

---

# Remoção de nuvens e reconstrução de imagens de satélite usando *deep learning*

---

Leandro Henrique Furtado Pinto Silva



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia  
2025



**Leandro Henrique Furtado Pinto Silva**

**Remoção de nuvens e reconstrução de imagens  
de satélite usando *deep learning***

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Dr. André Ricardo Backes

Coorientador: Dr. Mauricio Cunha Escarpinati

Coorientador: Dr. João Fernando Mari

Uberlândia

2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

S586  
2025

Silva, Leandro Henrique Furtado Pinto, 1991-  
Remoção de nuvens e reconstrução de imagens de  
satélite usando deep learning [recurso eletrônico] /  
Leandro Henrique Furtado Pinto Silva. - 2025.

Orientador: André Ricardo Backes.

Coorientador: Mauricio Cunha Escarpinati.

Coorientador: João Fernando Mari.

Tese (Doutorado) - Universidade Federal de Uberlândia,  
Pós-graduação em Ciência da Computação.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.te.2025.250>

Inclui bibliografia.

Inclui ilustrações.

1. Computação. I. Backes, André Ricardo, 1981-,  
(Orient.). II. Escarpinati, Mauricio Cunha, 1976-,  
(Coorient.). III. Mari, João Fernando, 1981-,  
(Coorient.). IV. Universidade Federal de Uberlândia.  
Pós-graduação em Ciência da Computação. V. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074





## UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Computação  
Av. João Naves de Ávila, 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpqfacom@ufu.br



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Tese, 07/2025, PPGCO				
Data:	15 de Abril de 2025	Hora de início:	14:00	Hora de encerramento:	17:40
Matrícula do Discente:	12113CCP008				
Nome do Discente:	Leandro Henrique Furtado Pinto Silva				
Título do Trabalho:	Remoção de nuvens e reconstrução de imagens de satélite usando deep learning				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	-----				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Mauricio Cunha Escarpinati - FACOM/UFU(Coorientador), João Fernando Mari - IEP/UFV(Coorientador), Henrique Coelho Fernandes - FACOM/UFU, Moacir Antonelli Ponti - ICMC/USP, João Batista Florindo - IMECC/UNICAMP, Kelen Cristiane Teixeira Vivaldini - DC/UFScar e André Ricardo Backes - DC/UFSCar, orientador do candidato.

Os examinadores participaram desde as seguintes localidades: André Ricardo Backes - São Carlos/SP, João Fernando Mari - Viçosa/MG, Moacir Antonelli Ponti - São Carlos/SP, João Batista Florindo - Campinas/SP, Kelen Cristiane Teixeira Vivaldini São Carlos/SP. Os outros membros da banca e o aluno participaram da cidade de Uberlândia.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. André Ricardo Backes, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir ao candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

#### Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **André Ricardo Backes, Usuário Externo**, em 16/04/2025, às 13:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **João Fernando Mari, Usuário Externo**, em 16/04/2025, às 14:11, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mauricio Cunha Escarpinati, Professor(a) do Magistério Superior**, em 16/04/2025, às 15:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Kelen Cristiane Teixeira Vivaldini, Usuário Externo**, em 16/04/2025, às 15:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **João Batista Florindo, Usuário Externo**, em 16/04/2025, às 20:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Moacir Antonelli Ponti, Usuário Externo**, em 17/04/2025, às 10:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Henrique Coelho Fernandes, Professor(a) do Magistério Superior**, em 18/04/2025, às 05:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **6236201** e o código CRC **FD5A769D**.

*Este trabalho é dedicado à minha mãe e minha avó (in memoriam) que me deram todo o amor do mundo.*



---

# Agradecimentos

Agradeço a Deus pelo dom da vida.

Agradeço à minha mãe, Iris de Fátima Furtado Pinto, que sempre me deu amor e apoio em todos os momentos da minha vida.

Agradeço aos meus orientadores, Professores Dr. André Ricardo Backes, Dr. Mauricio Cunha Escarpinati e Dr. João Fernando Mari, que com muita paciência me direcionam nesta pesquisa, além de me encantarem com os seus saberes.

Agradeço à Universidade Federal de Uberlândia e à Universidade Federal de Viçosa, as quais são Instituições que tenho imenso carinho e me proporcionaram incomensurável crescimento acadêmico e profissional.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



*“A educação é a arma mais poderosa que você pode usar para mudar o mundo.”*  
*(Nelson Mandela)*





---

# Resumo

O sensoriamento remoto tem se demonstrado uma importante ferramenta para aplicações fundamentais, como preservação ambiental, monitoramento e prevenção de desastres naturais, planejamento urbano e agricultura de precisão. Os satélites são fontes de aquisição de dados preponderantes para realização do sensoriamento remoto, em virtude da sua capacidade de oferecer imageamento em várias bandas espectrais e em alta qualidade. Entretanto, as imagens de satélite podem ser sensíveis a presença de artefatos, tais como nuvens, sombras e neve, que interferem nas aplicações mencionadas. Diante disso, este trabalho apresenta quatro principais contribuições. (i) Uma revisão sistemática da literatura, com estabelecimento de uma taxonomia, desafios, perspectivas e tendências para pesquisas relacionadas a tarefa de remoção de nuvens. (ii) Apresenta-se uma nova abordagem de difusão latente com a combinação de canal de atenção e bloco residual para remoção de nuvens, a qual possui entre 48% a 82% menos parâmetros que o estado da arte para a tarefa e preserva a qualidade da reconstrução em termos quantitativos e qualitativos. Além disso, (iii) investigou-se também como a interação entre otimizadores e funções de perda influenciam o processo de aprendizagem para reconstrução de áreas nubladas através de modelo de difusão latente. Por fim, (iv) apresenta-se também um estudo abrangente sobre o comportamento e impacto de diferentes arquiteturas e modelos de segmentação para um cenário de imagens de satélite com nuvens.

**Palavras-chave:** Imagens de Satélite. Modelos de Difusão. *Deep Learning*. Remoção de Nuvens.



---

# Abstract

Remote sensing has proven to be an essential tool for fundamental applications such as environmental preservation, monitoring and prevention of natural disasters, urban planning, and precision agriculture. Satellites are preponderant data acquisition sources for remote sensing due to their ability to offer high-quality images in several spectral bands. However, satellite images can be sensitive to artifacts, such as clouds, shadows, and snow, which interfere with the mentioned applications. Therefore, this work presents four main contributions. (i) A systematic literature review, establishing a taxonomy, challenges, perspectives, and trends for research related to cloud removal. (ii) This work presents a new latent diffusion approach with residual blocks and combined attention channels for cloud removal, reducing the number of parameters by 48% to 82% compared to the state-of-the-art while preserving reconstruction quality in quantitative and qualitative terms. In addition, (iii) investigated how the interaction between optimizers and loss functions influences the learning process for reconstructing cloudy areas through a latent diffusion model. Finally, (iv) this work presents a comprehensive study of the behavior and impact of different architectures and segmentation models for a satellite imagery scenario with clouds..

**Keywords:** Satellite Images. Diffusion Models. Deep Learning. Cloud Removal.



---

## Lista de ilustrações

Figura 1 – Exemplos de imagens de satélite com a presença de nuvens. . . . .	28
Figura 2 – Exemplo das operações de uma CNN para uma tarefa de classificação. .	36
Figura 3 – Estrutura de um modelo de difusão latente. . . . .	41
Figura 4 – Estrutura básica de métodos baseados em <i>deep learning</i> para segmentação de nuvens. . . . .	43
Figura 5 – Representação da Fronteira de Pareto como <i>trade-off</i> entre custo computacional e métrica de avaliação. . . . .	44
Figura 6 – Visão geral do protocolo de revisão deste trabalho. . . . .	49
Figura 7 – Distribuição de trabalhos considerando ano e técnicas usadas. . . . .	51
Figura 8 – Distribuição de trabalhos considerando técnicas e tipo de nuvens abordadas. . . . .	52
Figura 9 – Distribuição dos artigos considerando múltiplos parâmetros. . . . .	54
Figura 10 – Relacionamento das métricas com os trabalhos considerados. . . . .	57
Figura 11 – Utilização combinada de diferentes métricas ao longo dos anos para os artigos selecionados. . . . .	58
Figura 12 – Exemplos de cenas completas do conjunto de dados 38-Cloud. . . . .	62
Figura 13 – Exemplos de imagens em forma de <i>patch</i> do conjunto de treinamento do 38-Cloud. . . . .	62
Figura 14 – Configuração Experimental para avaliação da segmentação de nuvens. .	64
Figura 15 – Avaliação de segmentação de uma cena retirados do conjunto de teste.	67
Figura 16 – Avaliação de segmentação da região de interesse de uma cena, retirados do conjunto de teste. . . . .	68
Figura 17 – Amostras da base de dados multitemporal SEN2_MTC_New nas bandas RGB. . . . .	70
Figura 18 – Configuração experimental proposta para análise da combinação de otimizadores e função de perda. . . . .	72
Figura 19 – Validação dos modelos a cada 200 épocas para análise da combinação de otimizadores e função de perda. . . . .	73

Figura 20 – Visualização e comparação dos resultados das métricas SSIM e PSNR no conjunto de teste para cada modelo. . . . .	75
Figura 21 – Resultados qualitativos para análise da combinação de otimizadores e funções de perda. . . . .	76
Figura 22 – Resultados qualitativos para análise da combinação de otimizadores e funções de perda. . . . .	77
Figura 23 – Esquema dos blocos propostos para o modelo de <i>denoising</i> . . . . .	81
Figura 24 – Blocos CAB e CABRes propostos. . . . .	82
Figura 25 – Abordagem proposta. . . . .	82
Figura 26 – Validação dos modelos a cada 200 épocas para as abordagens propostas de remoção de nuvens. . . . .	85
Figura 27 – Resultados qualitativos considerando áreas florestais, estradas e lagos. .	86
Figura 28 – Resultados qualitativos considerando regiões montanhosas e áreas de floresta. . . . .	87
Figura 29 – Resultados qualitativos considerando áreas com terra, áreas plantadas e estradas. . . . .	88
Figura 30 – Resultados da avaliação com a Fronteira de Pareto usando métricas PSNR e SSIM. Resultados da avaliação com a Fronteira de Pareto usando métricas PSNR e SSIM. . . . .	89
Figura 31 – Resultados da avaliação com a Fronteira de Pareto usando métricas FID e LPIPS . . . . .	89
Figura 32 – Comparação de métodos propostos para remoção de nuvens. . . . .	90

---

## Lista de tabelas

Tabela 1	– Bandas do Sentinel-2 . . . . .	35
Tabela 2	– Bandas do Landsat 8 . . . . .	35
Tabela 3	– Bases de Dados para Remoção de Nuvens . . . . .	56
Tabela 4	– Resultados experimental para segmentação conjunto de teste. . . . .	66
Tabela 5	– Resultados experimentais para combinação de otimizadores e função de perda no conjunto de teste. . . . .	74
Tabela 6	– Comparação de resultados entre as abordagens propostas. . . . .	85
Tabela 7	– Resultados do DiffCR. . . . .	88
Tabela 8	– Resumo dos modelos presentes na Fronteira de Pareto para as métricas consideradas. . . . .	88
Tabela 9	– Comparação dos resultados entre as abordagens propostas após a otimização baseada na Fronteira de Pareto. . . . .	91
Tabela 10	– Comparação de resultados entre experimentos da literatura e abordagens selecionadas dos experimentos após a otimização baseada na Fronteira de Pareto. . . . .	91
Tabela 11	– Resultados experimentais para 32 e 128 canais para os modelos otimizados com a Fronteira de Pareto. . . . .	93





---

## Lista de algoritmos

1	Algoritmo para estabelecer a Fronteira de Pareto . . . . .	45
---	--	----



---

## Lista de siglas

**Acc** Acurácia

**ACM** *Association for Computing Machinery*

**CAB** *Channel Attention Block*

**CABRes** *Channel Attention with Residual Block*

**CC** *Correlation Coefficient*

**CNN** *Convolutional Neural Networks*

**DCP** *Dark Channel Prior*

**Dice** Coeficiente de Dice

**ESA** *European Space Agency*

**EMA** *Exponential Moving Average*

**FID** *Fréchet Inception Distance*

**FN** Falsos Negativos

**FP** Falsos Positivos

**GAN** *Generative Adversarial Network*

**GPU** *Graphics Processing Unit*

**IEEE** *The Institute of Electrical and Electronics Engineers*

**IoT** *Internet of Things*

**IoU** *Intersection over Union*

**LPIPS** *Learned Perceptual Image Patch Similarity*

**MACs** *Multiply-Accumulate Operations*

**MAE** *Mean Absolute Error*

**MSE** *Mean-Square Error*

**MDPI** *Multidisciplinary Digital Publishing Institute*

**NASA** *National Aeronautics and Space Administration*

**NIR** *Near-Infrared*

**nNIR** *Narrow NIR*

**NDVI** *Normalized Difference Vegetation Index*

**PIB** *Produto Interno Bruto*

**PRISMA** *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*

**PSNR** *Peak Signal-to-Noise Ratio*

**RE1** *Vegetation Red Edge*

**RE2** *Vegetation Red Edge*

**RE3** *Vegetation Red Edge*

**ReLU** *Rectified Linear Unit*

**ResNet** *Residual Network*

**RMSE** *Root-Mean-Square Error*

**SAM** *Similarity Assessment Metric*

**SAR** *Synthetic-aperture radar*

**SGD** *Stochastic Gradient Descent*

**SMP** *Segmentation PyTorch Library*

**SiLU** *Sigmoid Linear Unit*

**SSIM** *Structural Similarity Index Measure*

**SWIR1** *Shortwave Infrared*

**SWIR2** *Shortwave Infrared*

**TN** Verdadeiros Negativos

**TP** Verdadeiros Positivos

**USGS** *United States Geological Survey*

**VGG** *Visual Geometry Group*

---

# Sumário

1	INTRODUÇÃO . . . . .	27
1.1	Motivação . . . . .	28
1.2	Hipóteses . . . . .	29
1.3	Objetivos . . . . .	29
1.4	Contribuições . . . . .	30
1.5	Organização da Tese . . . . .	30
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	33
2.1	Sensoriamento Remoto . . . . .	33
2.2	Imagens de Satélite . . . . .	34
2.3	Redes Neurais Convolucionais . . . . .	36
2.3.1	Mecanismos de Atenção . . . . .	37
2.3.2	Funções de Perda . . . . .	38
2.3.3	Otimizadores . . . . .	39
2.4	Modelos de Difusão Latente . . . . .	40
2.5	Reconstrução de Imagens . . . . .	41
2.6	Segmentação de Imagens . . . . .	42
2.7	Fronteira de Pareto . . . . .	43
3	TRABALHOS RELACIONADOS . . . . .	47
3.1	Levantamento da Literatura . . . . .	48
3.2	Tipo de Remoção de Nuvens . . . . .	51
3.3	Número de Bandas Consideradas . . . . .	52
3.4	Integração com SAR . . . . .	53
3.5	Técnicas Emergentes . . . . .	54
3.6	Bases de Dados . . . . .	55
3.7	Métricas de Avaliação . . . . .	57

4	AVALIAÇÃO DE ARQUITETURAS E <i>ENCODERS</i> PARA SEGMENTAÇÃO DE NUVENS . . . . .	61
4.1	Base de Dados . . . . .	61
4.2	Arquiteturas e <i>Encoders</i> . . . . .	62
4.3	Configuração Experimental . . . . .	64
4.4	Resultados . . . . .	65
4.5	Considerações . . . . .	67
5	COMBINAÇÕES DE OTIMIZADORES E FUNÇÕES DE PERDA PARA REMOÇÃO DE NUVENS EM MODELOS DE DIFU- SÃO . . . . .	69
5.1	Base de Dados . . . . .	69
5.2	DiffCR . . . . .	71
5.3	Configuração Experimental . . . . .	71
5.4	Resultados . . . . .	72
5.5	Considerações . . . . .	75
6	MODELOS DE DIFUSÃO COMPACTOS PARA REMOÇÃO DE NUVENS . . . . .	79
6.1	Abordagem Proposta . . . . .	79
6.2	Configuração Experimental . . . . .	81
6.3	Resultados . . . . .	84
6.4	Considerações . . . . .	93
7	CONCLUSÃO . . . . .	95
7.1	Principais Contribuições . . . . .	96
7.2	Trabalhos Futuros . . . . .	97
	REFERÊNCIAS . . . . .	99



## Introdução

O sensoriamento remoto é um importante mecanismo para auxiliar em diversas aplicações, tais como, planejamento urbano, agricultura de precisão e monitoramento ambiental, em virtude da capacidade de extrair informações de uma área ou objeto sem a necessidade de contato físico (GHILDIYAL et al., 2024). Diversas fontes de aquisição, tais como satélites, drones e aviões, possibilitam realizar o sensoriamento remoto por meio de diferentes sensores.

Especificamente para os satélites, é possível considerar diversos sensores para realizar imageamento multi e hiperespectral. Entretanto, apesar deles possibilitarem um imageamento em alta resolução, muitos sensores são sensíveis à presença de artefatos, como nuvens, o que consiste em um desafio considerável para obtenção de índices acurados para as diferentes aplicações (JEPPESEN et al., 2019; ARAKAKI et al., 2023; FERREIRA et al., 2024). Por exemplo, na agricultura, a presença de nuvens dificulta a extração de índices de vegetação, como o *Normalized Difference Vegetation Index* (NDVI), e consequentemente impacta na estimativa de utilização de insumos agrícolas e na safra em geral (ABBASI; MARTINEZ; AHMAD, 2022). Para o monitoramento ambiental, essas oclusões podem dificultar a identificação de estradas e desmatamentos ilegais, o que impacta na fiscalização (FARIA et al., 2023). Já no planejamento urbano, tais artefatos podem afetar a identificação de loteamentos e construções irregulares, bem como atrapalhar a prevenção de desastres ambientais. Como pode ser visualizado na Figura 1, exemplos de imagens com a presença de nuvens, obtidas pelo satélite Sentinel-2, dificultam a identificação de diferentes coberturas de solo.

Na Figura 1 é possível ainda notar que as nuvens possuem dois tipos principais, espessas e finas. As nuvens espessas possuem um padrão mais definido e são mais fáceis de delinear, entretanto, carregam pouca ou nenhuma informação de solo abaixo dela. Já as nuvens finas, por mais que carreguem informações de transparência de solo, são complexas de segmentar e consequentemente dificultam um modelo computacional em reconstruir a área exata que estava obstruída, preservando eventuais áreas livres nuvens (HE; SUN; TANG, 2010). Ainda há de se mencionar que as nuvens geram sombras, o que também

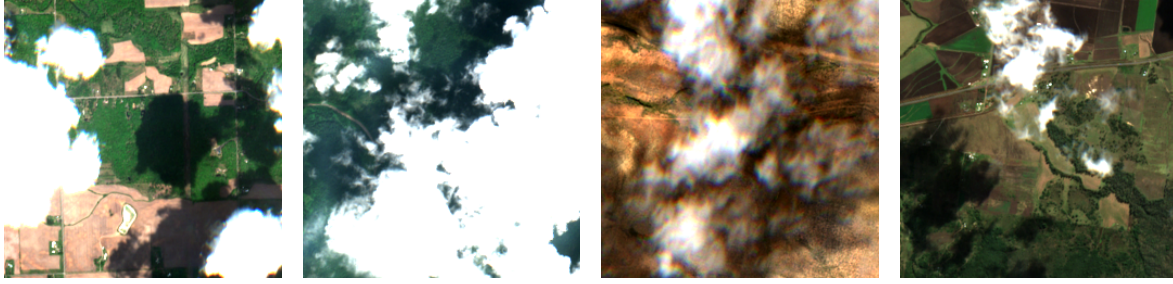


Figura 1 – Exemplos de imagens de satélite com a presença de nuvens.

impacta na análise da imagem, bem como na reconstrução, sendo, portanto, um desafio inerente a se considerar (GUO et al., 2020a; GUO et al., 2023).

Grandes esforços têm sido dedicados ao processamento digital de imagens para mitigar artefatos e aprimorar a qualidade das análises realizadas a partir de imagens de satélite. Na última década, pesquisas na área passaram a explorar consideravelmente técnicas baseadas em *deep learning* para resolver problemas diversos e atender a múltiplas aplicações (DONG; WANG; ABBAS, 2021; KAMILARIS; PRENAFETA-BOLDÚ, 2018; LITJENS et al., 2017; PICCIALLI et al., 2021). A incorporação dessas técnicas gerou um considerável impacto no sensoriamento remoto, impulsionando investigações sobre imagens de satélite e suas aplicações (SEGAL-ROZENHAIMER et al., 2020; GUO et al., 2020b).

Ainda assim, lacunas importantes persistem, especialmente relacionadas à remoção de nuvens, um desafio preponderante para a integridade e consequente utilização dessas imagens (BAI et al., 2016; JEPPESEN et al., 2019; BAI et al., 2016). Dentre essas lacunas, destacam-se a dificuldade de generalização dos modelos para diferentes sensores e condições atmosféricas, a preservação de detalhes e texturas nas regiões reconstruídas, a escassez de dados anotados para treinamento supervisionado, a falta de coerência espectral e radiométrica nas reconstruções, o alto custo computacional dos métodos e a limitação das métricas utilizadas para avaliação da qualidade das imagens processadas (XIE et al., 2023; STOCK et al., 2020; PODSIADLO; PARIS; BRUZZONE, 2020).

## 1.1 Motivação

O sensoriamento remoto por meio de satélites consiste em um dos principais mecanismos de monitoramento ambiental, análise de mudanças na cobertura do solo, agricultura de precisão e gestão de desastres naturais (MA et al., 2023). Entretanto, mais de 50% da superfície terrestre fica obstruída por nuvens espessas, além de outros 20% estarem cobertas por nuvens finas. Desse modo, em torno de 70% da Terra possui a presença de nuvens (ZHAO; OLSEN; CHANDRA, 2023). Essas nuvens se deslocam de forma ininterrupta ao redor do globo, o que gera aspectos intermitentes e sazonais na aquisição direta de imagens livres de nuvens (WYLIE; MENZEL, 1989; WYLIE et al., 2005; STUBENRAUCH

et al., 2013). Assim, soluções computacionais que consigam identificar e remover nuvens serão importantes para auxiliar diferentes aplicações,

A remoção eficaz de nuvens permite o acesso contínuo a imagens de satélite de alta qualidade, melhorando a confiabilidade das análises e tomadas de decisão baseadas nestes dados. Diante disso, as pesquisas têm se concentrado em diferentes métodos computacionais, incluindo *Convolutional Neural Networks* (CNN), *Generative Adversarial Network* (GAN), fusão de dados e, mais recente, modelos difusão latente para realizar a tarefa de remoção de nuvens (DONG; WANG; ABBAS, 2021). Como muitas vezes a disponibilidade de *hardware* é um fator limitador para a execução de modelos complexos de processamento de imagens, é fundamental investigar soluções que equilibrem eficiência computacional e desempenho na remoção de nuvens.

