERGER, O. *et al.* **U-Net: Convolutional Networks for Biomedical Image Segmentation**. 2015. Disponível em: <http://arxiv.org/abs/1505.04597>.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *American Psychological Association*, v. 65, n. 6, p. 386–408, 1958. ISSN 1939-1471.

RUMELHART, D. E. *et al.* Learning representations by back-propagating errors. v. 323, n. 6088, p. 533–536, 1986. ISSN 0028-0836, 1476-4687. Disponível em: <https://www.nature.com/articles/323533a0>.

SCHOTT, J. R. **Remote Sensing: The Image Chain Approach**. Oxford University Press, 2007. ISBN 978-0-19-517817-3. Disponível em: <https://doi.org/10.1093/oso/9780195178173.001.0001>.

SHEN, H. *et al.* Remote sensing image dehazing using generative adversarial network with texture and color space enhancement. v. 14, n. 1, p. 12382, 2024. ISSN 2045-2322. Disponível em: <https://www.nature.com/articles/s41598-024-63259-6>.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. Disponível em: <https://www.semanticscholar.org/paper/Very-Deep-Convolutional-Networks-for-Large-Scale-Simonyan-Zisserman/eb42cf88027de515750f230b23b1a057dc782108>.

SINGH, R. *et al.* Cloud Detection Methods for Optical Satellite Imagery: A Comprehensive Review. *Multidisciplinary Digital Publishing Institute*, v. 5, n. 3, p. 27, 2025. ISSN 2673-7418. Disponível em: <https://www.mdpi.com/2673-7418/5/3/27>.

SOHL-DICKSTEIN, J. *et al.* Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In: **Proceedings of the 32nd International Conference on Machine Learning**. PMLR, 2015. p. 2256–2265. ISSN 1938-7228. Disponível em: <https://proceedings.mlr.press/v37/sohl-dickstein15.html>.

- SRIVASTAVA, N. *et al.* Dropout: A simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, v. 15, n. 56, p. 1929–1958, 2014. Disponível em: <http://jmlr.org/papers/v15/srivastava14a.html>.
- SURYA, S. R.; RAHIMAN, M. A. Cloud detection from satellite images based on Haar wavelet and clustering. In: **2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2)**. [s.n.], 2017. p. 163–167. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8067921>.
- VASWANI, A. *et al.* Attention is All you Need. In: **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Disponível em: [https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html).
- WANG, M. *et al.* IDF-CR: Iterative Diffusion Process for Divide-and-Conquer Cloud Removal in Remote-sensing Images. v. 62, p. 1–14, 2024. ISSN 0196-2892, 1558-0644. Disponível em: <http://arxiv.org/abs/2403.11870>.
- WANG, X. *et al.* **A Survey on Curriculum Learning**. 2021. Disponível em: <http://arxiv.org/abs/2010.13166>.
- WANG, Z. *et al.* Image quality assessment: From error visibility to structural similarity. v. 13, n. 4, p. 600–612, 2004. ISSN 1941-0042.
- WOO, S. *et al.* **CBAM: Convolutional Block Attention Module**. 2018. Disponível em: <http://arxiv.org/abs/1807.06521>.
- XU, F. *et al.* **GLF-CR: SAR-Enhanced Cloud Removal with Global-Local Fusion**. 2022. Disponível em: <http://arxiv.org/abs/2206.02850>.
- YU, Z. *et al.* **DC4CR: When Cloud Removal Meets Diffusion Control in Remote Sensing**. 2025. Disponível em: <http://arxiv.org/abs/2504.14785>.
- ZHAO, H. *et al.* **Loss Functions for Neural Networks for Image Processing**. 2018. Disponível em: <http://arxiv.org/abs/1511.08861>.
- ZHU, X. X. *et al.* Deep Learning Meets SAR: Concepts, models, pitfalls, and perspectives. v. 9, n. 4, p. 143–172, 2021. ISSN 2168-6831. Disponível em: <https://ieeexplore.ieee.org/document/9351574/>.
- ZHU, Z.; WOODCOCK, C. E. Object-based cloud and cloud shadow detection in Landsat imagery. v. 118, p. 83–94, 2012. ISSN 0034-4257. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425711003853>.
- ZI, Y. *et al.* Wavelet integrated convolutional neural network for thin cloud removal in remote sensing images. v. 15, n. 3, p. 781, 2023.
- ZOU, X. *et al.* DiffCR: A Fast Conditional Diffusion Framework for Cloud Removal From Optical Satellite Images. v. 62, p. 1–14, 2024. ISSN 1558-0644. Disponível em: <https://ieeexplore.ieee.org/abstract/document/10436560>.