## 1.2 Hipóteses

Métodos de remoção de nuvens que utilizam arquiteturas mais eficientes computacionalmente podem alcançar um bom equilíbrio entre desempenho e custo computacional, tornando-se viáveis mesmo em cenários com restrições de *hardware*; além disso, a escolha do otimizador combinada à função de perda afeta significativamente a convergência e a qualidade da reconstrução, podendo melhorar a estabilidade do treinamento e reduzir artefatos. Os modelos de *deep learning* treinados com funções de perda adaptadas para priorizar regiões com nuvens apresentam melhor desempenho na remoção de nuvens em imagens de satélite, preservando melhor a qualidade visual, principalmente em cenários com diferentes coberturas de solo e tipos de nuvens; por fim, a escolha da arquitetura de segmentação e do *encoder* impacta significativamente a precisão da segmentação de nuvens.

## 1.3 Objetivos

Os objetivos deste trabalho consistem em avaliar diferentes métodos baseados em *deep learning* para remoção e segmentação de nuvens, além de apresentar uma nova abordagem capaz de remover nuvens de forma eficaz, preservando a qualidade visual das imagens. Busca-se, ainda, comparar o impacto de diferentes funções de perda e otimizadores no desempenho do modelo, fornecendo uma análise detalhada sobre as melhores configurações para essa tarefa.

Mais especificamente:

- Apresentar uma nova abordagem para remoção de nuvens, buscando um equilíbrio entre desempenho e utilização de recursos computacionais.

- ❑ Investigar a relação entre otimizadores e funções de perda na convergência e estabilidade do treinamento de modelos de difusão latente, identificando combinações que minimizam artefatos e melhoram a qualidade da reconstrução das imagens nubladas.
- ❑ Comparar o desempenho de diferentes funções de perda adaptadas para remoção de nuvens, analisando sua influência na preservação da qualidade visual e espectral das imagens, especialmente em cenários com variadas coberturas de solo e tipos de nuvens.
- ❑ Avaliar o impacto de combinações entre arquiteturas de segmentação e *encoders* na qualidade da segmentação de nuvens, investigando a influência neste processo.

## 1.4 Contribuições

Este trabalho traz as seguintes contribuições:

- ❑ Uma nova abordagem baseada em difusão latente que seja capaz de reconstruir áreas nubladas com desempenho semelhante à literatura, mas com um menor custo computacional.
- ❑ Uma revisão sistemática da literatura, com a proposição de uma taxonomia para remoção de nuvens, além da apresentação de tendências, desafios e perspectivas da área.
- ❑ Uma avaliação do impacto de diferentes combinações de otimizadores e funções de perda em modelo de difusão latente para remoção de nuvens.
- ❑ Uma análise abrangente do comportamento da combinação de diferentes arquiteturas e *encoders* para especificamente a tarefa de segmentação de nuvens, sob o ponto de vista de desempenho e custo computacional.

## 1.5 Organização da Tese

Esta Tese está organizada da seguinte forma: no Capítulo 2 são percorridos conceitos estabelecidos na literatura fundamentais para este trabalho. No Capítulo 3 é apresentada revisão sistemática de remoção de nuvens em imagens de satélite, de modo a estabelecer uma taxonomia para a tarefa. No Capítulo 4, diferentes configurações de redes são avaliadas para segmentação de nuvens. No Capítulo 5, são abordadas diferentes combinações de otimizadores e funções de perda que impactam modelos de difusão latente para remoção de nuvens. No Capítulo 6, uma nova abordagem baseada em difusão latente é proposta para remoção de nuvens capaz de realizar a tarefa com desempenho similar à literatura

---

com um custo computacional reduzido. Por fim, no Capítulo 7, tem-se as conclusões, limitações, perspectivas e trabalhos futuros à esta Tese.



---

## Fundamentação Teórica

### 2.1 Sensoriamento Remoto

O sensoriamento remoto pode ser definido como a obtenção de informações sobre uma determinada área ou objeto sem a necessidade de contato físico (LILLESAND; KIEFER; CHIPMAN, 2015). Para realizar o sensoriamento utiliza-se radiação eletromagnética, a qual irá sintetizar dados (incluindo imagens) de uma área analisada (NOGUEIRA; PENATTI; SANTOS, 2017). Tais dados podem ser dos mais diversos tipos, por exemplo, temperatura, umidade, pressão, índices de solo e vegetação. Assim, há de se destacar os dados obtidos através de imagens com natureza multi ou hiper espectral, que possibilitam diferentes tipos análises em diversas aplicações (TOTH; JÓZKÓW, 2016).

Para utilizar imagens de sensoriamento remoto em tomadas de decisão é necessário considerar alguns aspectos fundamentais, visto que diversos parâmetros são afetados pelo tipo de plataforma (satélite, aérea ou terrestre) e pelo sensor escolhido para a aquisição dos dados (KHANAL; FULTON; SHEARER, 2017). Cada um dos aspectos em questão pode ser sintetizado em suas respectivas características:

- ❑ **Precisão geométrica:** posicionamento do sensor e ângulo de visão.
- ❑ **Resolução radiométrica:** número de *bits* utilizados por sensor.
- ❑ **Resolução espectral:** bandas do espectro eletromagnético e intervalo de comprimento da onda.
- ❑ **Resolução espacial:** expressa pelo tamanho real da área em relação a cada pixel imageado.
- ❑ **Resolução temporal:** frequência de captura de imagens.
- ❑ **Informação espectral:** posição do sol, presença de nuvens e absorção atmosférica.

## 2.2 Imagens de Satélite

A utilização de imagens de satélite consiste em uma importante ferramenta na observação e compreensão da superfície da Terra (PAHLEVAN et al., 2019). Essa importância se deve ao fato de haver um fornecimento de dados contínuo e de larga escala, o que permite monitorar fenômenos ambientais, gerir recursos naturais e analisar mudanças inerentes ao tempo (ZHU et al., 2017; HEDLEY et al., 2018; DELLOYE; WEISS; DEFOURNY, 2018).

Para o monitoramento ambiental, esse tipo de imageamento é importante para acompanhar processos naturais e impactos da atividade humana no meio ambiente (LACROIX et al., 2018). Os sensores multi e hiperespetrais possibilitam monitorar a cobertura vegetal, detectar áreas desmatadas, avaliar qualidade da água para grandes extensões territoriais, além de proporcionarem acompanhamento de fenômenos climáticos severos, tais como secas, enchentes, incêndios e nevascas (ARNAUDO et al., 2023).

Outro setor fortemente impactado pela utilização de satélites é a agricultura, setor consolidado como o mais importante para o Produto Interno Bruto (PIB) do Brasil<sup>1</sup> nos últimos anos (PELLEGRINA, 2022). Em termos gerais, o agronegócio é potencializado pelas imagens de satélite através da possibilidade de uma gestão eficiente das lavouras, monitoramento da saúde das plantas, detecção de pragas e estimativas de produtividades (AKHTER; SOFI, 2022). Esses benefícios são possíveis através da identificação de variabilidades espaciais, o que permite tomadas de decisões baseadas em dados (KHANAL; FULTON; SHEARER, 2017).

O planejamento urbano também pode se beneficiar com o imageamento por satélites, visto que, por meio da detecção de mudanças temporais, é possível identificar transformações no uso e cobertura de solo (COUTTS et al., 2016). Esse tipo de análise auxilia estudos relacionados a degradação ambiental e planejamento territorial, de forma a proporcionar, por exemplo, a identificação e avaliação do impacto de construções irregulares e em áreas de preservação permanente.

Atualmente, diversos satélites<sup>2</sup> estão na órbita terrestre, os quais possibilitam realizar as atividades mencionadas. Entretanto, dois deles são comumente utilizados em pesquisas científicas, principalmente devido à possibilidade de obtenção gratuita dos dados, sendo: Sentinel-2<sup>3</sup> e o Landsat 8<sup>4</sup>. Ambos os satélites possibilitam imageamento multiespectral e em alta resolução. Entretanto, cada um possui suas respectivas características (ASTOLA et al., 2019; CHASTAIN et al., 2019).

O Sentinel-2 é operado pela *European Space Agency* (ESA)<sup>5</sup> e foi lançado em 2015. Em termos gerais esse satélite possui uma resolução espacial de 10m, 20m e 60m, com resolução

<sup>1</sup> <<https://www.ibge.gov.br/explica/pib.php>>

<sup>2</sup> <<https://directory.eoportal.org/>>

<sup>3</sup> <[https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2)>

<sup>4</sup> <<https://landsat.gsfc.nasa.gov/satellites/landsat-8/>>

<sup>5</sup> <<https://www.esa.int/>>



temporal de 5 dias e em 13 bandas (ZHANG et al., 2018). Devido às suas características, esse satélite apresenta maior resolução espacial nos espectros visíveis e Infravermelho Próximo o que possibilita aplicações em cenários relacionados à agricultura, gestão de recursos hídricos e monitoramento ambiental. Um resumo das bandas e das resoluções presentes no Sentinel-2 pode ser observado na Tabela 1.

Tabela 1 – Bandas do Sentinel-2

Banda	Descrição	Comprimento de Onda (nm)	Resolução (m)
1	Aerossol Costeiro	433–453	60
2	Azul	458–523	10
3	Verde	543–578	10
4	Vermelho	650–680	10
5	<i>Vegetation Red Edge</i> (RE1)	698–713	20
6	<i>Vegetation Red Edge</i> (RE2)	733–748	20
7	<i>Vegetation Red Edge</i> (RE3)	773–793	20
8	<i>Near-Infrared</i> (NIR)	785–900	10
8a	<i>Narrow NIR</i> (nNIR)	855–875	20
9	Vapor d'água	935–955	60
10	Infravermelho de Ondas Curtas - Cirrus	1360–1390	60
11	<i>Shortwave Infrared</i> (SWIR1)	1565–1655	20
12	<i>Shortwave Infrared</i> (SWIR2)	2100–2280	20

O Landsat 8 é operado pela *National Aeronautics and Space Administration* (NASA)<sup>6</sup> e *United States Geological Survey* (USGS)<sup>7</sup> e foi lançado em 2013 (PAHLEVAN et al., 2022). Ele possui uma maior abrangência histórica devido ao seu lançamento preceder ao Sentinel-2, o que pode ser preponderante para aplicações de mapeamento terrestre e estudos climáticos (KUMAR et al., 2022). Possui 9 bandas espectrais a uma resolução espacial de 30m, além de uma resolução temporal de 16 dias. As bandas do Landsat 8 e suas respectivas resoluções são sumarizadas na Tabela 2.

Tabela 2 – Bandas do Landsat 8

Banda	Descrição	Comprimento de Onda (nm)	Resolução (m)
1	Aerossol costeiro	433–453	30
2	Azul	450–515	30
3	Verde	525–600	30
4	Vermelho	630–680	30
5	NIR	845–885	30
6	SWIR1	1560–1660	30
7	SWIR2	2100–2300	30
8	Pancromática	500–680	30
9	Infravermelho de Ondas Curtas - Cirrus	1360–1390	30

<sup>6</sup> <https://www.nasa.gov/>

<sup>7</sup> <https://www.usgs.gov/>

## 2.3 Redes Neurais Convolucionais

Métodos baseados em *deep learning*, um ramo do aprendizado de máquina que utiliza redes neurais profundas para modelar e extrair padrões complexos a partir de grandes volumes de dados, têm como objetivo encontrar modelos paramétricos adequados para resolver problemas específicos. Em termos gerais, o processo de ajuste dos parâmetros ocorre através de exemplos que orientam e refinam o aprendizado do modelo, para que possa posteriormente atuar em dados ainda não apresentados (GOODFELLOW et al., 2016).

As CNN consistem em um dos modelos de *deep learning* mais utilizados, especialmente em problemas relacionados a imagens. A característica de maior destaque em uma CNN é alusiva ao seu nome: ser composta por camadas convolucionais (LECUN et al., 1989). Tais camadas são responsáveis por processar as entradas, de modo a basear-se nos campos receptivos locais, os quais irão explorar correlações espaciais através da conectividade entre pixels (GOODFELLOW et al., 2016). Além da camada convolucional, as operações de *pooling* também são comumente observadas em uma CNN e reduzem a dimensionalidade espacial das imagens, de modo a representar as principais características obtidas nas operações de convolução. Após sucessivas operações de convolução, *pooling* e funções de ativação, temos uma camada densa, sendo, em linha geral, uma rede neural multicamada convencional (PONTI et al., 2017), conforme a Figura 2.

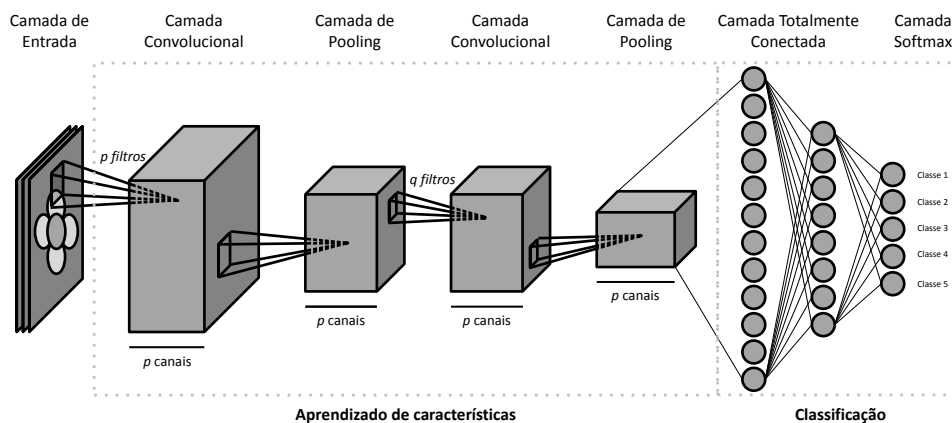


Figura 2 – Exemplo das operações de uma CNN para uma tarefa de classificação. Adaptado de Mari (2023).

Há diversas arquiteturas baseadas em CNN, com suas devidas particularidades, entretanto, podemos destacar LeNet-5 (LECUN et al., 1998), AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; KRIZHEVSKY; SUTSKEVER; HINTON, 2017), *Visual Geometry Group* (VGG) (SIMONYAN; ZISSERMAN, 2014), Inception (SZEGEDY et

al., 2015) e as *Residual Network* (ResNet) (HE et al., 2016). Apesar de aplicáveis em outras áreas, como classificação de pacotes de redes (MOREIRA et al., 2020) e detecção de similaridade de código-fonte (MARASTONI; GIACOBazzi; PREDA, 2018), a principal aplicação de CNNs consiste no processamento de imagens. Tal aplicabilidade acontece pois as operações de convoluções possibilitam filtrar imagens considerando sua estrutura espacial, bi ou multidimensional (LECUN; BENGIO; HINTON, 2015; GUO et al., 2016).

As redes residuais, especificamente, foram preponderantes para o treinamento de redes neurais profundas, visto que auxiliam a evitar o problema da atenuação do gradiente, através de conexões residuais. A utilização de blocos de residuais em uma rede neural facilita a propagação de características e gradientes ao incorporar a diferença entre a saída desejada e a entrada diretamente à entrada original. Pode-se definir os blocos residuais conforme a Equação 1:

$$y = F(x, \{W_i\}) + x, \quad (1)$$

onde  $y$  é a saída do bloco;  $x$  é a entrada original da camada;  $F(x, \{W_i\})$  é função de transformação realizada;  $\{W_i\}$  consiste nos pesos da camada.

### 2.3.1 Mecanismos de Atenção

Os mecanismos de atenção consistem em um componente preponderante para o treinamento de redes neurais profundos, pois possibilitam que os modelos foquem em regiões ou características mais relevantes dos dados de entrada para realizar uma determinada predição. Nesse sentido, ao invés de tratar todas as informações de maneira uniforme, a atenção atribui pesos diferentes para cada parte da entrada a ser processada (NIU; ZHONG; YU, 2021). Na literatura, há variações dos mecanismos de atenção, sendo cada uma projetada para atender diferentes necessidades computacionais e estruturais.

A atenção global é uma técnica que examina todas as posições da entrada de forma simultânea, atribuindo pesos a cada elemento em relação aos outros. Esse tipo de atenção é particularmente utilizado em modelos do tipo *Transformer*, que processa sequências de dados de forma eficiente e paralela. A atenção local, por outro lado, restringe o campo de atenção para uma janela fixa ao redor de cada posição. Esse tipo de mecanismo é particularmente útil em ambientes com recursos computacionais reduzidos. O *self-attention* é outro mecanismo de atenção onde cada posição de uma determinada sequência examinará todas as outras posições, de forma a aprender relações internas (WANG et al., 2020b). Por fim, o *cross-attention* calcula a relação entre duas entradas diferentes e de tipos diferentes, sendo comumente utilizado em tarefas de tradução automática e quando há a mistura de dados de tipo texto e imagem (CAI; WEI, 2020).

Especificamente para tarefas de visão computacional, uma abordagem alternativa consiste na atenção às características extraídas pelas convoluções. Essa abordagem aprende

a destacar canais ou regiões espaciais de interesse, a fim de aprimorar a capacidade do modelo em realizar a identificação dos detalhes mais importantes para a tarefa a ser realizada. Uma das maneiras de executar essa abordagem é através da ponderação da contribuição de cada mapa de características, de forma a permitir que o modelo destaque os canais com informações mais preponderantes (ZHANG et al., 2021). O método *Squeeze and Excitation* possui as características mencionadas em três principais etapas: (i) redução da dimensão espacial através de uma operação de *pooling* global, (ii) aplicação de camadas totalmente conectadas para aprimorar o processo de aprendizagem da importância de cada canal e (iii) a recalibração através da multiplicação dos pesos aprendidos pelos canais originais, gerando uma ênfase nas características mais relevantes (HU; SHEN; SUN, 2018).

### 2.3.2 Funções de Perda

Outro componente importante para o treinamento de modelos é a função de perda. A escolha da função de perda é preponderante para que o modelo tenha a capacidade de aprender e conseguir extrapolar conhecimento (WANG et al., 2020a). Basicamente, a função de perda penaliza o modelo na atualização dos pesos durante o processo de treinamento e, a depender da função escolhida, o erro é tratado de diferentes formas (TIAN et al., 2022). Uma das funções de perda é o *Mean Absolute Error* (MAE), que calcula a diferença absoluta entre a diferença entre o valor predito e o esperado. Outra função de perda é o erro médio quadrático *Mean-Square Error* (MSE), o qual é uma variação o MAE através da utilização do cálculo da raiz quadrada da diferença entre o previsto e o esperado. As Equações 2 e 3 apresentam as relações mencionadas.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |x_i - y_i|, \quad (2)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2, \quad (3)$$

onde  $x_i$  é o valor predito,  $y_i$  é o valor esperado, e  $N$  é o número de elementos considerados.

Além das funções de perda mencionadas, há também formas de combinar as funções, de forma a extrair o potencial máximo de cada uma delas. Nesse sentido, a função Huber possui aspectos de MAE e MSE, onde a definição de qual função a ser utilizada é guiada por um limiar  $\delta$ . Essa estratégia permite penalizar erros maiores enquanto suaviza erros menores. A função Huber é definida de acordo com a Equação 4.

$$\text{Huber} = \begin{cases} 0.5 \cdot (x_i - y_i)^2, & \text{se } |x_i - y_i| < \delta \\ \delta \cdot (|x_i - y_i| - 0.5 \cdot \delta), & \text{outro caso} \end{cases} \quad (4)$$

Outra função de perda considerada, principalmente em tarefas discretas de classificação e segmentação, é a Entropia Cruzada ( $\ell$ ), que é definida conforme a Equação 5.

$$\ell = - \sum_{n=1}^N \sum_{c=1}^C y_{n,c} \log p_{n,c} \quad (5)$$

onde  $N$  é o número de amostras no lote,  $C$  é o número de classes,  $y_{n,c}$  é a matriz de rótulos; em que  $y_{n,c} = 1$  se a amostra  $n$  pertence à classe  $c$ , e 0 caso contrário; por fim,  $p_{n,c}$  é a probabilidade predita para a classe  $c$ , calculada pela função *softmax*.

### 2.3.3 Otimizadores

Em virtude da complexidade inerente aos modelos baseados em *deep learning*, utiliza-se de otimizadores para ajustar os parâmetros dos modelos, buscando minimizar a função de perda, proporcionando encontrar uma solução eficiente para a tarefa avaliada (RUDER, 2017).

Um algoritmo muito comum para otimização é o *Stochastic Gradient Descent* (SGD) que, basicamente, atualiza os pesos através do movimento em direção oposta ao gradiente da função de perda (ROBBINS; MONRO, 1951). O SGD atualiza os pesos de acordo com Equação 6.

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t), \quad (6)$$

onde  $\eta$  é a taxa de aprendizagem,  $\theta$  são os pesos e  $\nabla_{\theta} J(\theta_t)$  é o gradiente para atualização dos pesos.

No algoritmo do SGD, o *momentum* é um hiperparâmetro que é capaz de auxiliar a reduzir grandes oscilações e, conseqüentemente, acelerar a convergência do treinamento. O *momentum* acumula os gradientes passados para realizar uma atualização de pesos mais suave e é definido de acordo com a Equação 7.

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_t), \quad (7)$$

onde,  $v_t$  é o gradiente acumulado e  $\gamma$  armazena informação do gradiente anterior ( $v_{t-1}$ ).

Por fim, ao integrar o uso do *Momentum* no SGD, temos otimizador definido de acordo com a Equação 8.

$$\theta_{t+1} = \theta_t - v_t, \quad (8)$$

Outro otimizador muito utilizado é o Adam, que busca uma média dos gradientes passados para possibilitar uma atualização de pesos mais adaptativa (KINGMA, 2014). Desse modo o Adam armazena respectivamente: (i) a média exponencial dos gradientes ( $m_t$ ) e (ii) a média exponencial dos quadrados dos gradientes ( $s_t$ ), conforme definido pelas Equações 9 e 10:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} J(\theta_t), \quad (9)$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2)(\nabla_{\theta} J(\theta_t))^2, \quad (10)$$

onde  $\beta_1$  e  $\beta_2$  são fatores de decaimento exponencial da atualização dos pesos.

Diante disso, a atualização dos pesos no otimizador Adam é definida de acordo com as Equações 11 e 12.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{s}_t = \frac{s_t}{1 - \beta_2^t}, \quad (11)$$

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{s}_t + \epsilon}}, \quad (12)$$

onde  $\epsilon$  é um valor extremamente pequeno para evitar divisões por zero.

Outro otimizador muito utilizado para o treinamento de modelos baseados em *deep learning* é o AdamW, uma adaptação do Adam, mas que utiliza também o decaimento de pesos para direcionar a atualização dos pesos e, desse modo, também mitigar a possibilidade de *overfitting* (LOSHCHILOV; HUTTER, 2019). A partir disso, os componentes do AdamW são similares ao Adam e definidos de acordo com a Equação 13.

$$\theta_{t+1} = \theta_t - \eta \left( \frac{m_t}{\sqrt{v_t + \epsilon}} + \lambda \theta_t \right), \quad (13)$$

onde,  $\lambda$  é o fator de decaimento de pesos.

## 2.4 Modelos de Difusão Latente

Os modelos de difusão latente basicamente combinam a robustez de modelos baseados em difusão com a eficiência do aprendizado em representações em espaços latentes, o que gera uma redução considerável no custo computacional, preservando a qualidade da imagem. Os modelos de difusão latente possuem dois componentes principais: (i) o processo de difusão direta e (ii) o processo de reversão (HO; JAIN; ABBEEL, 2020). No processo de difusão direta, um ruído gaussiano é adicionado de forma progressiva a uma imagem codificada no espaço latente até que sua representação se aproxime de uma distribuição normal. Por outro lado, no processo de reversão, um modelo treinado aprende a reverter a imagem corrompida, removendo o ruído, a fim de reconstruí-la, gerando uma nova imagem cujo conteúdo está de acordo com o que foi treinado. (YANG et al., 2023). Por envolver a aplicação de uma rede neural, o processo de reversão é onde há o maior custo computacional de um modelo de difusão latente.

Nesse cenário os modelos de difusão latente demandam uma arquitetura de *autoencoder* a fim de representar uma determinada imagem e alta dimensão em um espaço latente comprimido. Na referida arquitetura, o *encoder* é responsável por transformar a imagem em uma representação compacta em um espaço latente e o *decoder* reconstrói a imagem

a partir dessa representação (ROMBACH et al., 2022). Assim, o mecanismo possibilita ao processo de difusão latente operar em uma dimensionalidade reduzida, corroborando com a redução do custo computacional supracitado (KAWAR et al., 2022). A estrutura básica de um modelo de difusão latente pode ser visualizada na Figura 3.

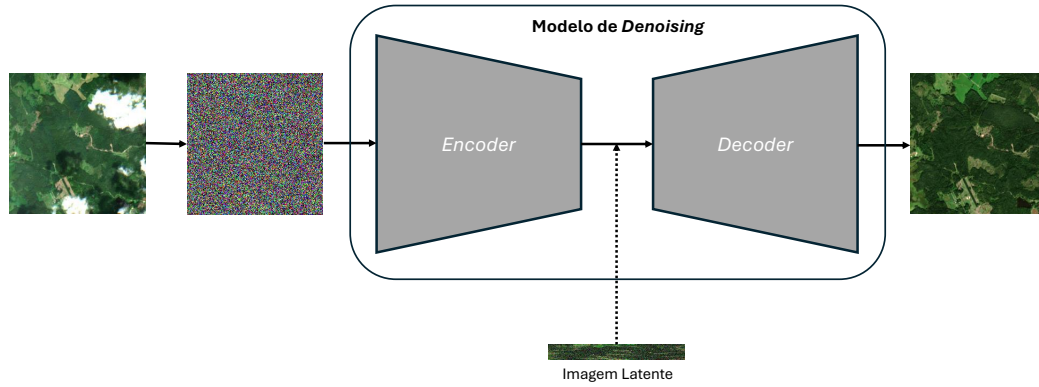


Figura 3 – Estrutura de um modelo de difusão latente.

Diante de suas características, esses modelos tem se tornado populares em diversas aplicações de visão computacional, tais como o preenchimento de regiões ausentes (*in-painting*), extensão de imagens (*uncropping*), super-resolução de imagens e transformação entre domínios (*image-to-image translation*) (SAHARIA et al., 2022). Tal versatilidade possibilita que esses modelos sejam adaptados também a várias aplicações como, por exemplo, análise de imagens médicas, restauração de imagens e sensoriamento remoto (PINAYA et al., 2022; XIAO et al., 2023; XIA et al., 2023).

## 2.5 Reconstrução de Imagens

A reconstrução de imagens visa preencher informações ausentes em imagens corrompidas, incompletas, de baixa resolução ou obstruídas por artefatos. O objetivo principal da reconstrução é gerar imagens visualmente coerentes, principalmente em relação à imagem original, e, por isso, essa técnica é aplicada em áreas como medicina e sensoriamento remoto (DESPRÉS; JIA, 2017; HE et al., 2019).

Alguns métodos tradicionais são baseados em estatística e determinismo (GONZALEZ, 2009). A interpolação espacial, por exemplo, emprega algoritmos como interpoladores para estimar valores de pixels ausentes com base em informações de pixels vizinhos. Os modelos variacionais minimizam uma determinada função que representa a regularização e a fidelidade aos dados. As transformadas de Fourier e Wavelets restauram informações a partir do domínio da frequência e, deste modo, é possível reconstruir regiões ausentes com base em padrões espectrais (SHEPP; LOGAN, 1974). Apesar das técnicas tradicionais serem eficazes em diferentes cenários, elas possuem limitações relacionadas, especialmente, a detalhes sutis, texturas complexas, ruídos intensos e regiões com grande oclusão.

Diante das limitações de métodos tradicionais, modelos baseados em *deep learning* têm alcançado resultados expressivos para a tarefa. As redes neurais convolucionais conseguem aprender padrões espaciais complexos que permitem reconstrução detalhada de imagens com artefatos, oclusões ou ausências. Por exemplo, uma GAN explora sua abordagem competitiva entre um gerador e um discriminador para fomentar o processo de geração de imagens realísticas e detalhadas, principalmente para tarefas de *inpainting*, que consiste na restauração de regiões ausentes ou danificadas de uma imagem com base nas informações ao redor (YATES et al., 2022). Por fim, os modelos de difusão latente aprendem a reconstruir as imagens através da remoção iterativa do ruído de imagens corrompidas em um espaço latente.

## 2.6 Segmentação de Imagens

A segmentação consiste, essencialmente, em particionar uma imagem em diferentes regiões ou categorias com base em características compartilhadas entre os elementos presentes. Em linhas gerais, podemos dividir segmentação em cinco tipos principais, cada um com suas respectivas técnicas. A limiarização é uma técnica que divide a imagem com base em um valor de intensidade, utilizando limiar local, global ou múltiplos limiares para separar objetos do fundo. A detecção de bordas visa identificar mudanças abruptas na intensidade da imagem, usando operadores como Roberts, Prewitt, Sobel, além de técnicas avançadas como a Transformada de Hough, Marr-Hildreth e Canny. A segmentação por regiões envolve a divisão da imagem em áreas homogêneas, aplicando métodos como crescimento de regiões, *split-merge* e *watershed* para identificar regiões distintas. A segmentação semântica consiste em atribuir uma classe a cada pixel da imagem, permitindo a segmentação com base no significado ou conteúdo de cada região. Já a segmentação por instâncias segmenta objetos individualmente dentro de uma mesma classe, diferenciando instâncias da mesma categoria em uma imagem (GONZALEZ, 2009).

Atualmente, métodos baseados em *deep learning* constituem o estado da arte para o problema da segmentação semântica, onde uma das mais conhecidas é a arquitetura U-Net. De forma geral, a U-Net pode ser dividida em duas partes: (i) *encoder* e (ii) *decoder* (RONNEBERGER; FISCHER; BROX, 2015; EPPENHOF et al., 2019).

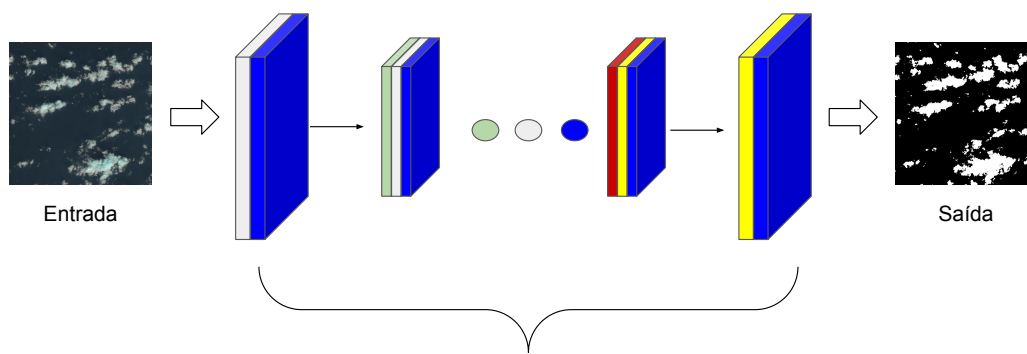
O *encoder* atua como uma sequência de camadas de convolução e *pooling*. Assim como nas CNNs tradicionais, nessa primeira etapa da U-Net há a extração das características (camada de convolução) da imagem de entrada e, posteriormente, a dimensionalidade da imagem é reduzida na camada de *pooling*. Tais instruções permitem a extração de características em múltiplas resoluções (RONNEBERGER; FISCHER; BROX, 2015).

O *decoder* realiza a expansão em níveis, de modo a realizar interconexões entre imagens e dimensões equivalentes, o que se refere ao processo de aumentar gradualmente a resolução da imagem durante a decodificação, permitindo a recuperação detalhada das



características espaciais enquanto se preserva a estrutura hierárquica da entrada. Além disso, a arquitetura U-Net possui como característica as conexões de salto entre a saída de um bloco de *encoder* e a entrada de um bloco de *decoder* equivalente, de forma que a imagem de saída possua dimensão equivalente à entrada (RONNEBERGER; FISCHER; BROX, 2015).

Além da U-Net, pode-se destacar também outras redes de segmentação disponíveis na literatura, tais como: Linknet (CHAURASIA; CULURCIELLO, 2017), PSP-Net (LI et al., 2018) e MANet (LI et al., 2021). A estrutura geral para redes de segmentação pode ser visualizada na Figura 4.



**Típicas arquiteturas baseadas em CNN para Segmentação**

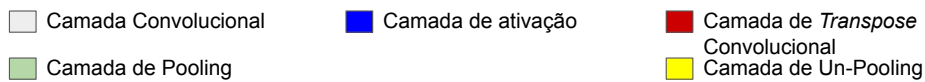


Figura 4 – Estrutura básica de métodos baseados em *deep learning* para segmentação de nuvens.

## 2.7 Fronteira de Pareto

A Fronteira de Pareto é uma técnica utilizada para analisar e avaliar experimentos, principalmente em termos de performance e recursos computacionais. Em linhas gerais, a fronteira de Pareto possibilita a análise e a comparação de soluções em problemas multi-objetivos e com métricas de avaliações conflitantes. Assim, um conjunto de experimentos é comparados com uma *baseline* para verificar quais deles podem manter desempenho de forma similar a esta *baseline*, considerando, principalmente, a relação entre eficácia e utilização de recursos computacionais (RAHMAN; SZABÓ, 2021).

A interpretação da Fronteira de Pareto pode ser através de gráficos bidimensionais, onde cada um dos eixos corresponde a um dos objetivos, por exemplo custo computacional e uma das métricas de avaliação do modelo. Nessa representação os pontos (experimentos) que estão na fronteira representam as melhores relações possíveis entre os objetivos,

enquanto os pontos fora da fronteira representam soluções subótimas (RUCHTE; GRABOCKA, 2021), conforme apresentado na Figura 5.

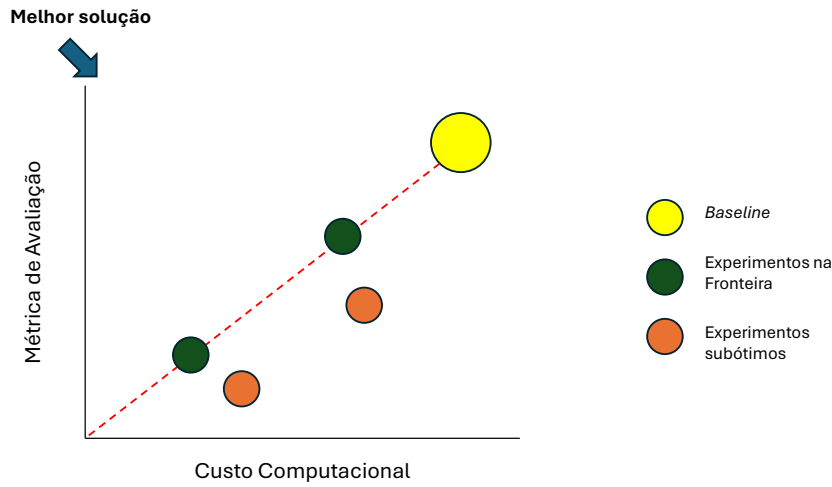


Figura 5 – Representação da Fronteira de Pareto como *trade-off* entre custo computacional e métrica de avaliação.

O Algoritmo 1 descreve instruções para determinar modelos presentes na Fronteira de Pareto em um exemplo onde considera-se número de parâmetros treináveis e acurácias (DEB et al., 2002). Entretanto, a análise baseada na Fronteira de Pareto pode apresentar desafios, especialmente quando há um número elevado de métricas a serem consideradas. Representações bidimensionais são úteis para visualização, mas podem não capturar toda a complexidade do problema. Dentre as alternativas para contornar essa situação pode-se (i) utilizar de projeções multidimensionais e (ii) avaliar cada métrica individualmente e considerar vários gráficos bidimensionais para a análise (PETCHROMPO et al., 2022; NGATCHOU; ZAREI; EL-SHARKAWI, 2005).

---

**Algoritmo 1** Algoritmo para estabelecer a Fronteira de Pareto

---

```

1: Entrada: Vetores  $n\_parametros$  (minimizar) e  $acuracias$  (maximizar)
2: Ordene os índices de  $acuracias$  em ordem decrescente:
    $sorted\_indices = \text{argsort}(-acuracias)$ 
3: Ordene os vetores  $acuracias$  e  $n\_parametros$  de acordo com os índices ordenados:
4:    $sorted\_acuracias = acuracias[sorted\_indices]$ 
    $sorted\_n\_parametros = n\_parametros[sorted\_indices]$ 
5: Inicialize:
    $pareto\_front = []$ 
    $pareto\_front\_acuracias = []$ 
    $pareto\_front\_n\_parametros = []$ 
    $min\_n\_parametros = \infty$ 
6: for  $i = 1$  to  $\text{len}(sorted\_acuracias)$  do
7:   if  $sorted\_n\_parametros[i] < min\_n\_parametros$  then
8:     Adicione  $sorted\_indices[i]$  a  $pareto\_front$ 
9:     Adicione  $sorted\_acuracias[i]$  a  $pareto\_front\_acuracias$ 
10:    Adicione  $sorted\_n\_parametros[i]$  a  $pareto\_front\_n\_parametros$ 
11:    Atualize  $min\_n\_parametros = sorted\_n\_parametros[i]$ 
12:   end if
13: end for
14: Saída:  $pareto\_front$ 

```

---



---

## Trabalhos Relacionados

Na literatura científica, uma das formas mais robustas de consolidar o conhecimento em uma determinada área é por meio de estudos de revisão. Em termos gerais, as revisões buscam por trabalhos de impacto através de bases acadêmicas relevantes, a fim de extrair e sumarizar dados para análise, interpretação, conclusões e proposições de perspectivas em uma área do conhecimento (ARMSTRONG et al., 2011). Especificamente em reconhecimento de padrões e aprendizado de máquina, esses estudos têm ganhado destaque por sintetizar uma área de pesquisa e assim fomentar novas pesquisas experimentais (XU; ZHU; CLIFTON, 2023; CROITORU et al., 2023; HAN et al., 2022; MINAEE et al., 2021; HOSPEDALES et al., 2021).

A estrutura de busca, critério de inclusão, critério de exclusão e análise deste trabalho foi baseada no método *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA). Essa abordagem é amplamente utilizada na literatura médica e possui itens recomendados para serem verificados durante a realização de um estudo de revisão (MOHER et al., 2010; TRICCO et al., 2018). Portanto, para aderir ao escopo desta pesquisa, foram feitas algumas adaptações na estrutura original do PRISMA para adaptá-la a um estudo de computação aplicada ao sensoriamento remoto. Esse uso adaptado para estudos em ciência da computação e outras áreas pode ser observado nos trabalhos de Huang et al. (2023) e Paula et al. (2023).

Para realizar a busca por trabalhos, as seguintes bases acadêmicas foram consideradas: Scopus<sup>1</sup>, *Association for Computing Machinery* (ACM)<sup>2</sup>, Springer<sup>3</sup>, *The Institute of Electrical and Electronics Engineers* (IEEE)<sup>4</sup> e *Multidisciplinary Digital Publishing Institute* (MDPI)<sup>5</sup>. A exploração dessas bases abrangeu o período entre outubro de 2023 e setembro de 2024. Foram considerados apenas estudos revisados por pares, escritos em inglês e publicados de 2019 a 2024. Assim, não foram incluídos estudos *preprints*, livros e

---

<sup>1</sup> <<https://www.scopus.com/>>

<sup>2</sup> <<https://libraries.acm.org/>>

<sup>3</sup> <<https://link.springer.com/>>

<sup>4</sup> <<https://ieeexplore.ieee.org/Xplore/home.jsp>>

<sup>5</sup> <<https://www.mdpi.com/librarians>>

capítulos de livros. É fundamental mencionar que, no escopo deste trabalho, os conceitos de remoção e reconstrução são considerados sinônimos. Embora, em algumas áreas de aplicação, a remoção não implique necessariamente na reconstrução, uma vez que certos trabalhos a tratam como um processo de segmentação seguido pela exclusão da região segmentada, neste trabalho ambos os conceitos são equivalentes. Portanto, quaisquer estudos que contenham o termo remoção e não realizem reconstrução foram descartados da amostra de obras a serem analisadas.

Partindo da premissa de que os métodos baseados em *deep learning* são atualmente o estado da arte para abordagem da visão computacional, esta revisão seguiu a estratégia de buscar primeiramente trabalhos que utilizassem essa técnica para remoção de nuvens. Para evitar ambiguidade, foram removidos os trabalhos que abordavam “nuvem de pontos” ou “computação em nuvem” em um contexto específico de *Internet of Things* (IoT). Assim, a busca foi realizada por meio de operadores lógicos (definidos pelos critérios mencionados), considerando a estrutura disponível em cada base científica (AHMED; DHUBAIB, 2011).

Após a aplicação da *string* de busca, foram encontrados 47 artigos, dos quais 7 duplicados e, conseqüentemente, removidos. Posteriormente, descartou-se outros 12 artigos, 5 por tratarem exclusivamente do problema de segmentação e detecção, e outros 7 por estarem completamente fora do escopo deste trabalho (por exemplo, considerar aplicações estritamente em séries temporais e sem aplicações em imagens). Portanto, inicialmente foram selecionados 28 estudos.

A partir desta primeira pesquisa, foram considerados outros métodos para remoção de nuvens além daqueles baseados em *deep learning*. Esse processo foi apoiado pelas referências bibliográficas fornecidos por estes 28 artigos. Desse modo, foi conduzida uma busca exaustiva nas respectivas referências bibliográficas para refinar os artigos considerados para esta revisão. Foram encontrados outros 44 artigos que satisfazem os mesmos critérios de inclusão e, assim, inseridos na amostra explorada nesta revisão. Por fim, uma nova busca exaustiva por trabalhos publicados entre 2023 e 2024 foi conduzida para dar uma maior robustez para a revisão, a fim de incluir trabalhos, aplicações e técnicas emergentes para a tarefa de remoção de nuvens. Essa nova busca resultou na inclusão de outros 48 artigos no conjunto de estudos analisados e, deste modo, tem-se um total de 120 trabalhos selecionados para serem considerados após a aplicação dos protocolos de inclusão e exclusão. O resumo do protocolo de revisão descrito pode ser visualizado na Figura 6.

### 3.1 Levantamento da Literatura

A literatura apresenta muitas propostas para remoção de nuvens em imagens de satélite, o que reflete o desafio para a tarefa (WEN et al., 2021; ZHANG et al., 2023; WANG

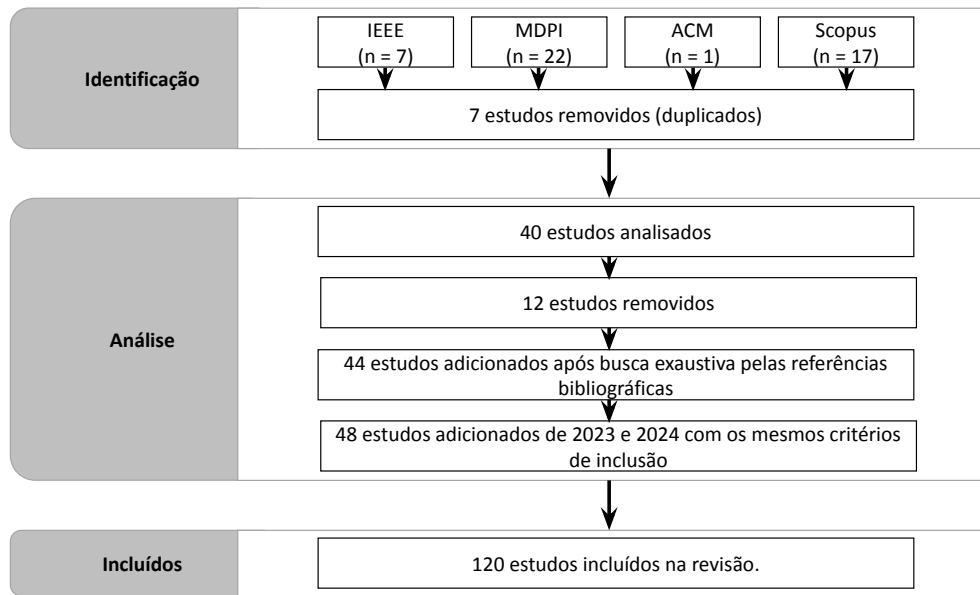


Figura 6 – Visão geral do protocolo de revisão deste trabalho. Adaptado de: Page et al. (2021).

et al., 2019). As GANs são constantemente consideradas, principalmente pela capacidade de gerar imagens de alta resolução por meio do aprendizado de distribuições complexas, o que corrobora para geração de imagens sem nuvens (LI; LIU; LI, 2023; WEN et al., 2021; ZHOU et al., 2022). As GANs, particularmente, têm sido utilizadas para remoção de nuvens finas e espessas, além de combinadas com outras técnicas, tais como fusão de imageamento óptico com *Synthetic-aperture radar* (SAR) (XIA et al., 2019; EBEL; SCHMITT; ZHU, 2020; GAO et al., 2020). As GANs possibilitam gerar imagens sem nuvens realísticas e preservam detalhes espaciais (ZHENG; LIU; WANG, 2020; LI et al., 2020; LI; LIU; LI, 2023; GHOZATLOU; DATCU, 2021).

As Redes Neurais Convolucionais (CNN) também são comumente utilizadas especialmente para remoção de nuvens finas pela capacidade de capturar padrões espaciais, onde as texturas finas são preservadas (YU; ZHANG; PUN, 2022; KHANDELWAL et al., 2022; EBEL; SCHMITT; ZHU, 2021; MERANER et al., 2020). A eficiência para remoção de nuvens pode ser relacionada à estrutura hierárquica, onde as camadas iniciais capturam os detalhes maiores da imagem, enquanto as camadas mais profundas conseguem extrair padrões mais complexos, o que gera versatilidade para tratar nuvens de diferentes densidades e tipos (DAI; JI; ZHANG, 2020; JING et al., 2023; CZERKAWSKI et al., 2022; HAN; WANG; ZHANG, 2023). As arquiteturas baseadas em CNN são frequentemente adaptadas, especialmente com conexões residuais ou mecanismos de atenção, para melhorar a capacidade de remoção de nuvens (WU et al., 2022; LI et al., 2021). Outros métodos, como fusão de dados, demandam a combinação intrínseca de múltiplos sensores para melhorar a precisão da remoção de nuvens; já a utilização do *Dark Channel Prior* (DCP)

e otimizadores de tensores representam métodos mais tradicionais que podem também oferecer alternativas robustas para contextos específicos (ZHANG et al., 2022; SHI et al., 2021; LI; HU; AI, 2018).

Ainda no contexto de CNN, diversas abordagens exploram arquiteturas do tipo *encoder-decoder* (LI; LI; CHAN, 2019; WEN et al., 2022; MERANER et al., 2020). A U-Net é uma arquitetura recorrente para solucionar tarefas de remoção de nuvens. Essa arquitetura, embora originalmente tenha sido projetada para segmentação de imagens biomédicas, tem sido efetivamente adaptada para remoção de nuvens, principalmente por meio de módulos especializados, tais como o Slope-Net e blocos residuais (ZI et al., 2021b; ZI et al., 2021a; LI et al., 2019a; LI et al., 2021). O sucesso desses módulos sugere que a simetria e as conexões de salto da U-Net auxiliam a preservar detalhes finos, o que garante a capacidade de remoção de nuvens (ZHANG et al., 2024). Essas características mencionadas demonstram a adaptabilidade da U-Net em delinear onde há obstruções relacionadas à nuvens, principalmente em contextos de nuvens finas, onde essa definição é extremamente sutil (SHEN et al., 2019).

Ainda há o foco em desenvolvimento de abordagens híbridas com, por exemplo, mecanismos de atenção e redes baseadas em *Transformers* (XU et al., 2021; LI et al., 2023). Esses modelos focam no aprendizado do espaço temporal para tratar grandes áreas ocluídas por nuvens. Por ser mais flexível, esse tipo de abordagem é visto como alternativa para situações em que há grande diferença temporal entre as imagens referência e alvo.

Um tema bastante observado pelos trabalhos analisados é relativo à importância de construir modelos robustos que trabalhem com variações atmosféricas. A capacidade de generalizar dentre diferentes tipos (finas e espessas) e em diferentes conjuntos de dados (por exemplo, SAR e imagens óticas) é o objetivo principal em muitos trabalhos. É possível notar na Figura 7 a distribuição dos trabalhos em termos de técnicas usadas por ano, onde, apesar de uma queda do número de publicações observadas até 2023, há um claro avanço a partir de 2024, principalmente pela popularização de modelos generativos (SARUKKAI et al., 2020; ZHAO et al., 2021).

Ao analisar o conjunto de artigos ao longo do tempo, é possível encontrar tendências na evolução e utilização de técnicas de remoção de nuvens (JI et al., 2020; ZHANG et al., 2022). Antes de 2020 era mais comum o uso de métodos tradicionais, como fusão de dados e fatoração de matriz (BERMUDEZ et al., 2019; LI et al., 2019c; LI et al., 2019; DU et al., 2019; LI et al., 2019b; ZHANG et al., 2019). Por mais que esses métodos tenham trazido desempenho considerável por determinado período, eles foram paulatinamente sendo substituídos em virtude das limitações no processo de aprendizagem e flexibilidade para adaptar aos diferentes tipos de nuvens e condições ambientais (ZHANG et al., 2021; JI et al., 2022; CHEN et al., 2019a).

A partir de 2020, as abordagens baseadas em *deep learning* se tornaram dominantes



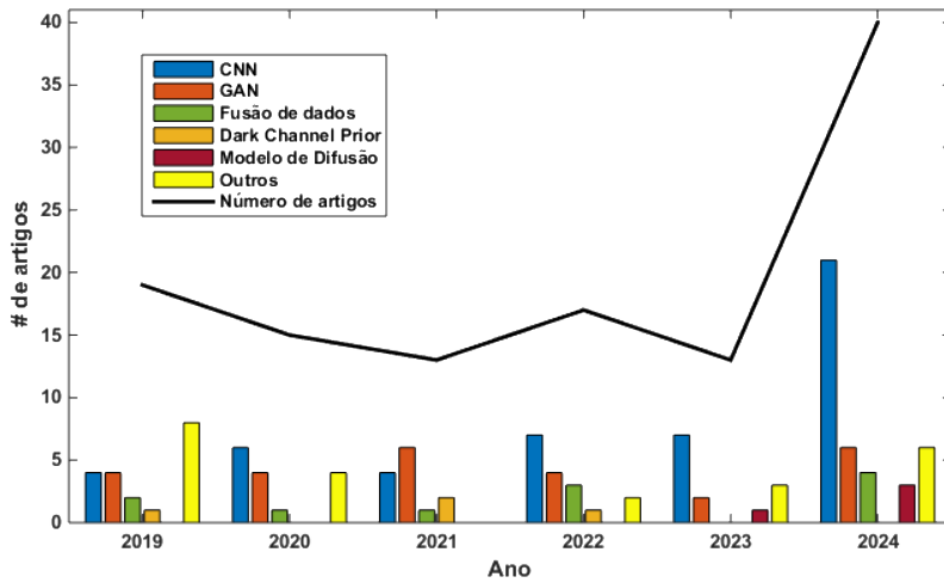


Figura 7 – Distribuição de trabalhos considerando ano e técnicas usadas.

para a tarefa por sua considerável adaptabilidade, capacidade de aprendizado de características diretamente dos dados, em vez de depender de operações predefinidas (YANG et al., 2020). A partir de 2022, há o aumento de modelos híbridos, principalmente pela utilização de técnicas baseadas em *Transformers* e Modelos de Difusão Latente. Esses modelos, que originalmente foram desenvolvidos para processamento de linguagem natural e geração de imagens, foram adaptados para tarefas específicas e complexas, como cenários de remoção de nuvens (SARÊ et al., 2024; LI et al., 2024a). Outra tendência observada desde 2022 é a fusão de dados multisensor, principalmente para situações onde há proeminência de nuvens espessas. Esse tipo de fusão pode fomentar a reconstrução de detalhes em imageamento por sensores que são sensíveis à presença de nuvens (DUAN; BELGIU; STEIN, 2024c; ZHAO et al., 2024).

## 3.2 Tipo de Remoção de Nuvens

A partir da revisão foi possível estabelecer que literatura é unânime ao distinguir entre remoção de nuvens finas e espessas, onde cada um dos tipos demanda diferentes níveis de técnicas e sofisticação para a imagem reconstruído. As nuvens finas aparentam ser um problema mais facilmente tratável, já que envolve um nível de transparência, podendo o solo ser visto através das nuvens, o que possibilita que informações de solo sejam consideradas (LI et al., 2022; WEN et al., 2022; TAN et al., 2024; DING et al., 2024). Assim, abordagens para esse tipo são motivadas pela remoção de interferência que esse tipo de nuvem ou neblina causa, ainda que alcançar um alto nível de acurácia por pixel seja tarefa desafiadora (XU et al., 2019; LI et al., 2024b; JIANG et al., 2024). As CNNs são usadas frequentemente para remoção de nuvens finas já que tendem a ter

um desempenho considerável para variações de textura e luminosidade. Assim, diversos modelos, como os que empregam redes multiescala ou aprendizado residual, objetivam encontrar características espectrais e espaciais para remover nuvens das imagens de satélite (LEE; SIM, 2019; MA et al., 2023).

Por outro lado, a remoção de nuvens espessas tende a ser uma tarefa mais complexa, visto que pouca ou nenhuma informação de solo está disponível devido ao nível de oclusão que esse tipo de artefato gera (ZHANG et al., 2020; CHEN et al., 2019b; DUAN; PAN; LI, 2020; LI et al., 2019; LI et al., 2019b). Assim, a revisão da literatura sugere que soluções para remover nuvens espessas geralmente demandam por modelos sofisticados, utilização de imageamento multitemporal e fusão de sensores (por exemplo o SAR, que é capaz de penetrar as nuvens) (CAO et al., 2020; LIU et al., 2024b; XIANG; TAN, 2024). A distribuição de artigos em termos de lista de técnicas utilizadas e tipo de nuvem considerada pode ser visualizada na Figura 8.

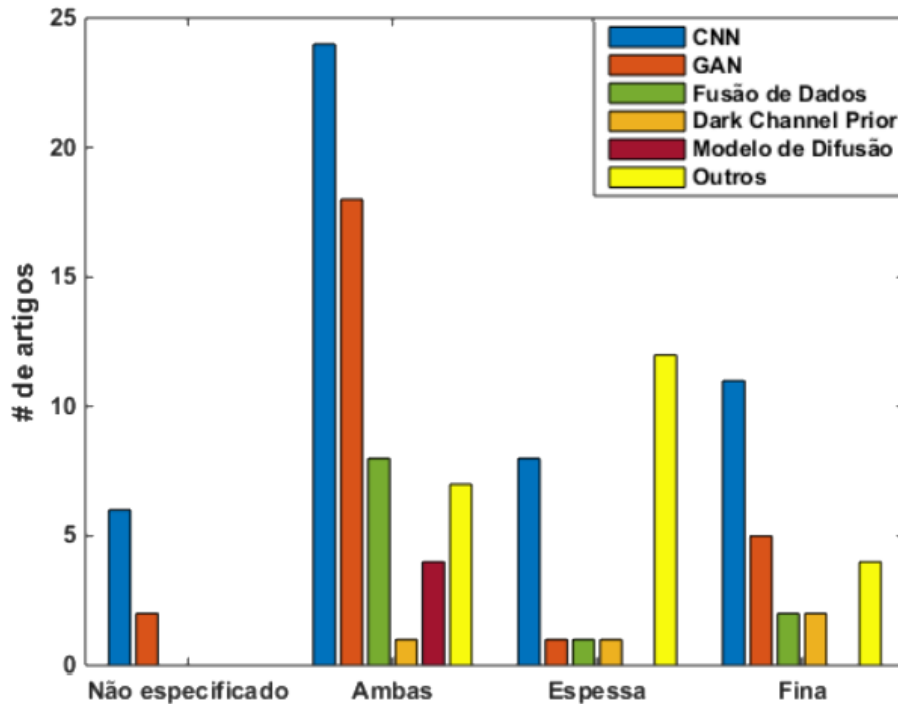


Figura 8 – Distribuição de trabalhos considerando técnicas e tipo de nuvens abordadas.

### 3.3 Número de Bandas Consideradas

Outro ponto de destaque é o número de bandas utilizada nos diferentes métodos considerados para a tarefa. Esse ponto é bastante variável, principalmente a depender do tipo de dado que está sendo processado (LIN et al., 2022). Algumas técnicas, particularmente as que consideram imagens hiperespectrais, lidam com até 166 bandas, o que possibilita uma reconstrução detalhada de áreas afetadas por nuvens através das informações das

múltiplas bandas (ANGEL; HOUBORG; MCCABE, 2019). As imagens multiespectrais, que tipicamente consideram menos bandas (geralmente entre 3 e 13) são muito utilizadas, inclusive integrando informações de SAR. Assim, trabalhos que consideram um número menor de bandas (entre 3 a 7) tendem a focar em um processamento mais rápido e um custo computacional menor, tornando os métodos, portanto, aplicáveis em cenários reais e com limitação de disponibilidade de dados (SUN et al., 2019; HASAN et al., 2022; ZHANG et al., 2024).

O número de bandas também impacta fortemente na complexidade da remoção de nuvens (ZHENG et al., 2023). Dados multi e hiperespectrais exigem por métodos mais avançados de processamento, visto que possuem um grande volume de informação e, deste modo, é observado uso de técnicas baseadas em *deep learning*, GANs e difusão latente, em detrimento de técnicas tradicionais. Por fim, a revisão também aponta para um considerável número de trabalhos onde não é possível inferir a quantidade de bandas utilizada nos experimentos, o que é uma informação preponderante, principalmente ao levar em conta o aspecto de reprodutibilidade.

## 3.4 Integração com SAR

Uma das tendências encontradas a partir da revisão é a integração com dados de SAR que, devido a sua característica de penetrar a cobertura da nuvem, podem ser uma informação complementar aos sensores óticos tradicionais (ARP et al., 2024; CHEN et al., 2020; SEBASTIANELLI et al., 2022). Cerca de 1/3 das técnicas de remoção de nuvens, particularmente as que usam GAN e CNN, usam fusão de dados com SAR para tratar as lacunas deixadas pela limitação de imageamento óptico (JING et al., 2022; DARBAGHSHAHI; MOHAMMADI; SORYANI, 2021). Assim, a fusão com imageamento de SAR possibilita uma reconstrução mais robusta da área nublada, visto que dados de radar geram detalhes estruturais da região, enquanto os dados ópticos complementam a reconstrução com informações espectrais (DUAN; BELGIU; STEIN, 2024b). Tal integração potencializa resultados robustos e são particularmente úteis em áreas que são naturalmente propensas à presença constante de nuvens (ZHAO; OLSEN; CHANDRA, 2023; EBEL et al., 2023; LI et al., 2023; LIU; HUANG; CAI, 2023).

Ainda neste contexto, o uso do SAR é ainda mais proeminente em contextos onde há nuvens espessas e o imageamento óptico sozinho tende a ser insuficiente (XIAO; LU, 2021; XIANG; TAN; YAN, 2024). Desse modo, através da fusão de diferentes sensores, os métodos de reconstrução podem melhorar a precisão da reconstrução para geração de imagens cada vez mais nítidas, independente da densidade da nuvem (EBEL et al., 2020a; XU et al., 2022). O uso do SAR é especialmente comum em trabalhos mais recentes por empregar a fusão multisensor e grafos de agregação de características, o que possibilita a modelagem mais complexas de interações entre duas fontes diferentes de aquisição de

dados (ZHOU; WANG, 2019; GAO et al., 2021; CHEN et al., 2022). Uma visão geral dos trabalhos avaliados, onde é possível observar que a um aumento do uso de SAR nos últimos anos, principalmente para nuvens espessas e com métodos baseados em *deep learning* pode ser observada na Figura 9.

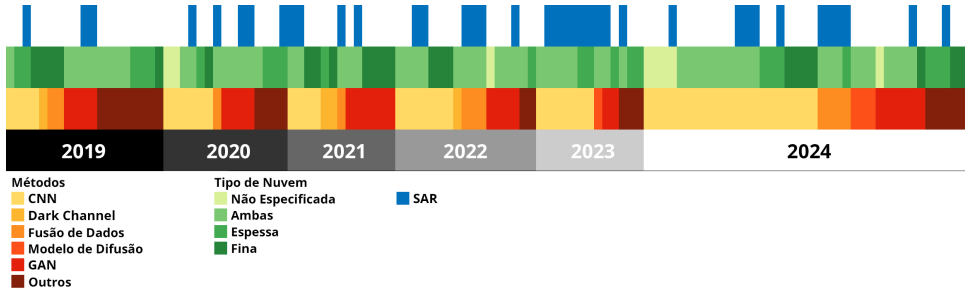


Figura 9 – Distribuição dos artigos considerando múltiplos parâmetros.

### 3.5 Técnicas Emergentes

Nos últimos anos é possível notar que houve um aumento em técnicas de ponta, principalmente métodos generativos, para realizar a remoção de nuvens (HIRAHARA et al., 2020; ZHAO; JIA, 2023; WANG et al., 2023; LYU; QIAN; WANG, 2024; CHEN et al., 2024). Assim, uma tendência observada é a aplicação métodos de difusão latente que tem tido considerável destaque nas mais diversas áreas de aplicação para geração de imagens sintéticas. A abordagem baseada em mecanismos de *denoising* tem demonstrado robustez em remoção de nuvens, como destacado em Sui et al. (2024a). Assim, estes modelos tem gradualmente melhorado a geração de imagens, por meio de seus sucessivos passos iterativos que podem, portanto, aprimorar a reconstrução para imagens livres nuvens com grande precisão (YANG et al., 2024; ZOU et al., 2024).

*Transformers* é outra técnica emergente para remoção de nuvens por sua capacidade de capturar grandes intervalos de dependências e informações contextuais. Assim, muitos estudos demonstram que abordagens baseadas em *Transformers*, principalmente pelos mecanismos de *self-attention*, podem melhorar a remoção de nuvens através da melhor captura de relações espaciais e espectrais dentre imagens principalmente de natureza multitemporal ou entre fusão de SAR e dados ópticos (SUN; JI; DING, 2024; WU et al., 2024; ZI et al., 2024). Ainda é importante mencionar que essas arquiteturas apresentam flexibilidade e escalabilidade, principalmente visto que dados de sensoriamento remoto aumentam o volume e a resolução, consequentemente demandam por recursos computacionais mais robustos (LIU; PAN; SHI, 2024).

Além dos métodos supracitados, há também a inserção de modelos híbridos, que fornecem uma abordagem abrangente e complementar para remoção de nuvens (TU et al., 2024). Por combinarem *Transformers* com GAN ou CNN, diversos pesquisadores têm de-

monstrado resultados melhores para remoção de nuvens, principalmente onde há cenários complexos (por exemplo, quando há nuvens finas e espessas simultaneamente) (GHILDIYAL et al., 2024; MA et al., 2024; BASKAR; PARAMBALATH; KRISHNANUNNI, 2024). Estes modelos híbridos podem aproveitar das vantagens de cada modelo em particular, conforme exposto em Zhao et al. (2021), já que podem combinar treinamento de natureza adversaria com fusão multimodal para uma melhor generalização. Outra tendência observada é a crescente incorporação de mecanismos de atenção em arquiteturas baseadas em *deep learning*, visto que possibilita o modelo focar o processo de aprendizado em partes mais relevantes para o processo de reconstrução (WANG et al., 2024; AJAYI; OLAIDE, 2024; JIN; WANG; LI, 2024; SUI et al., 2024b).

Assim, essas técnicas emergentes apontam para uma mudança para modelos cada vez mais flexíveis e poderosos para que sejam capazes de adaptar a diferentes coberturas de solo, tipos de nuvens e condições ambientais (ZHENG et al., 2024; LI et al., 2024; ZHANG et al., 2024). Portanto, à medida que esse campo continua mostrando crescimento, é esperado que esses modelos emergentes avancem ainda mais e desempenhem um papel fundamental na melhoria da qualidade e eficiência da geração de imagens de satélite sem nuvens (DUAN; BELGIU; STEIN, 2024a; YAN et al., 2024; DOU et al., 2024).

## 3.6 Bases de Dados

A partir do protocolo de revisão, coletou-se quatro trabalhos que abordam exclusivamente a proposição de bases de imagens, sendo: Schmitt et al. (2019), Ebel et al. (2022), Aybar et al. (2022) e Czerkawski et al. (2023). Há também artigos que realizam os experimentos em datasets sintéticos ou em ambientes com nuvens simuladas, como observado em Ji et al. (2020) e Ma et al. (2023), o que produz ambientes controlados para treinamento, mas com restrições quando considerados ambientes reais. Especificamente em Czerkawski et al. (2023), os autores abordam a geração de nuvens artificiais para construção de datasets de imagens nubladas a partir de imagens sem nuvens. Esse tipo de alternativa é particularmente interessante para avaliar a reconstrução, visto que a imagem original exata está disponível.

Além das bases de dados mencionados, há trabalhos levantados nesta revisão onde os autores constroem suas bases a partir de imagens públicas de satélites. Nesses casos, as imagens foram coletadas através das agências de monitoramento espaciais, tais como como EarthExplorer<sup>6</sup>, SentinelSat<sup>7</sup>, Theia<sup>8</sup>, The European Space Agency<sup>9</sup>, Apollo Mapping<sup>10</sup> e Natural Earth<sup>11</sup> e, deste modo, as referidas imagens são usadas para criar bases de dados

<sup>6</sup> <<https://earthexplorer.usgs.gov/>>

<sup>7</sup> <<https://sentinelsat.readthedocs.io/en/stable/>>

<sup>8</sup> <<https://theia.cnes.fr/atdistrib/rocket>>

<sup>9</sup> <<https://earth.esa.int/eogateway/catalog/worldview-2-european-cities>>

<sup>10</sup> <<https://apollomapping.com/>>

<sup>11</sup> <<https://www.naturalearthdata.com/>>

específicos para o treinamento do respectivo trabalho. Também foi possível levantar outras bases de dados utilizados em diversos artigos, mas que podem não ter sido capturados pela *string* de busca em virtude do período que foi considerado no protocolo de revisão, mas que podem ser explorados em futuras pesquisas. A maioria dessas bases de imagens apresentam dados de natureza multiespectral, ou seja, em diferentes bandas tais como, visíveis, infravermelho e outras. Essas bases de dados com respectivas características principais podem ser observadas na Tabela 3.

Tabela 3 – Bases de Dados para Remoção de Nuvens

Bases de Dados	# de Bandas	SAR
Rice Dataset (LIN et al., 2019)	4	Não
MEcGANs - WorldView-2 European Cities (HASAN et al., 2022)	4	Não
Paired Sentinel-1 and Sentinel-2 Images for 2 Locations in Scotland and India for 2019 and 2020 (CZERKAWSKI et al., 2022)	3	Sim
WHUS2-CR - A Thin Cloud Removal Dataset for Sentinel-2 Images (LI et al., 2020; LI et al., 2021)	13	Não
WHUS2-CRv - Sentinel-2A Cloud Removal Dataset (LI et al., 2020; LI et al., 2021; LI et al., 2022)	12	Não
SEN1-2 Dataset (SCHMITT; HUGHES; ZHU, 2018)	12	Sim
SEN12MS Dataset (SCHMITT et al., 2019)	13	Sim
Sen12MS-CR Dataset (EBEL et al., 2020b)	13	Sim
Sen2_MTC Dataset (HUANG; WU, 2022)	4	Não
Sen12MS-CRP-TS Dataset (EBEL et al., 2020b)	13	Sim
Campo Verde Database (SANCHES et al., 2018)	-	Sim
CloudSEN12 (AYBAR et al., 2022)	12	Sim
Landsat 8 Cloud Cover Assessment Validation Data (XIE et al., 2023)	11	Sim
Thin cloud removal dataset for Sentinel-2 images (LI et al., 2020; LI et al., 2021)	13	Não
WHU Cloud Dataset (DAI; JI; ZHANG, 2020)	3	Não
EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Cover Classification (HELBER et al., 2019)	13	Não
Spatio-Temporal SAR-Optical Data Fusion for Cloud Removal Dataset (SEBASTIANELLI et al., 2021)	3	Sim
Cloud Removal Dataset (SARUKKAI et al., 2020)	4	Não

### 3.7 Métricas de Avaliação

A revisão da literatura também possibilitou listar as principais métricas de avaliação em trabalhos que envolvem reconstrução de nuvens. As principais métricas para reconstrução obtidas foram: *Structural Similarity Index Measure* (SSIM), *Peak Signal-to-Noise Ratio* (PSNR), *MSE/Root-Mean-Square Error* (RMSE), *Correlation Coefficient* (CC), *Similarity Assessment Metric* (SAM) e MAE, as quais serão descritas na sequência. Enquanto na Figura 10 observa-se o relacionamento dentre as métricas e suas frequências de uso, em um contexto em que um determinado trabalho pode ter usado mais de uma métrica; na Figura 11 nota-se a combinação de uso das métricas de modo a relacioná-las com os trabalhos no decorrer dos anos. Nesse ponto, é possível evidenciar que a maioria dos trabalhos utilizam as métricas SSIM e PSNR para avaliar a reconstrução de imagens nubladas, bem como a ampla maioria dos trabalhos utilizam ao menos duas métricas para uma avaliação pertinente das abordagens propostas.

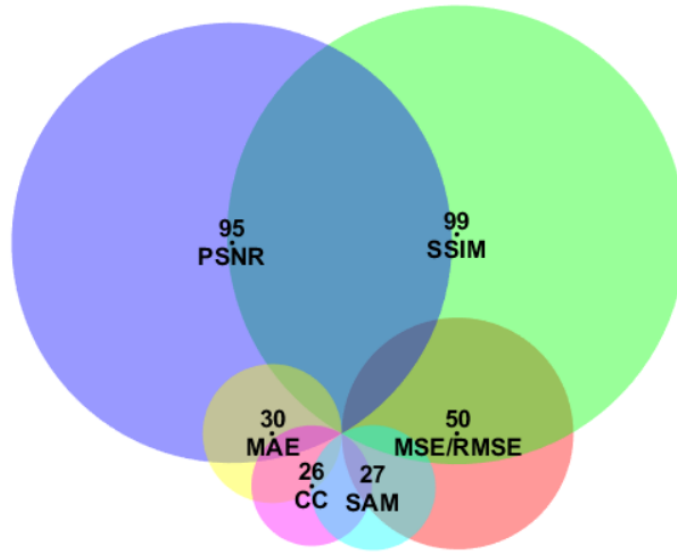


Figura 10 – Relacionamento das métricas com os trabalhos considerados.

O MAE apresenta as diferenças médias absolutas (pixel a pixel) entre os valores previstos e esperados. O MAE determina uma medida objetiva de discrepância, independentemente da direção do erro (positiva ou negativa). MAE é calculado de acordo com a Equação 14:

$$\text{MAE}(x, y) = \frac{\sum_{i=1}^n |x_i - y_i|}{n}, \quad (14)$$

onde:  $n$  é o número total de pixels;  $x$  é o valor real do pixel  $i$ ;  $y$  é o valor previsto do pixel  $i$ .

O SAM também avalia a semelhança entre duas imagens e considera o quadrado da diferença, pixel a pixel, entre uma imagem. Essa métrica compara principalmente

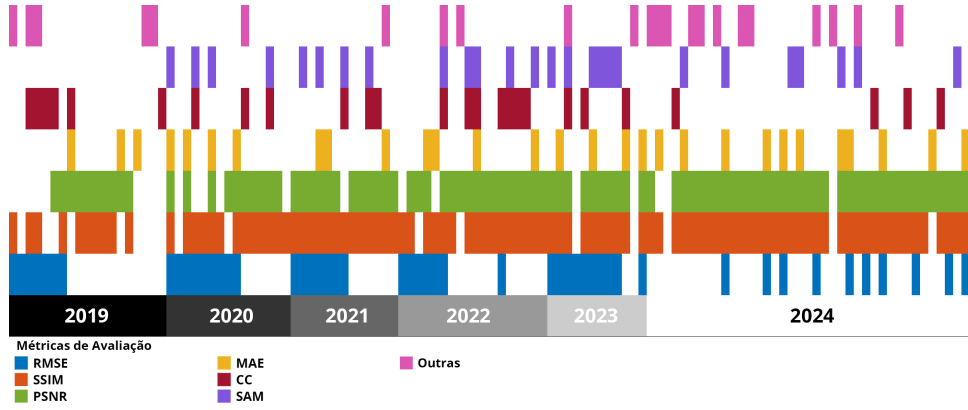


Figura 11 – Utilização combinada de diferentes métricas ao longo dos anos para os artigos selecionados.

características de alta frequência por meio de transformadas, como Fourier ou Wavelet. O cálculo do SAM é descrito na Equação 15.

$$\text{SAM}(x,y) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{n}, \quad (15)$$

onde também:  $n$  é o número total de pixels;  $x$  é o valor real do pixel  $i$ ;  $y$  é o valor previsto do pixel  $i$ .

O RMSE, uma variação MAE, é uma métrica para avaliar a diferença entre os valores previstos e reais por meio do erro estimado para cada observação. O RMSE é calculado como a média quadrada das diferenças entre os valores observados e previstos (WANG; BOVIK, 2009). Dado uma imagem de referência  $I$  de dimensões  $m \times n$ , o RMSE de outra imagem  $K$ , de  $I$ , é dado pela Equação 16:

$$\text{RMSE}(x,y) = \sqrt{\frac{1}{mn} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (x_{(m,n)} - y_{(m,n)})^2}, \quad (16)$$

onde:  $x_{(m,n)}$  é um valor de pixel previsto e  $y_{(m,n)}$  é um valor de pixel verdadeiro.

O PSNR é outra métrica para avaliar a distorção entre imagens. De modo geral, esta métrica é definida em termos de RMSE (HORE; ZIOU, 2010). A Equação 17 define o PSNR:

$$\text{PSNR}(x,y) = 20 \log_{10} \left( \frac{MAX_T}{RMSE} \right), \quad (17)$$

onde  $MAX_T$  é o valor máximo de pixel da imagem.

O SSIM consiste em uma métrica para avaliar a qualidade de uma imagem utilizando uma imagem de referência (livre de distorção e compressão) (WANG et al., 2004). Em termos gerais, a métrica SSIM é definida de acordo com a Equação 18:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (18)$$



onde:  $x$  e  $y$  são dois *patches* obtidos das respectivas imagens a serem avaliadas. Valores de  $x$  e  $y$  são definidos como valores não negativos que representam um sinal, que deve estar alinhado;  $\mu_x$  é a média da amostra de pixels do sinal  $x$ ;  $\mu_y$  é a média da amostra de pixel do sinal  $y$ ;  $\sigma_x^2$  é a variância do sinal  $x$ ;  $\sigma_y^2$  é a variância do sinal  $y$ ;  $\sigma_{xy}$  é a covariância dos sinais  $x$  e  $y$ ;  $C_1$  e  $C_2$  são duas constantes incluídas para evitar instabilidade para valores muito próximos de zero.

O CC é uma métrica de avaliação linear entre os valores de intensidade dos pixels em duas imagens. Em geral, o CC mede o quão semelhantes são duas imagens. O valor CC varia entre  $-1$  e  $1$ , onde valores próximos de  $1$  indicam correlação positiva, valores próximos de  $-1$  indicam correlação negativa e valores próximos de  $0$  indicam nenhuma correlação. O CC é calculado de acordo com a Equação 19:

$$CC(x,y) = \frac{(\sum_{i=1}^n x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}}, \quad (19)$$

onde:  $n$  é o número total de pixels;  $x_i$  e  $y_i$  são valores de intensidade dos respectivos pixels para imagens  $x$  e  $y$ ;  $\mu_x$  e  $\mu_y$  referem-se aos valores médios de intensidade em ambas as imagens, respectivamente.



## Avaliação de Arquiteturas e *Encoders* para Segmentação de Nuvens

A segmentação de nuvens é parte importante no processo de reconstrução da área nublada. Diante disso, investigou-se também o comportamento de arquiteturas tradicionais de segmentação com diferentes combinações de *encoders* nessa tarefa. Desse modo, este capítulo apresenta uma análise abrangente da combinação de quatro arquiteturas de segmentação (U-Net, LinkNet, PSP-Net e MA-Net) combinada com seis *encoders* pré-treinados (VGG-16, ResNet-50, Inception V4, Densenet-121, MobileNet V2 e EfficientNet B2) e, deste modo, traz contribuições e inspirações para análise de imagens de satélite.

### 4.1 Base de Dados

A base de dados 38-Cloud<sup>1</sup> é uma fonte de dados importante e reconhecida para a área de sensoriamento remoto e visão computacional, sendo projetado especificamente para a segmentação e detecção de nuvens em imagens de satélite (MOHAJERANI; KRAMMER; SAEEDI, 2018; MOHAJERANI; SAEEDI, 2019). Esta base de dados possui 38 cenas completas extraídas do satélite Landsat 8 em quatro bandas (azul, verde, vermelho e infravermelho próximo) que compreendem diferentes regiões da Terra, condições climáticas e coberturas de solo. Cada cena da base de dados é dividida em *patches* de  $384 \times 384$ , com separação em conjuntos de treinamento e teste, nos quantitativos de 8400 e 9201 *patches* respectivamente. Como o 38-Cloud não possui explicitamente um conjunto de validação, optou-se por selecionar aleatoriamente 30% do conjunto de treinamento para compor a validação. É importante mencionar que o conjunto de treinamento possui imagens manualmente marcadas delineando áreas nubladas para serem utilizados como *ground-truth*. Exemplos de, respectivamente, cenas e *patches* do 38-Cloud podem ser observados nas Figuras 12 e 13.

<sup>1</sup> <<https://github.com/SorourMo/38-Cloud-A-Cloud-Segmentation-Dataset>>

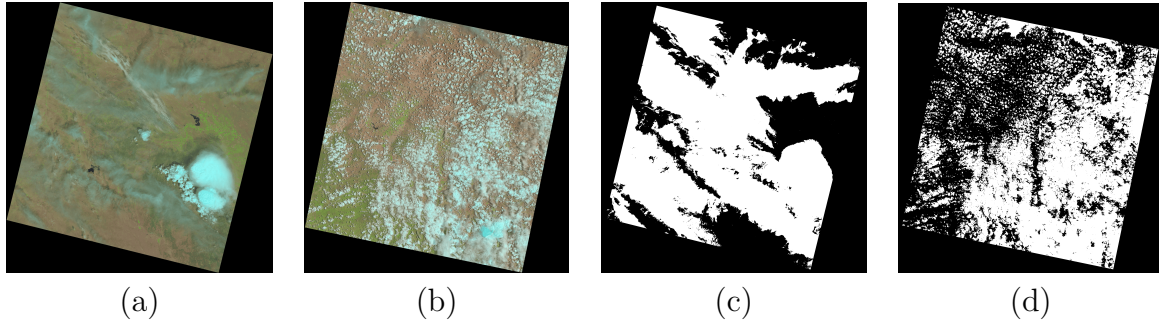


Figura 12 – Exemplos de cenas completas do conjunto de dados 38-Cloud. Em pseudo-cores (a - b) e seus respectivos *ground-truths* (c - d).

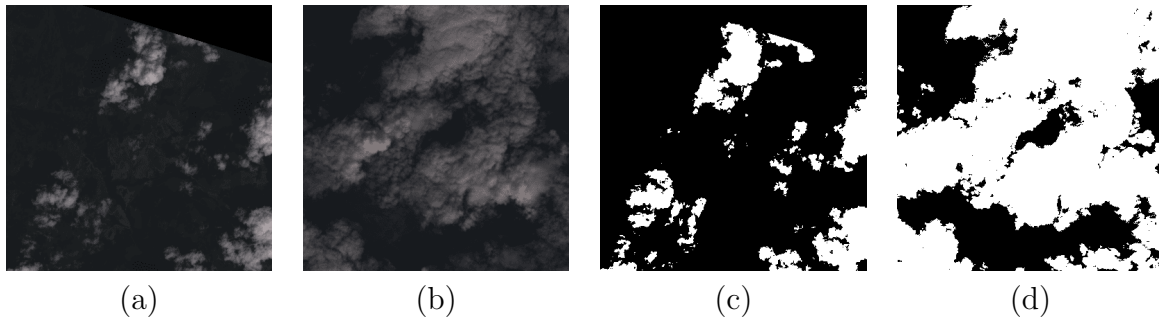


Figura 13 – Exemplos de imagens em forma de *patch* do conjunto de treinamento do 38-Cloud. Em tons de cinza (a - b) e seus respectivos *ground-truths* (c - d).

## 4.2 Arquiteturas e *Encoders*

A U-Net é uma arquitetura clássica para segmentação de imagens, sendo consolidada também para tarefas de sensoriamento remoto (EPPENHOF et al., 2019; SILVA et al., 2022). A U-Net é considerada neste trabalho como uma arquitetura do projeto experimental de segmentação, juntamente com outras arquiteturas inspiradas nela, as quais são expostas a seguir.

A LinkNet é outra arquitetura do tipo *encoder-decoder* projetada para segmentação semântica (CHAURASIA; CULURCIELLO, 2017). Nessa arquitetura, o *encoder* extrai características da imagem de entrada, enquanto o *decoder* reverte a informação espacial, de modo a realizar a segmentação. A LinkNet utiliza blocos residuais para auxiliar a estabilização dos gradientes e, consequentemente, melhora a capacidade de aprender características complexas da imagem de entrada (FAN et al., 2020).

A PSPNet é outra abordagem baseada em CNN para segmentação de imagens (ZHAO et al., 2017). Essa arquitetura é capaz de capturar informações de contexto global efetivamente, o que é adequado para tarefas onde a compreensão das relações entre objetos e seu entorno é crucial, como o reconhecimento de objetos em imagens. O módulo de *pooling* em pirâmide proposto nessa arquitetura é capaz de capturar informações contextuais multiescala de diferentes regiões da imagem. Essa operação ocorre pela divisão do mapa de características em um *grid* de tamanho fixo para aplicar o agrupamento médio.

A MA-Net é uma variação da U-Net que incorpora de blocos de atenção posicionais e blocos de atenção de fusão multiescala para melhorar a capacidade de extrair características importantes das imagens. Através dos modelos de atenção, a MA-Net tende a melhorar acurácia, especialmente em termos espaciais e entre canais.

Como mencionado, arquiteturas de segmentação consistem em estruturas do tipo *encoder-decoder*. Desse modo, diferentes *encoders* podem ser avaliados para a tarefa de segmentação. Diante disso, a partir das arquiteturas supracitadas, considerou-se combiná-las com seis diferentes *encoders* para avaliação de um cenário onde há a presença de nuvens. Esses seis *encoders* são expostos a seguir.

A VGG-16 é uma arquitetura clássica e conhecida pela sua simplicidade e eficiência (SIMONYAN; ZISSERMAN, 2014). Ela consiste em 16 camadas empilhadas com pequenos filtros convolucionais de  $3 \times 3$  em cada camada. Devido a sua simplicidade, é uma arquitetura de fácil compreensão e adaptação e, desse modo, é comumente considerada como *baseline* para diversas tarefas de visão computação e como *benchmark* em problemas relacionados com classificação de imagens.

A Inception V4 é uma arquitetura projetada com base nas versões anteriores do Inception para tarefas de classificação de imagens (SZEGEDY et al., 2015). Essa arquitetura apresentou avanços significativos pela integração com conexões residuais, o que auxilia o treinamento de redes profundas no tratamento do gradiente. Basicamente, a Inception V4 combina camadas de convolução, operações de *pooling* e módulos Inceptions para gerar uma extração de características eficiente e em múltipla escala.

A ResNet-50 consiste em uma variação da ResNet com 50 camadas com blocos residuais e conexões de atalho para possibilitar o treinamento de redes com grande profundidade. Essas características auxiliam a otimização e a melhoria da precisão para a tarefa a ser abordada (HE et al., 2016).

A DenseNet é uma arquitetura de CNN que emprega blocos densos para concatenar as saídas das múltiplas camadas convolucionais a fim de melhorar o desempenho. Nessa arquitetura, as camadas de transições reduzem o número de parâmetros e, juntamente com as operações de *pooling*, reduzem a dimensionalidade. Os blocos densos e as camadas de transição proporcionam que o gradiente flua de forma mais sutil, levando a uma melhor troca de informações entre camadas e eficiência de aprendizagem. Para este estudo, foi considerado a DenseNet com 121 camadas (HUANG et al., 2017).

Outra CNN considerada como *encoder* é a MobileNet V2 (SANDLER et al., 2018). Ela, basicamente, objetiva reduzir o custo computacional, enquanto mantém desempenho competitivo. A MobileNet V2 é projetada para dispositivos com restrição de recursos computacionais, como dispositivos móveis e embarcados. Essa arquitetura reduz consideravelmente o número de parâmetros comparada às arquiteturas tradicionais através da utilização de convoluções separáveis em profundidade.

Por fim, a EfficientNet B2 possui um número de parâmetros balanceado e moderado em

relação a outros modelos maiores, ao mesmo tempo em que fornece resultados consistentes (TAN; LE, 2019). A EfficientNet B2 é considerada por muitos pesquisadores por equilibrar a profundidade, largura e resolução do modelo para otimizar o desempenho e a eficiência.

### 4.3 Configuração Experimental

A combinação das arquiteturas com os *encoders* foi realizada através da biblioteca *Segmentation PyTorch Library* (SMP)<sup>2</sup> que possui os modelos pré-treinados pela Imagenet (DENG et al., 2009). Além disso, utilizou-se o otimizador Adam com taxa de aprendizagem inicial de  $10^{-4}$  com a função de perda entropia cruzada. O tamanho do lote foi de 16 para todos os experimentos, exceto quando se utilizou o *encoder* MobileNet V2 que, devido ao menor número de parâmetros e possibilidade de execução em uma placa de 11 GB de memória, considerou-se 24. Através da redução da taxa de aprendizagem nos platôs, configurou-se um decaimento da taxa de aprendizagem a um fator de 0,1 quando a perda da validação não melhorasse por 10 épocas consecutivas. Por fim, para mitigar a possibilidade de *overfitting*, optou-se por utilizar a estratégia de parada antecipada, onde, caso o modelo não melhorasse a perda da validação por 21 épocas consecutivas, o treinamento era encerrado, preservando os parâmetros da época com menor perda de validação. Um resumo da configuração proposta pode ser observado na Figura 14.

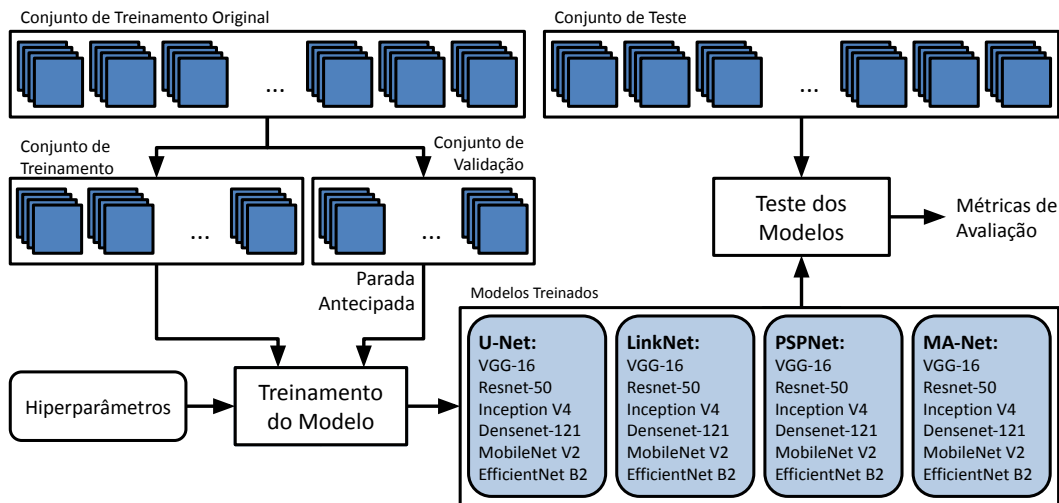


Figura 14 – Configuração Experimental para avaliação da segmentação de nuvens.

Os experimentos foram conduzidos em três PCs com processadores i5 com 3.0 GHz. Dois dos PCs possuem NVIDIA GTX 1080 ti *Graphics Processing Unit* (GPU) com 11 GB de memória e o terceiro uma NVIDIA Titan XP com 12 GB. O sistema operacional era Ubuntu 22.04 LTS e os experimentos desenvolvidos com Python 3.9, Scikit-learn

<sup>2</sup> <[https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch)>

1.2.0, Matplotlib 3.7.1, PyTorch 2.0.1, torchvision 0.15.12 e CUDA 11.0. Utilizou-se as arquiteturas de segmentação e os modelos pré-treinados da biblioteca SMP 0.3.2.

Para a avaliação dos resultados, utilizou-se a Acurácia (Acc), *Intersection over Union* (IoU) e Coeficiente de Dice (Dice). Todas essas métricas são relacionadas em termos de Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), Falsos Positivos (FP) e Falsos Negativos (FN).

A acurácia mede a proporção de classificação correta, pixel a pixel, de uma imagem e é definida de acordo com a Equação 20.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (20)$$

O IoU é uma métrica capaz de mensurar a sobreposição entre a segmentação realizada e o *ground-truth*, considerando a razão entre a interseção e a união dessas regiões. Com valores em um intervalo entre 0 e 1, um IoU próximo a 1 indica uma melhor segmentação. Essa métrica é definida de acordo com a Equação 21.

$$\text{IoU} = \frac{\text{TP}}{\text{FP} + \text{TP} + \text{FN}} \quad (21)$$

O coeficiente de Dice é semelhante ao IoU em um contexto em que ambos avaliam a sobreposição entre a segmentação realizada e o *ground-truth*. Especificamente, o coeficiente de Dice é obtido através da média harmônica entre precisão e *recall*, como descrito na Equação 22.

$$\text{Dice} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}} \quad (22)$$

## 4.4 Resultados

Todos os resultados em termos de acurácia, IoU e Dice podem ser observados na Tabela 4. Nessa tabela é possível comparar o desempenho de cada combinação em termos das métricas consideradas, a quantidade de épocas treinadas e o número de parâmetros.

O resultados demonstram que independente da arquitetura considerada, o *encoder* VGG-16 alcança os melhores resultados para todas as métricas consideradas. Esse resultado corrobora com o reportado na literatura, onde o VGG-16 é constantemente mencionado como um excelente extrator de características e, diante disso, utilizado em um grande número de aplicações e confirmado também para a segmentação de nuvens. Por outro lado, os *encoders* MobileNet V2 e EfficientNet B2 tiveram os piores resultados dentre os avaliados. Esse resultado pode ser atribuído à menor quantidade de parâmetros, o que impacta no processo de aprendizagem.

Em termos de arquitetura, a PSPNet utiliza um menor número de parâmetros, independente do *encoder* considerado. Entretanto, isso gera um desempenho inferior. Assim,

Tabela 4 – Resultados experimental para segmentação conjunto de teste. “Ép.” representa o número de épocas de treinamento até a parada antecipada. “# Params” representa o número de parâmetros treináveis.

Arquitetura	Encoder	# Params (M) ↓	Ép.	Acc ↑	IoU ↑	Dice ↑
U-Net	VGG-16	23,75	196	<b>0,9603</b> ± 0,0393	<b>0,8227</b> ± 0,0983	<b>0,8789</b> ± 0,0714
	ResNet-50	32,52	168	0,9579 ± 0,0374	0,7956 ± 0,1318	0,8542 ± 0,1052
	Inception V4	48,79	133	0,9593 ± 0,0341	0,7904 ± 0,1276	0,8479 ± 0,1065
	Densenet-121	13,61	141	0,9553 ± 0,0428	0,7911 ± 0,1248	0,8512 ± 0,0982
	MobileNet V2	6,63	139	0,9569 ± 0,0367	0,7791 ± 0,1360	0,8381 ± 0,1163
	EfficientNet B2	10,05	75	0,9588 ± 0,0331	0,7724 ± 0,1600	0,8309 ± 0,1417
PSPNet	VGG-16	15,51	132	<b>0,9571</b> ± 0,0372	<b>0,7959</b> ± 0,1276	<b>0,8572</b> ± 0,0999
	ResNet-50	24,31	129	0,9558 ± 0,0371	0,7823 ± 0,1363	0,8431 ± 0,1098
	Inception V4	41,69	131	0,9549 ± 0,0370	0,7756 ± 0,1404	0,8366 ± 0,1181
	Densenet-121	7,75	190	0,9561 ± 0,0366	0,7838 ± 0,1351	0,8452 ± 0,1084
	MobileNet V2	2,27	92	0,9433 ± 0,0335	0,7120 ± 0,1997	0,7786 ± 0,1876
	EfficientNet B2	7,76	78	0,9465 ± 0,0320	0,7297 ± 0,1932	0,7941 ± 0,1777
LinkNet	VGG-16	15,98	147	<b>0,9619</b> ± 0,0357	<b>0,8143</b> ± 0,1139	<b>0,8706</b> ± 0,0863
	ResNet-50	31,18	151	0,9584 ± 0,0376	0,7857 ± 0,1397	0,8442 ± 0,1155
	Inception V4	46,16	69	0,9587 ± 0,0346	0,7851 ± 0,1393	0,8441 ± 0,1162
	Densenet-121	10,43	104	0,9590 ± 0,0363	0,7962 ± 0,1310	0,8547 ± 0,1050
	MobileNet V2	4,32	115	0,9543 ± 0,0397	0,7778 ± 0,1303	0,8390 ± 0,1073
	EfficientNet B2	7,89	111	0,9548 ± 0,0365	0,7767 ± 0,1410	0,8379 ± 0,1165
MA-Net	VGG-16	27,93	129	<b>0,9625</b> ± 0,0339	<b>0,8240</b> ± 0,1037	<b>0,8794</b> ± 0,0762
	ResNet-50	147,44	121	0,9592 ± 0,0357	0,7929 ± 0,1347	0,8508 ± 0,1080
	Inception V4	114,49	132	0,9554 ± 0,0397	0,7855 ± 0,1243	0,8456 ± 0,1012
	Densenet-121	44,77	159	0,9547 ± 0,0400	0,7833 ± 0,1323	0,8432 ± 0,1087
	MobileNet V2	48,89	122	0,9531 ± 0,0401	0,7735 ± 0,1420	0,8340 ± 0,1190
	EfficientNet B2	13,44	68	0,9590 ± 0,0331	0,7878 ± 0,1322	0,8466 ± 0,1113

é possível notar que a quantidade parâmetros é preponderante para alcançar resultados melhores em termos quantitativos, visto que os piores resultados também são unânimes para os *encoders* MobileNet V2 e EfficientNet B2.

Por outro lado, os melhores resultados foram com a combinação MA-Net e VGG-16. Esse resultado sugere que os blocos de atenção possuem papel fundamental para melhorar a capacidade de de extrair informações relevantes das imagens. Entretanto, há de se considerar que esses resultados melhores tiveram um considerável número de parâmetros se comparados com outros modelos avaliados. Ainda é importante observar que, apesar de uma arquitetura mais simplificada, a U-Net reporta resultados competitivos comparados à MA-Net, mesmo com menos parâmetros demandados. Entretanto, a U-Net demanda por mais épocas de treinamento para conseguir as melhores características responsáveis por realizar a segmentação. Outro destaque é que se considerar apenas o índice IoU, a U-Net reporta três dos seis melhores resultados para os *encoders* analisados, seguida pela MA-Net com dois resultados. Cada combinação de arquitetura e *encoders* com um mapa de segmentação para avaliação qualitativa de uma cena do conjunto de teste pode ser visualizado na Figura 15. Cada pixel está colorido de acordo com a classificação, sendo: verdadeiro positivo são brancos, verdadeiros negativos são pretos, falsos positivos são vermelhos e falsos negativos estão em laranja.

Independente da arquitetura, alguns *encoders* tendem a ter dificuldade de distinguir entre regiões de fundo e nuvens, o que gera uma grande proporção de falsos negativos e falsos positivos. Também se observa que, especificamente, o *encoder* VGG-16 combinado



com a arquitetura U-Net gera uma taxa maior de falsos positivos em relação aos outros modelos avaliados. Conforme já discutido, o desempenho do *encoder* VGG-16 é destacado, e isso é evidenciado por uma menor área com presença de falsos positivos e falsos negativos.

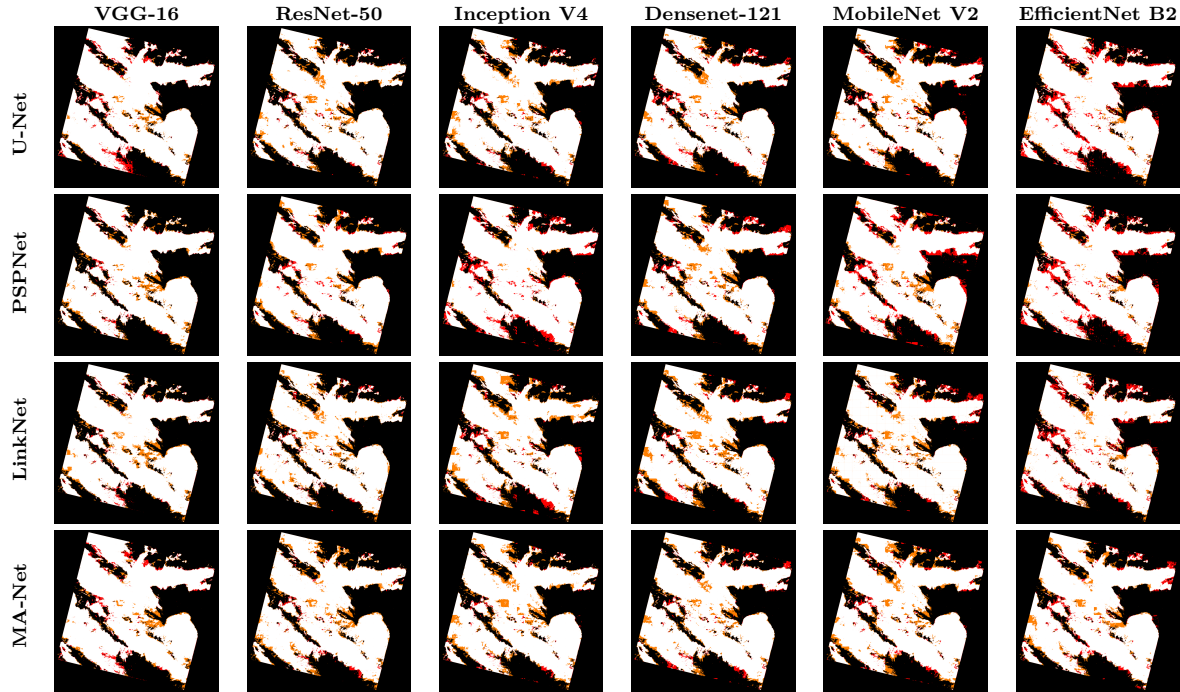


Figura 15 – Avaliação de segmentação de uma cena retirados do conjunto de teste.

A avaliação qualitativa da segmentação de uma cena a partir de uma região de interesse pode ser observada na Figura 16. Note que, apesar da arquitetura utilizada, todos os *encoders* focam em diferentes atributos das imagens, o que impacta os resultados da segmentação. Esse resultado corrobora o fato de que, independentemente da combinação de arquitetura e *encoder*, a tendência de classificar o fundo como nuvem é mais frequente do que o oposto. Essa tendência pode ser justificada pelo fato de as nuvens serem finas em determinadas áreas e, como consequência, informações de solo através da nuvem podem estar disponíveis, o que gera um padrão misto e que tende a confundir o modelo e o desempenho aquém para este cenário específico.

## 4.5 Considerações

A investigação de diferentes configurações experimentais compostas por arquiteturas e *encoders* para segmentação mostrou que a escolha de ambas tem impacto direto na segmentação de nuvens. Os *encoders* com maior profundidade, tais como, ResNet-50, Densenet-121 e VGG-16 apresentam desempenho substancialmente melhor independente da arquitetura considerada. Os melhores resultados foram observados na arquitetura MA-Net que incorpora mecanismos de atenção, o que tende a melhorar a segmentação. Outro destaque específico é a combinação MA-Net com *encoder* VGG-16 que tende a produzir

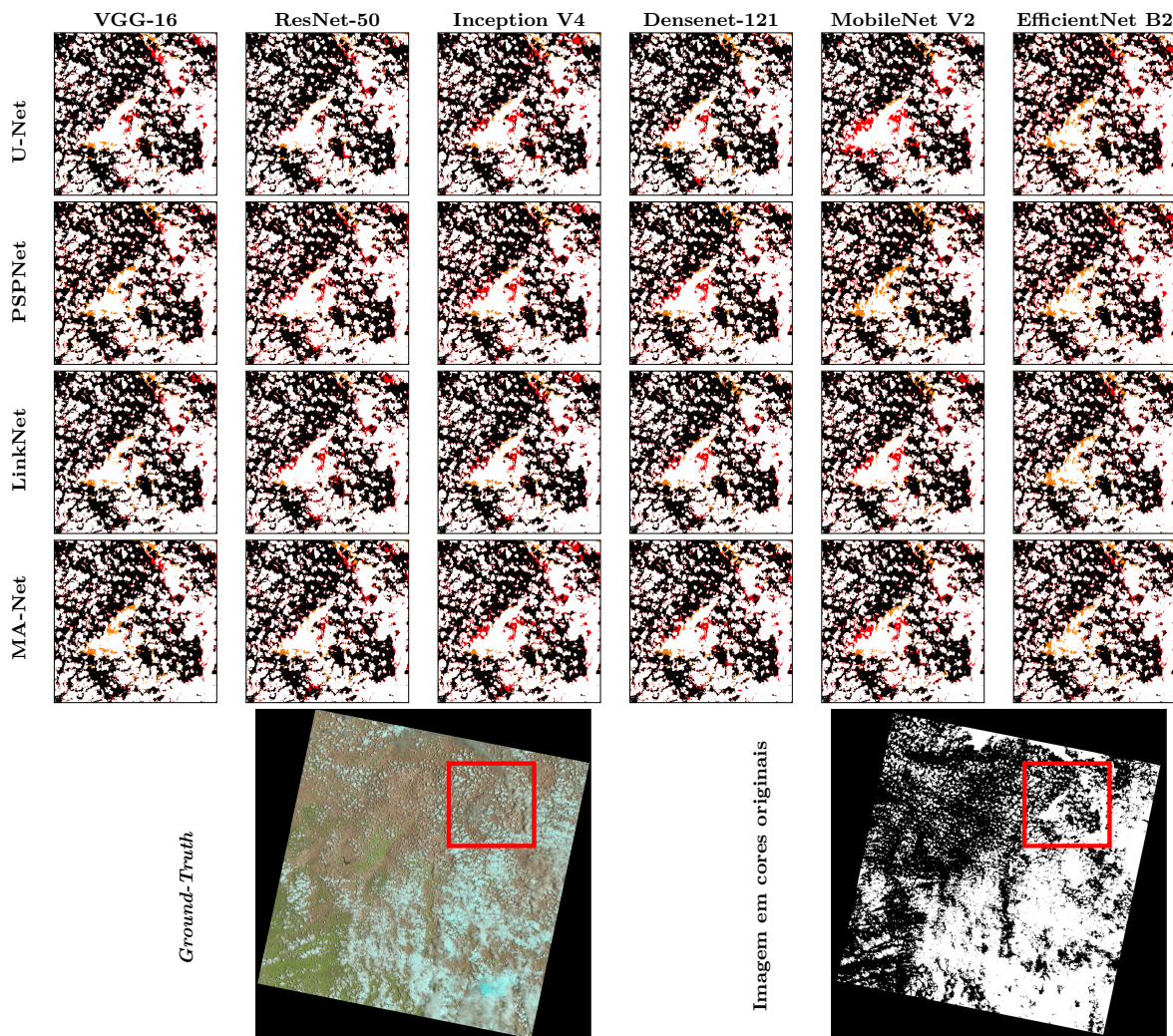


Figura 16 – Avaliação de segmentação da região de interesse de uma cena, retirados do conjunto de teste.

resultados com menos falsos positivos e falsos negativos, o que é interessante, visto que há um considerável desafio onde há nuvens finas com certa transparência de solo.

## Combinações de otimizadores e funções de perda para remoção de nuvens em modelos de difusão

Conforme exposto na revisão da literatura (Capítulo 3), vários pesquisadores têm empregado diversas abordagens para realizar a tarefa de remoção de nuvens, tais como CNN, GAN e, mais recentemente, modelos de difusão latente (DONG; WANG; ABBAS, 2021). Apesar destas abordagens terem características próprias, elas são baseadas em *deep learning*, logo treinar e ajustar esses modelos envolvem aspectos empíricos (BARBOSA et al., 2024). A escolha correta do otimizador e da função de perda é fundamental para aprimorar o processo de treinamento, atualização mais precisa dos pesos e, consequentemente, melhorar a capacidade de generalização e extrapolação (SEYREK; UYSAL, 2024). Assim, enquanto a escolha de otimizador e função de perda é crucial para redes convolucionais tradicionais, esses elementos carecem de investigação para técnicas emergentes, como modelos de difusão latente, ainda mais em cenários específicos como remoção de nuvens.

Deste modo, a partir da avaliação de nove modelos, investigou-se qual combinação de otimizador e função de perda produz melhores resultados para remoção de nuvens. Para isso, considerou-se o modelo DiffCR, proposto por Zou et al. (2024), como *baseline* para a investigação proposta.

### 5.1 Base de Dados

O SEN2\_MTC\_New é uma base de dados proposta por Huang e Wu (2022) que possui imagens nubladas em quatro bandas em três momentos temporais diferentes com o respectivo *ground-truth* (imagem sem nuvens). As quatro bandas considerados são azul, verde, vermelho e infra-vermelho próximo e contemplam diferentes regiões do planeta com consequente distintas condições climáticas e coberturas de solo. A base de dados possui

imagens de  $256 \times 256$ , em um total de 3417 imagens livres de nuvens com seus três pares multitemporais e é dividida originalmente na proporção de 70%, 20% e 10% para respectivamente treinamento, validação e teste. Exemplos de *patches* do SEN\_MTC\_New podem ser visualizados na Figura 17.

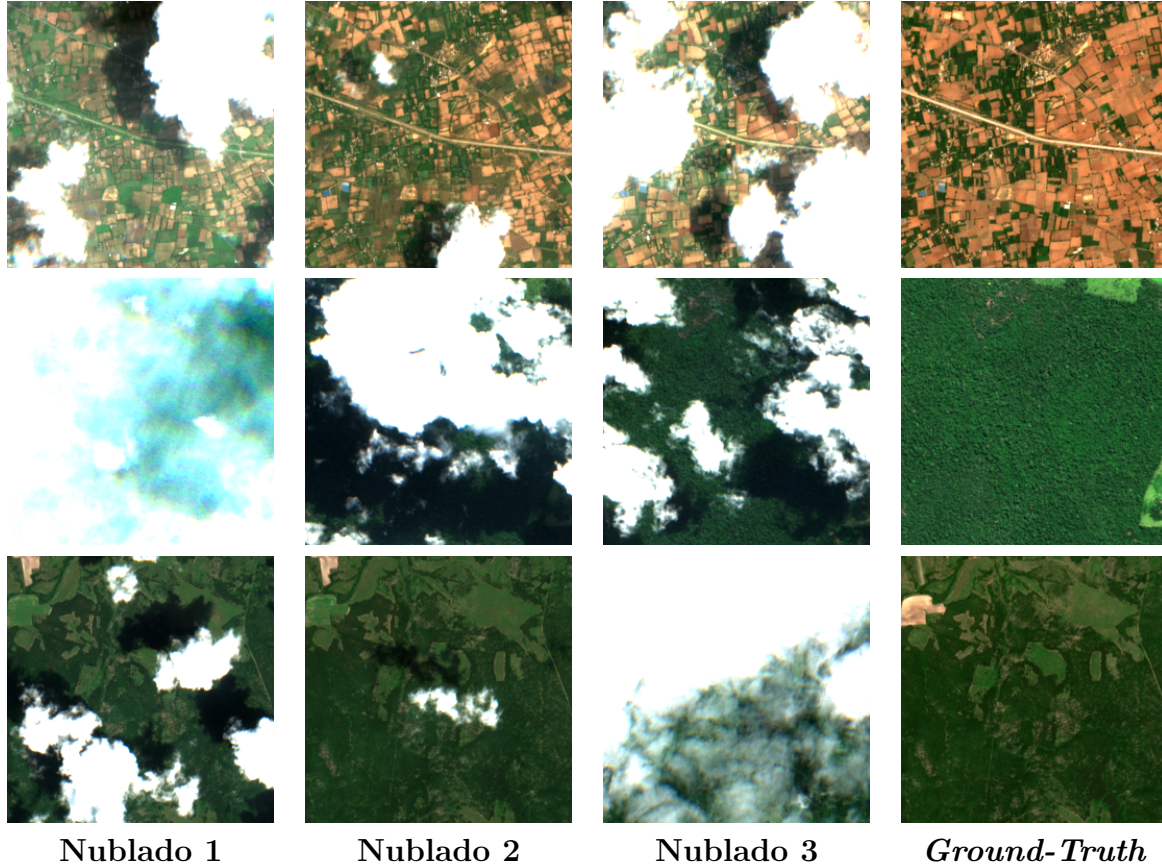


Figura 17 – Amostras da base de dados multitemporal SEN2\_MTC\_New nas bandas RGB (HUANG; WU, 2022).

Huang e Wu (2022), no mesmo trabalho, também propuseram uma nova abordagem baseada em GAN combinada com *Transformers*, denominada CTGAN<sup>1</sup>, para remoção de nuvens em imagens de satélite. A abordagem foi projetada para trabalhar em um contexto multiespectral com um extrator de características responsável por preservar as regiões sem nuvens, enquanto reduz o peso da influência da área nublada. Essa proposta garante uma reconstrução consistente e alinhada com aspectos globais da imagem considerada. Os autores avaliaram a abordagem no SEN2\_MTC\_New em termos de PSNR e SSIM e demonstraram a robustez da abordagem proposta em comparação à literatura e da importância da proposição da base de dados para tarefas deste domínio.

<sup>1</sup> <<https://github.com/come880412/CTGAN>>



## 5.2 DiffCR

O DiffCR<sup>2</sup> é uma abordagem baseada em difusão latente para remoção de nuvens proposta por Zou et al. (2024) que produz imagens sintéticas livres de nuvens através de um ruído gaussiano. A estrutura do DiffCR possui três componentes principais: (i) *encoder* condicional das imagens multitemporais, (ii) o modelo de *denoising* e (iii) o *encoder* temporal incorporado aos blocos que compõem o modelo de *denoising*.

A principal novidade é o uso de um novo bloco de operações, denominado TCFBlock, incorporado no mecanismo de *denoising*. Esse conjunto de operações consiste em uma alternativa aos mecanismos baseados em *Transformers*, e a utilização do TCFBlock possibilita reduzir consideravelmente o custo computacional, enquanto preserva a capacidade de reconstrução e, inclusive, supera outros métodos da literatura para remoção de nuvens. Outra característica do bloco proposto é a incorporação implícita de tempo, a qual realiza operações senoidais para preservar características temporais e, conseqüentemente, mantém informações temporais importantes para a reconstrução.

O DiffCR foi avaliado no SEN2\_MTC\_New e apresentou excelentes resultados em termos quantitativos e qualitativos. Desse modo, a abordagem de Zou et al. (2024) se consolida na área por produzir bom resultado em um custo computacional acessível, o que era um limitador de métodos de reconstrução.

## 5.3 Configuração Experimental

Para avaliar a melhor combinação de otimizadores e funções de perda para remoção de nuvens em modelo de difusão latente considerou-se três pontos principais: (i) o modelo DiffCR, (ii) os otimizadores SGD, Adam e AdamW e (iii) as funções de perda MAE, MSE e Huber. Para a função Huber, utilizou-se a configuração padrão do PyTorch<sup>3</sup> e um  $\delta = 1,0$ . Diante disso, foram conduzidos nove experimentos, onde todos possuíam uma taxa de aprendizagem de  $5 \times 10^{-5}$ , um decaimento de pesos de 0,01 para mitigar a possibilidade de *overfitting* e um tamanho de lote de 16. É importante mencionar que o decaimento de pesos é um hiperparâmetro que atua de forma diferente em cada otimizador e afeta diretamente na atualização dos pesos. Por exemplo, o decaimento de pesos impacta apenas os pesos no AdamW, já quando se considera o SGD e o Adam há impacto direto também nos gradientes. Por fim, para o SGD também se utilizou um *momentum* de 0,9.

O treinamento foi conduzido por 3000 épocas, com validação a cada 200 época, a fim de monitorar o processo de aprendizagem e eventual *overfitting*. A validação em todos os experimentos foi aferida de acordo com as perdas em termos de MAE, MSE e Huber. A proporção de divisão da base de dados SEN2\_MTC\_New considerou a divisão original de

<sup>2</sup> <<https://xavierjiezou.github.io/DiffCR/>>

<sup>3</sup> <<https://pytorch.org/docs/stable/nn.html#loss-functions>>

70%, 20% e 10% para respectivamente treinamento, validação e teste. Por fim, utilizou-se uma semente 42 em todos os experimentos para garantir a reprodutibilidade. Para testar os experimentos, considerou-se as métricas PSNR e SSIM. O projeto experimental deste estudo pode ser observado na Figura 18.

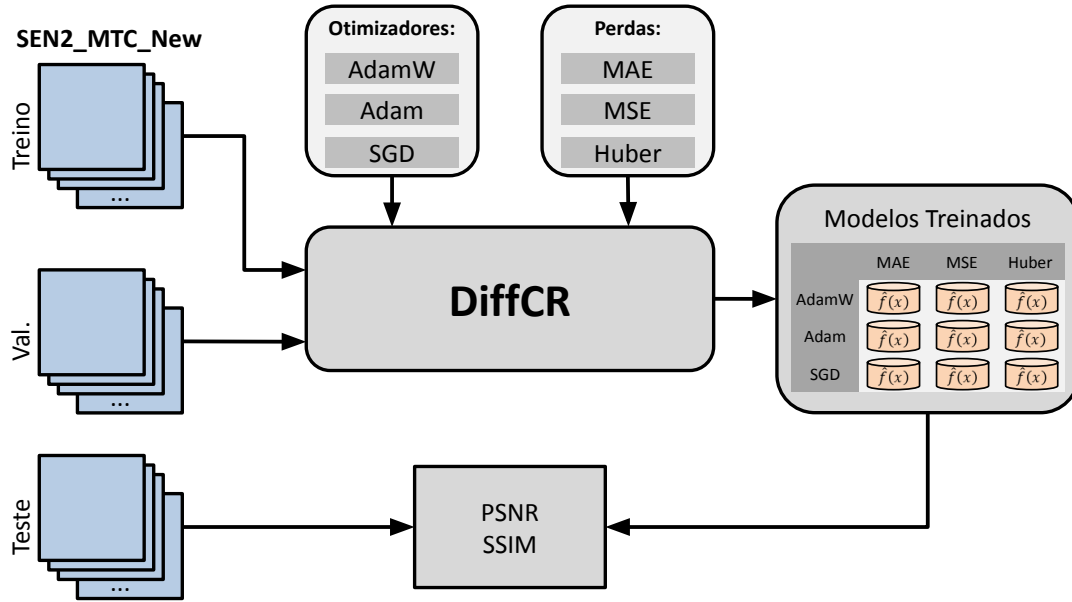


Figura 18 – Configuração experimental proposta para análise da combinação de otimizadores e função de perda.

Os experimentos foram realizados em um PC com CPU Core i5-12400 de 4,4 GHz e 32 GB de RAM com uma GPU NVIDIA RTX 4090 (memória de 24 GB). Os experimentos usaram a linguagem de programação Python 3.10.14 e as bibliotecas NumPy e Matplotlib para processamento numérico e visualização de imagens e dados. A biblioteca Scikit-learn foi usada para manipular o conjunto de imagens e analisar os resultados. A biblioteca usada para implementar os modelos foi PyTorch 2.3.1 e CUDA 12.1.

## 5.4 Resultados

A etapa de validação é importante para monitorar o processo de aprendizagem para entender o comportamento do modelo para a configuração experimental e, eventualmente, realizar ajustes baseando nesse conjunto. Como mencionado, considerou-se avaliação no conjunto de validação em termos das perdas de MAE, MSE e Huber, conforme pode ser visto na Figura 19. Essa figura indica que o modelo entrou em *overfitting* apenas na configuração SGD + MAE, entretanto, para as demais configurações há uma redução na perda da validação, o que sugere a ocorrência do processo de aprendizagem. Além disso, é possível observar que os modelos desempenham melhor com os otimizadores Adam

e AdamW para todas as funções de perda, onde, particularmente, o AdamW apresenta oscilações menos abruptas.

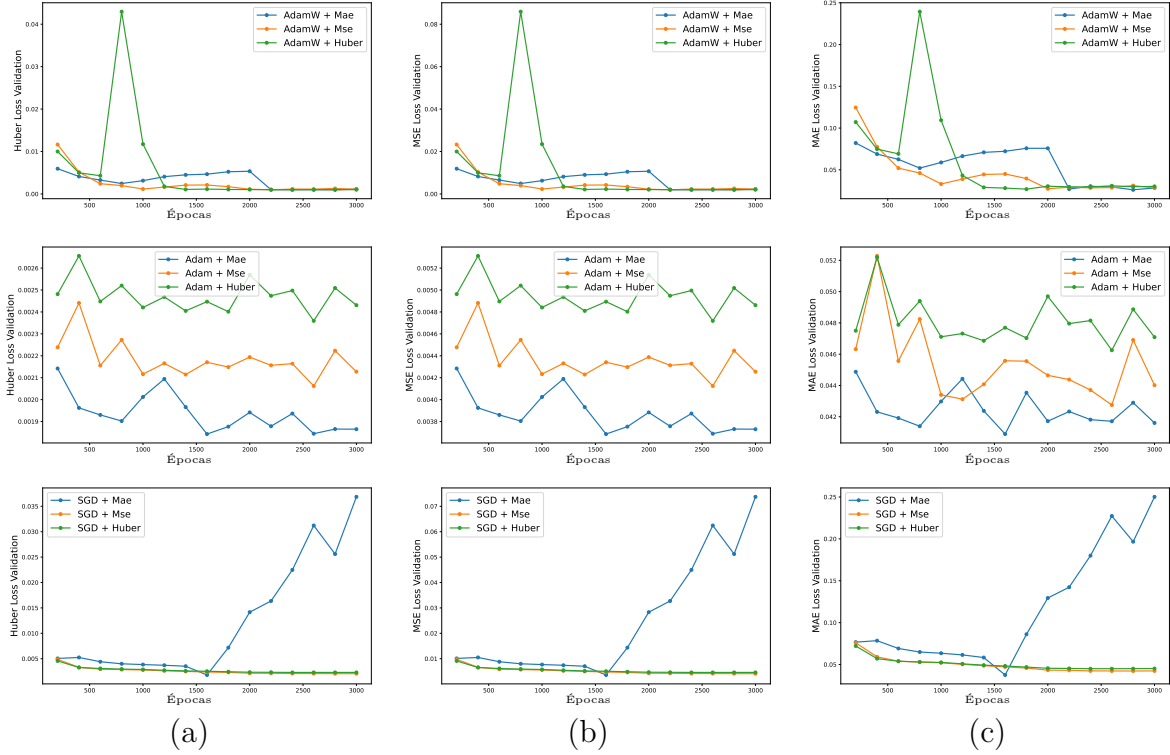


Figura 19 – Validação dos modelos a cada 200 épocas. Para cada gráfico, o eixo  $x$  representa o número de épocas e o eixo  $y$  representa a respectiva perda. As colunas (a), (b) e (c) apresentam a perda da validação em termos de Huber, MSE e MAE, respectivamente. Cada linha representa os otimizadores AdamW, Adam e SGD, respectivamente, com cada otimizador com as três funções de perda de treinamento.

Os resultados dos experimentos, em termos de PSNR e SSIM com média e desvio padrão podem ser observados na Tabela 5. Complementar à essa tabela, uma representação qualitativa destes resultados pode ser visualizada na Figura 20.

Em geral, os melhores resultados foram observados com o otimizador AdamW, que apresenta valores estáveis de PSNR e SSIM para as três funções de perda. Entretanto, é destacável que embora os valores de PSNR permaneçam relativamente próximos (entre 17-18 dB) para MAE e MSE, a função Huber apresenta tendência superior. Isso pode ser justificado pela penalização seletiva dos erros maiores o que contribui para um desempenho melhor, já que evita penalizações uniformes dos erros, independente da magnitude. Esse tipo de comportamento da função de perda pode ser eficiente para tratar remoção de nuvens em diferentes contextos (por exemplo, rios, florestas e áreas urbanas) em diversos tipos e densidades de nuvens, o que tende a gerar imagens de reconstrução mais realísticas. Além disso, esses resultados sugerem que a função Huber pode lidar melhor com imagens ruidosas ou discrepantes, o que pode explicar o aumento significativo no PSNR.

Para os experimentos com o otimizador Adam, a diferença entre as funções de perda é relativamente menor. Embora haja alguma variação, nenhuma configuração apresenta um desempenho significativamente melhor. Para o cenário de remoção de nuvens, é possível inferir que o Adam não se beneficia muito das funções de perda, o que sugere que este otimizador tem flexibilidade limitada para otimização e melhoria dos resultados. Outro destaque é que, ao contrário do AdamW, o pior resultado foi com a combinação com a função Huber, enquanto o melhor foi com o MAE, o que sugere que o Adam inseriu atuações abruptas durante o processo de treinamento. Em suma, Adam pode não ser tão eficaz para otimizar as métricas de qualidade de imagem para remoção de nuvens. Por fim, para o SGD é observado um desempenho estável e moderado em termos quantitativos, exceto quando combinado com a função MAE, quando resultou em *overfitting*.

Além dos pontos mencionados, observa-se que, em função da natureza heterogênea da base de dados, há um desvio padrão considerável entre todos os experimentos, o que evidencia o desafio de se considerar diferentes coberturas de solo. Assim, estudos adicionais sobre ajuste e comportamento de hiperparâmetros podem auxiliar a melhorar o desempenho em contextos tão variados, refinando o modelo para reduzir os desvios e aprimorar a generalização.

Tabela 5 – Resultados experimentais no conjunto de teste. Em negrito são os melhores resultados gerais para PSNR e SSIM, enquanto o sublinhado são os melhores resultados para cada otimizador considerado.

Otimizador	Função de Perda	PSNR $\uparrow$	SSIM $\uparrow$
AdamW	MAE	$18,0040 \pm 4,2008$	$0,5615 \pm 0,1148$
	MSE	$17,7674 \pm 4,1784$	$0,5614 \pm 0,1123$
	Huber	<b><u><math>18,4395 \pm 4,4706</math></u></b>	<b><u><math>0,6059 \pm 0,1245</math></u></b>
Adam	MAE	<u><math>15,2478 \pm 3,0464</math></u>	<u><math>0,4130 \pm 0,0892</math></u>
	MSE	<u><math>14,9229 \pm 2,9440</math></u>	<u><math>0,4103 \pm 0,0933</math></u>
	Huber	$14,4507 \pm 2,9423$	$0,4047 \pm 0,0959$
SGD	MAE	$5,3825 \pm 3,2556$	$0,2333 \pm 0,1045$
	MSE	<u><math>15,0421 \pm 3,1325</math></u>	<u><math>0,4147 \pm 0,0960</math></u>
	Huber	$14,5604 \pm 3,0442$	$0,4081 \pm 0,0979$

Para uma avaliação qualitativa, resultados de duas imagens para todos os experimentos podem ser visualizados nas Figuras 21 e 22. Essas figuras corroboram com os resultados quantitativos, onde é possível observar que diferentes coberturas de solo, densidade de nuvens e sobras, evidenciando os desafios de generalização da remoção de nuvens. Conforme já exposto, em todos os experimentos, lidar com áreas sombreadas é um desafio, semelhante à dificuldade observada com nuvens espessas. A combinação AdamW + Huber atinge um resultado visual que corresponde de perto ao resultado esperado, com as diferenças mais notáveis ocorrendo em áreas de cobertura de nuvens mais densas. As demais combinações apresentam problemas de reconstrução, incluindo perda de informações que



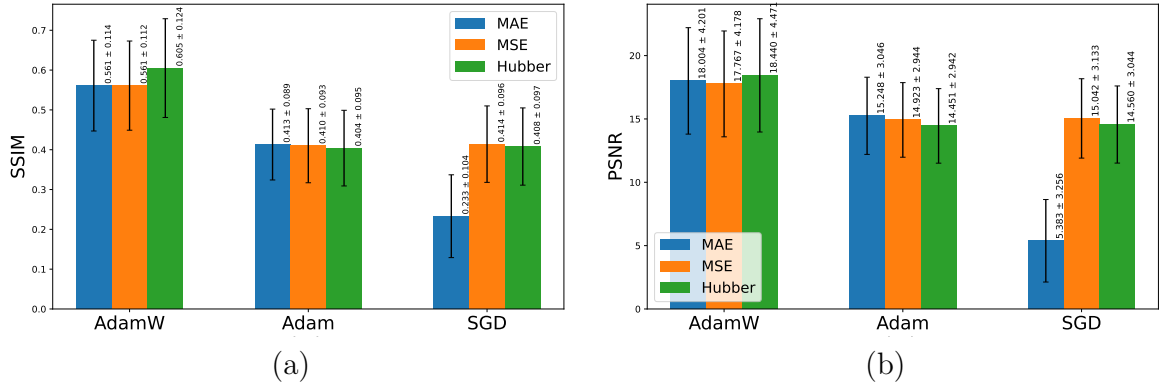


Figura 20 – Visualização e comparação dos resultados das métricas (a) SSIM e (b) PSNR no conjunto de teste para cada modelo.

eram originalmente visíveis e desobstruídas por nuvens. Além disso, outras combinações de otimizadores apresentam muita dificuldade para reconstruir com precisão aspectos da imagem, como tonalidade e textura, mesmo em regiões visíveis, com destaque para SGD + MAE que falha totalmente em gerar uma imagem coerente.

## 5.5 Considerações

A seleção de um otimizador combinado a uma função de perda tem impacto considerável em modelos de difusão latente. Para tarefas de remoção de nuvens, os experimentos demonstraram que essas escolhas afetam diretamente o desempenho do modelo, em termos positivos ou negativos.

Particularmente, a combinação da função de perda Huber com o otimizador AdamW apresentou eficiência para remoção de nuvens, principalmente quando se considera a complexidade do problema, o qual envolve variações de nuvens, coberturas de solo e neblinas, com cada um desses aspectos representando um desafio único para a reconstrução. O otimizador AdamW proporcionou um treinamento mais suave, enquanto a função de perda de Huber enfatizou efetivamente regiões com maior oclusão de nuvens, preservando áreas que já estavam livres de nuvens.

Por fim, este estudo considerou o DiffCR como modelo de difusão, logo há espaço para proposição de alternativas ao TCFBlock que apresentem desempenho melhor ou semelhante com um custo computacional menor, explorando as lacunas deixadas por essa abordagem. Além disso, este estudo pode fomentar estudos relacionados com busca de melhores hiperparâmetros em um espaço de busca, além da implementação de parada antecipada para evitar um treinamento tão longo.

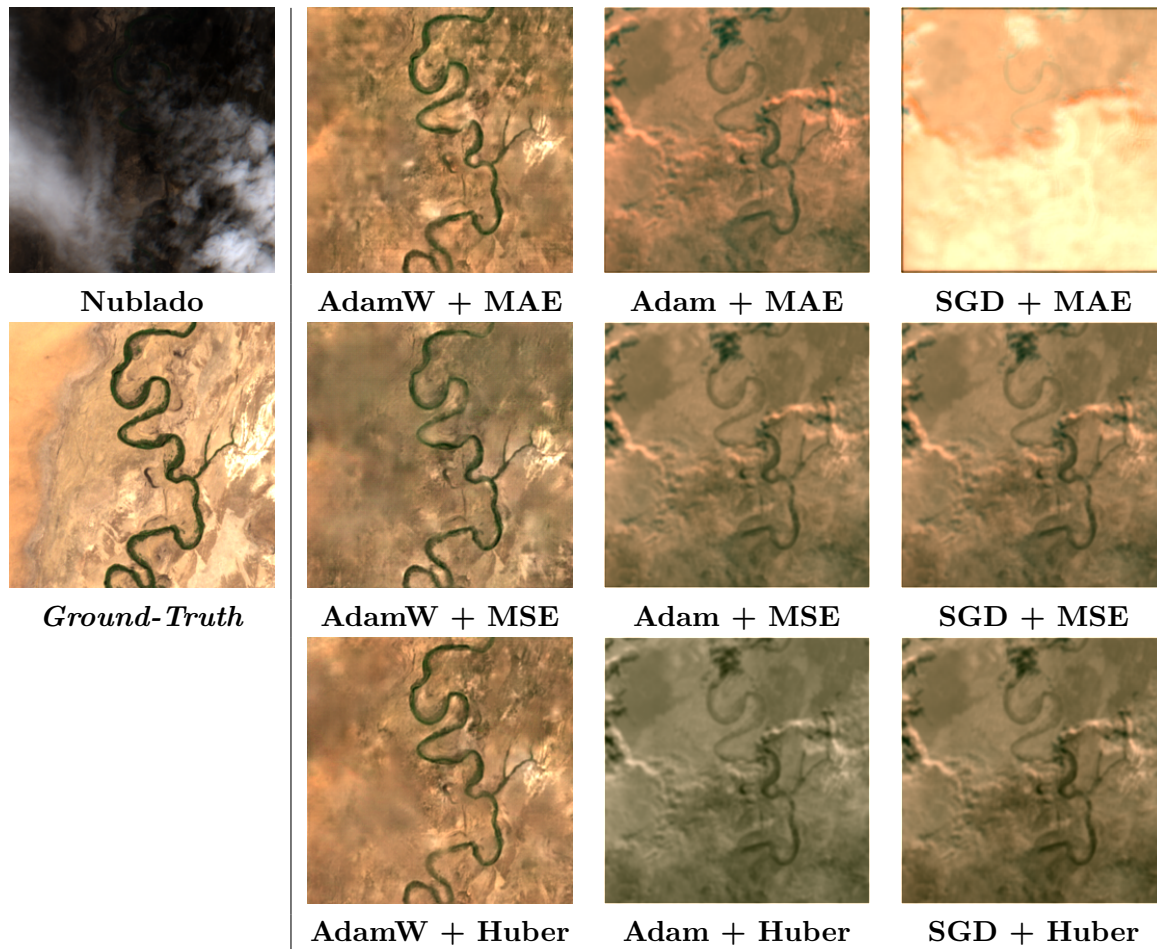


Figura 21 – Resultados qualitativos nas bandas RGB. A primeira coluna apresenta um dos três *patches* multitemporais com nuvens e respectivo *Ground-Truth*. Os outros *patches* apresentam visualizações de cada um dos nove experimentos deste trabalho.

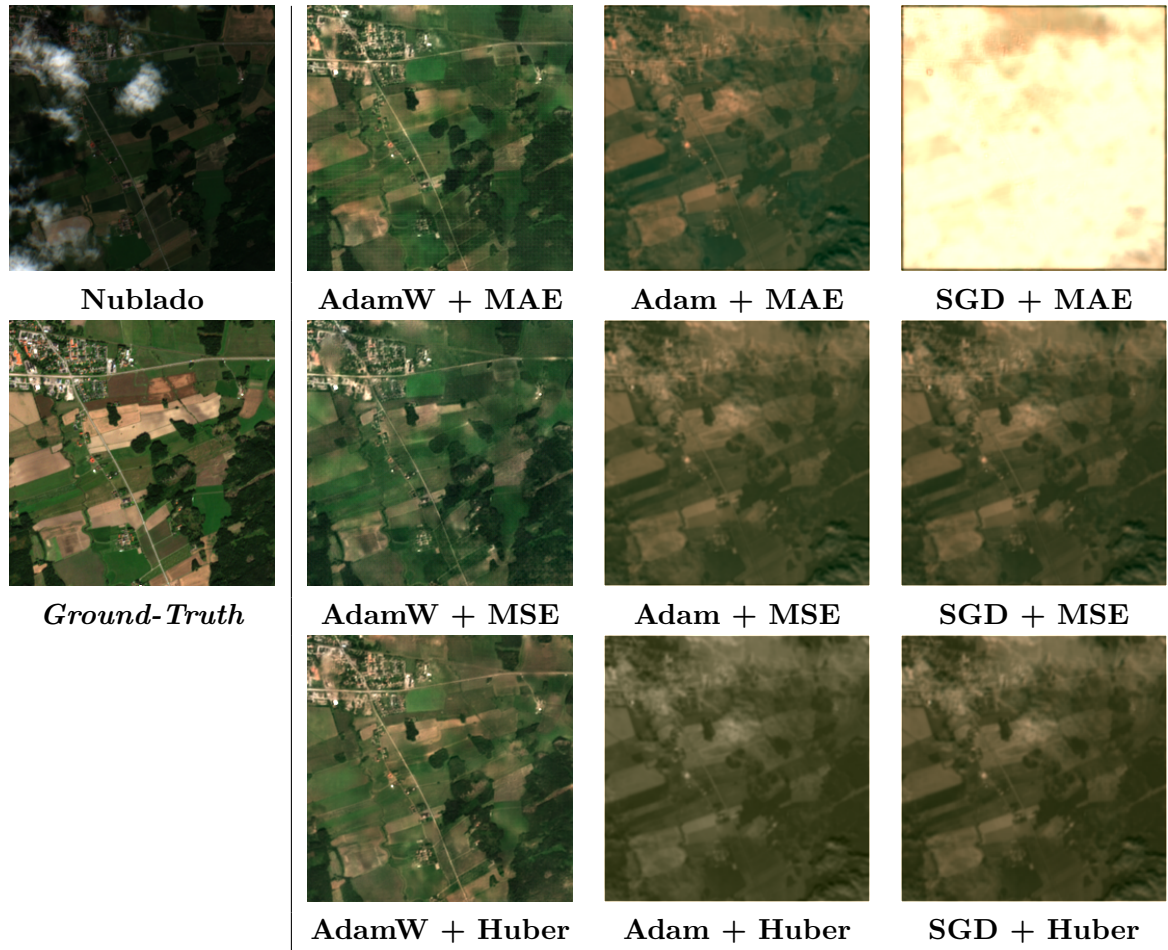


Figura 22 – Resultados qualitativos nas bandas RGB. A primeira coluna apresenta um dos três *patches* multitemporais com nuvens e respectivo *Ground-Truth*. Os outros *patches* apresentam visualizações de cada um dos nove experimentos deste trabalho.



# Modelos de difusão compactos para remoção de nuvens

Apesar de o DiffCR se mostrar robusto e eficiente para realizar a tarefa de remoção de nuvens, há espaço para investigação de blocos alternativos e alterações na estrutura do modelo de difusão latente para esta tarefa. Baseando-se nessa motivação, este capítulo apresenta uma nova abordagem baseada em blocos residuais e canais de atenção para compor um modelo de *denoising* para realizar efetivamente a tarefa de remoção de nuvens, através de um baixo custo computacional e sem a dependência explícita de incorporação temporal nos referidos blocos. Além disso, conforme já investigado, há um impacto considerável da função de perda para guiar o modelo no processo de aprendizado e, neste sentido, investigou-se a abordagem proposta sob a perspectiva das funções de perda MAE, MSE e Huber.

## 6.1 Abordagem Proposta

A abordagem proposta consiste em três pilares: (i) os *encoders* condicional e multiescala, como proposto por Zou et al. (2024), (ii) modelo de *denoising* e (iii) a função de perda que guia o modelo para realizar a reconstrução.

O *encoder* condicional empilha imagens multitemporais para que o modelo possa tratá-las de forma única e, conseqüentemente, explorar as informações de cada captura temporal considerada. Para os experimentos deste trabalho, considerou-se três imagens com três canais cada, o que totaliza, portanto, nove canais empilhados. Já o *encoder* condicional multiescala extrai as características espaciais das imagens nubladas, principalmente para aspectos multitemporais, de modo a utilizar as informações disponíveis de todos os momentos temporais considerados para gerar imagens sem nuvens com alta fidelidade (ZOU et al., 2024).

O modelo de *denoising* é inspirado na U-Net com as modificações nos blocos para realizar a remoção de nuvens com um custo computacional reduzido (RONNEBERGER;

FISCHER; BROX, 2015). A abordagem é composta por canal de atenção e operações residuais em uma arquitetura composta de três blocos de *encoder*, um bloco intermediário e três blocos de *decoder* (TREBING; STANCZYK; MEHRKANOON, 2021). Os blocos de *decoder* foram construídos através da operação de *PixelShuffle*, que tem se demonstrado mais robusta em relação à operação de convolução transposta em termos de mitigar perda de informação (SHI et al., 2016).

Cada bloco da abordagem proposta inicia-se com duas convoluções com filtros de  $3 \times 3$ , *stride* e *padding* de 1 para preservar a resolução espacial de entrada. Entre as duas camadas há a aplicação de uma função de ativação *Sigmoid Linear Unit* (SiLU) que combina aspectos suaves e diretos da não linearidade para obter gradientes melhores no treinamento. Após este primeiro processo convolucional, os atributos passam por um canal de atenção para enfatizar as características mais importantes de cada um, ainda mais em um contexto em que há uma entrada com três imagens multitemporais empilhadas (GUO et al., 2024).

O canal de atenção recalibra a importância de cada canal para o *tensor* de entrada e, desse modo, o modelo pode focar nos canais onde há informação mais relevante. Esse processo se inicia com uma operação de *pooling* global por canal para combinar informação de cada canal em um único valor para produzir uma representação global de cada canal. Na sequência duas operações de convolução são aplicadas, sendo essas operações intermediadas por uma função de ativação *LeakyReLU* para possibilitar ao modelo capturar relacionamento não linear entre os canais. A escolha dessa função visa impedir que os neurônios assumam valores nulos, o que poderia resultar na sua neutralização. Por fim, uma função de ativação *sigmoid* é aplicada para um intervalo entre 0 e 1 e, desse modo, é possível multiplicar esses pesos por canal e enfatizar a importância de cada um, baseando-se nessa ponderação (HU; SHEN; SUN, 2018). Essas operações são apresentadas de forma resumida na Figura 23(a).

Após o canal de atenção, a abordagem é direcionada para o bloco residual (LIU et al., 2024a), onde o conjunto de operações inicia com duas camadas de convolução com filtros de  $3 \times 3$ , *stride* e *padding* de 1. A primeira convolução extrai as características espaciais e contextuais e a segunda é responsável por refinar tais características, preservando as dimensões espaciais. Entre as operações de convolução há uma função de ativação *Rectified Linear Unit* (ReLU) para possibilitar obter aspectos de não-linearidade. Este bloco possui uma conexão de atalho que é responsável por replicar a entrada e adicionar à saída das convoluções. Essa operação tende a preservar a identidade da entrada e auxilia na propagação do gradiente. Por fim, a abordagem adiciona as saídas dos caminhos principal e de atalho, concatenando-as e seguindo por outra função de ativação ReLU para refinar as características combinadas (HE et al., 2016), conforme Figura 23(b).

A partir das estruturas do canal de atenção e bloco residual, é possível construir o modelo de *denoising* de diferentes formas e, assim, avaliou-se três tipos de modelos de

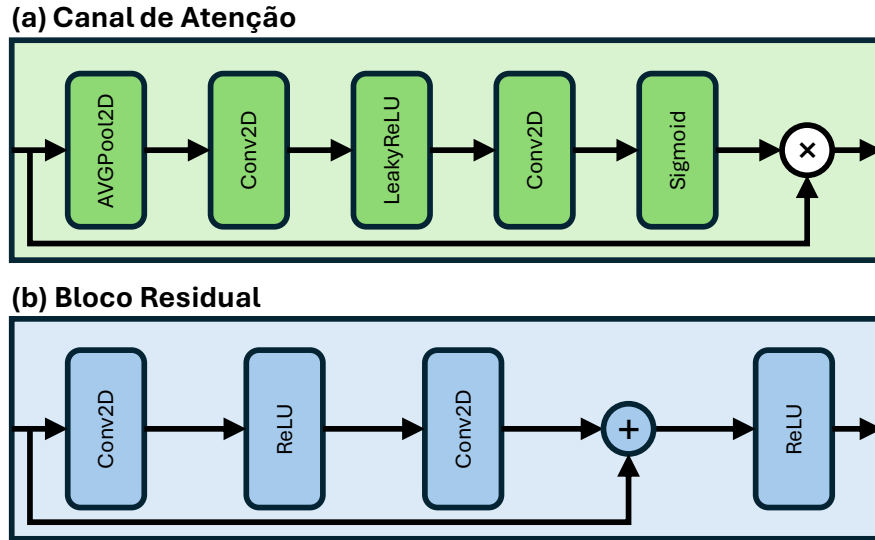


Figura 23 – Esquema dos blocos propostos para o modelo de *denoising*. (a) Canal de Atenção: recalibra a importância dos canais via *pooling* global, convoluções com *LeakyReLU* e ativação *sigmoid*. (b) Bloco Residual: utiliza convoluções com ReLU e conexão de atalho, seguido por concatenação e nova ativação ReLU.

*denoising*. O primeiro tipo, denominado *Channel Attention Block* (CAB) (Figura 24(a)), possui operações de atenção em todos os blocos de *encoder*, intermediário e *decoder*. O segundo tipo, denominado *Channel Attention with Residual Block* (CABRes) (Figura 24(b)), possui operações de atenção e residuais em todos os blocos de *encoder*, intermediário e *decoder*. Por fim, o terceiro tipo, CAB + CABRes, possui os blocos de *encoder* e *decoder* com canal de atenção e o bloco intermediário com operações residuais e atenção.

A função de perda, que desempenha um papel importante para medir a diferença entre a saída do modelo e a imagem de referência, é fundamental em um cenário onde diferentes coberturas de solo estão envolvidas. Para tanto, avaliou-se a combinação de cada um dos modelos com as funções de perda MAE, MSE e Huber (com  $\delta = 1,0$ ). A depender da função usada no processo de treinamento, o erro é propagado através do *backpropagation* para que os parâmetros da rede sejam ajustados e, dessa forma, minimizar a perda e melhorar a remoção de nuvens. Por fim, a abordagem proposta é treinada, validada e testada através da base de dados SEN2\_MTC\_New (HUANG; WU, 2022) com divisão original de 70%, 20% e 10% para, respectivamente, treinamento, validação e teste. A abordagem proposta pode ser visualizada na Figura 25.

## 6.2 Configuração Experimental

Foram conduzidos para três conjuntos de experimentos para avaliar o impacto dos blocos propostos para remoção de nuvens, onde cada um deles é combinado com uma das

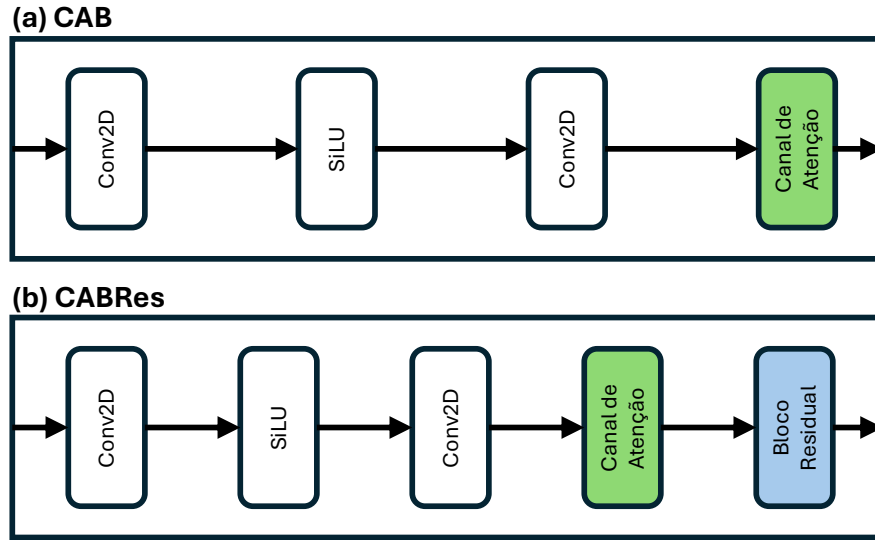


Figura 24 – Estrutura dos blocos CAB e CABRes. (a) CAB: composto por duas convoluções seguidas por ativação SiLU, e blocos com canal de atenção. (b) CABRes: possui a mesma estrutura do modelo CAB, com a adição de conexões residuais.

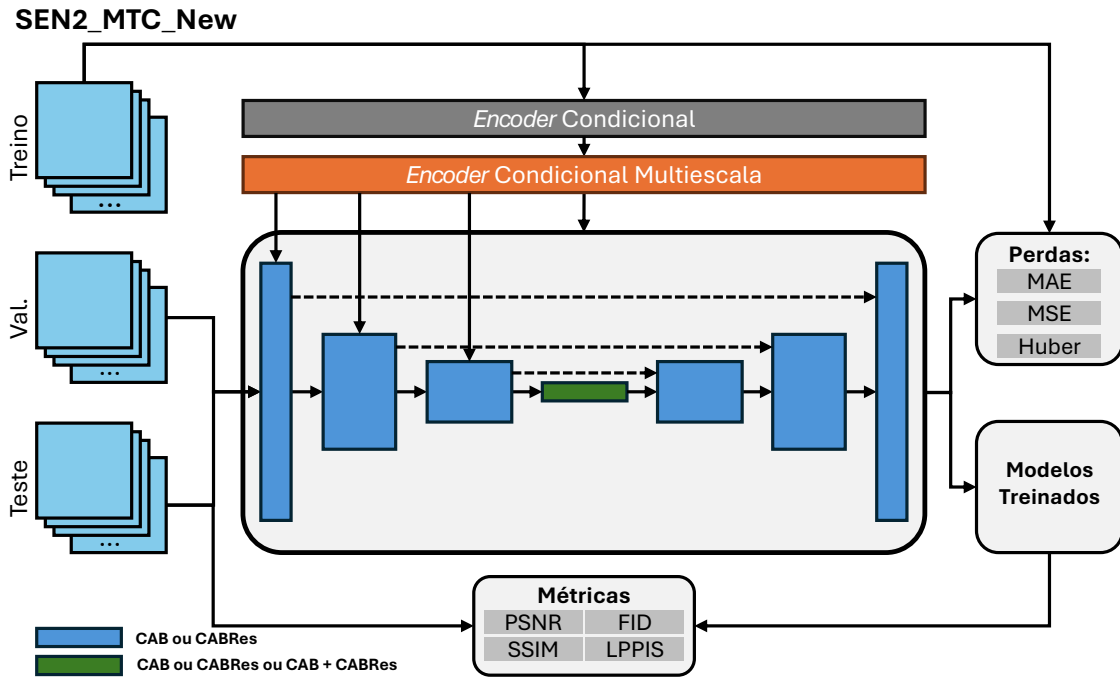


Figura 25 – Arquiteturas com três tipos de modelos de *denoising* propostos. CAB utiliza blocos com canal de atenção em todas as etapas. O modelo CABRes incorpora operações de atenção e conexões residuais em todas as etapas. Por fim, o modelo CAB + CABRes combina blocos de atenção nas etapas de *encoder* e *decoder*, e blocos com atenção e residuais na etapa intermediária. Cada modelo é treinado através de três funções de perda diferentes.



três funções de perda consideradas. Assim, avaliou-se nove experimentos no total. A partir dos estudos realizados sobre otimizadores, decidiu-se utilizar exclusivamente o otimizador AdamW, com uma taxa de aprendizagem de  $5 \times 10^{-5}$ , um decaimento de pesos de 0,01 para mitigar a possibilidade de *overfitting* e um tamanho de lote de 12. O treinamento foi conduzido por 3000 épocas, com validação a cada 200 para verificar e avaliar o processo de aprendizagem, sendo a perda da validação também medida em termos de MAE, MSE e Huber (SILVA et al., 2025). Também se considerou 64 canais para as operações do modelo de *denoising*, uma semente 42 para garantir a reprodutibilidade e um esquema de ruído do tipo *sigmoid* para guiar o modelo de difusão. Além disso, de acordo com Haynes, Corns e Venayagamoorthy (2012) e Zou et al. (2024), também se utilizou a técnica de *Exponential Moving Average* (EMA) para refinar a convergência do modelo. O EMA mantém uma versão suavizada dos pesos do modelo ao longo das iterações, atualizando-os de acordo com a Equação 23.

$$\theta_{\text{EMA}} = \beta \cdot \theta_{\text{EMA}} + (1 - \beta) \cdot \theta, \quad (23)$$

onde  $\theta$  é o peso;  $\beta$  controla a taxa de atualização. Neste trabalho, definiu-se  $\beta = 0,999$ .

Os experimentos são avaliados no conjunto de testes em termos de PSNR, SSIM, *Fréchet Inception Distance* (FID) e *Learned Perceptual Image Patch Similarity* (LPIPS) considerando a última época de treinamento. PSNR, SSIM, já definidos, abordam aspectos de correlação e nível de ruído entre as imagens gerada e esperada. O FID e o LPIPS complementam a avaliação em termos perceptuais, de modo a tentar capturar as diferenças perceptuais para os olhos humanos e a distribuição estatística em um espaço latente. Essas métricas se consolidaram como avaliação de modelos generativos. O FID é definido de acordo com a equação 24.

$$\text{FID}(x, y) = \|\mu_x - \mu_y\|_2^2 + \text{Tr} \left( \Sigma_x + \Sigma_y - 2(\Sigma_x \Sigma_y)^{\frac{1}{2}} \right), \quad (24)$$

onde  $\mu_x$  e  $\mu_y$  são as médias dos vetores das distribuições de características para as imagens reais ( $x$ ) e geradas ( $y$ );  $\Sigma_x$  e  $\Sigma_y$  são matrizes de covariância das distribuições de características para imagens reais e geradas;  $\text{Tr}$  é a soma dos elementos na diagonal principal.

O LPIPS considera a soma das diferenças ponderadas de camadas em diferentes camadas em redes neurais pré-treinadas. Para essa métrica, considerou-se a AlexNet pré-treinado pela ImageNet. O LPIPS é definido de acordo com a equação 25.

$$\text{LPIPS}(x, y) = \sum_l w_l \cdot \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \|\phi_l(x)_{h,w} - \phi_l(y)_{h,w}\|_2^2, \quad (25)$$

onde  $l$  é o índice de camadas na rede neural usada para extração de características;  $w_l$  é o peso atribuído à camada  $l$ ;  $H_l$  e  $W_l$  são a altura e a largura da camada  $l$ ;  $h$  e  $w$  são índices de pixels nas características extraídas;  $\phi_l(x)$  e  $\phi_l(y)$  são representações de características extraídas da camada  $l$  para imagens esperadas ( $x$ ) e preditas ( $y$ ).

Por fim, avaliou-se as abordagens propostas para cada métrica considerada através da Fronteira de Pareto (RAHMAN; SZABÓ, 2021). Como *baseline* para avaliação, selecionou-se o DiffCR para uma avaliação em termos quantitativos relativos ao custo computacional. Diante disso, para cada modelo que atingiu a Fronteira de Pareto em pelo menos uma das métricas, continuou-se o treinamento até a época 5000 para refinar e auxiliar o processo de reconstrução da área nublada, sendo essa última época também considerada para avaliação no conjunto de testes.

Além disso, realizou-se uma etapa de pós-processamento para melhorar a qualidade visual das imagens, ajustando o contraste e a nitidez (SILVA et al., 2023). Especificamente para o pós-processamento considerou-se um ajuste sutil por meio de filtro de realce de bordas para nitidez (fator de *sharpness* de 1,02) para destacar detalhes sem introduzir ruído excessivo e outro ajuste de contraste de um fator de 0,9 para suavizar transições abruptas. Assim, os modelos considerados após a Fronteira de Pareto são melhorados por reduzir pequenas discrepâncias entre as imagens reconstruída e a referência. Além disso, leves ajustes de textura e contraste minimizam diferenças perceptuais nas imagens reconstruídas que surgem durante o processo de treinamento.

Os experimentos também foram realizados em um PC com CPU Core i5-12400 de 4,4 GHz e 32 GB de RAM com uma GPU NVIDIA RTX 4090 (memória de 24 GB). Os experimentos usaram a linguagem de programação Python 3.10.14 e as bibliotecas NumPy e Matplotlib para processamento numérico e visualização de imagens e dados. A biblioteca Scikit-learn foi usada para manipular o conjunto de imagens e analisar os resultados. A biblioteca usada para implementar os modelos foi PyTorch 2.3.1 e CUDA 12.1.

## 6.3 Resultados

As perdas da validação, preponderantes para acompanhar o processo de aprendizagem dos modelos e evitar o vazamento de dados, são consideradas conforme visualizado na Figura 26. Observa-se redução ao longo do processo de treinamento até uma estabilização, o que sugere um processo efetivo de aprendizado ocorrendo em todos os modelos avaliados.

A avaliação quantitativa para o conjunto de testes é apresentada na Tabela 6 e demonstra a capacidade das abordagens experimentais em realizar a remoção de nuvens em baixo custo computacional, mensurado através do número de parâmetros e de *Multiply-Accumulate Operations* (MACs). Complementar ao resultado quantitativo, as Figuras 27, 28 e 29 apresentam os resultados da reconstrução corroborando os resultados reportados e demonstrando a capacidade dos modelos treinados. Esses resultados demonstram que a diferença de cobertura consegue ser absorvida no treinamento dos modelos. Nesse ponto, é possível destacar que os experimentos do tipo CAB + CABRes possuem melhoria limitada em relação ao tipo CAB e essa avaliação é ainda mais evidente quando considera-se que

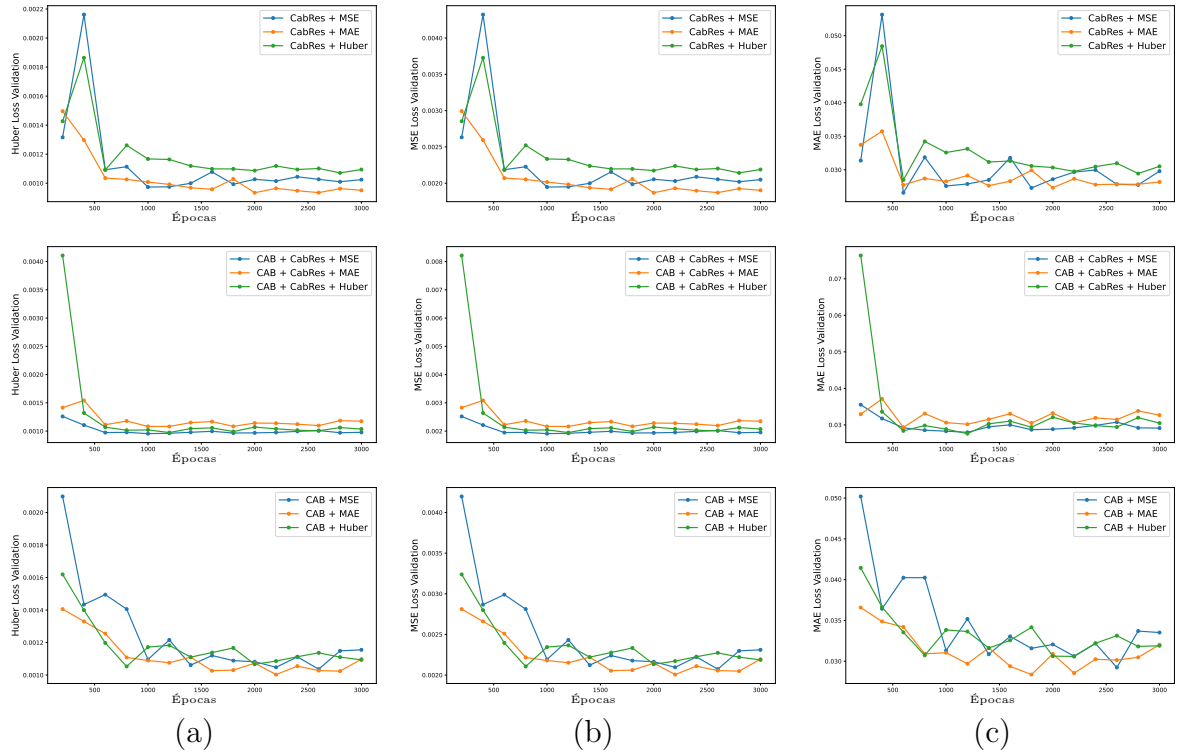


Figura 26 – Validação dos modelos a cada 200 épocas. O eixo  $x$  representa o número de épocas e o eixo  $y$  representa a respectiva perda. As colunas (a), (b) e (c) apresentam a perda da validação em termos de Huber, MSE e MAE, respectivamente. Cada linha representa os experimentos dos conjuntos CABRes, CAB + CABRes e CAB, com cada conjunto possuindo três funções de perda de treinamento.

CAB + CABRes demanda aproximadamente o dobro de parâmetros treináveis e MACs em relação ao CAB.

Tabela 6 – Comparação de resultados entre as abordagens propostas. Em negrito, estão os melhores resultados para todos os experimentos considerados em todas as métricas, e o sublinhado refere-se ao melhor resultado para cada grupo de experimentos considerados.

Modelo	PSNR $\uparrow$	SSIM $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$	# Params (M) $\downarrow$	MACs (G) $\downarrow$
CABRes + MSE	18,380	<b>0,636</b>	<b>95,845</b>	<b>0,335</b>	11,90	97,65
CABRes + MAE	18,487	0,632	100,900	0,342	11,90	97,65
CABRes + Huber	<b>18,495</b>	0,632	99,273	0,339	11,90	97,65
CAB + CABRes + MSE	18,009	0,610	116,614	0,373	8,81	39,61
CAB + CABRes + MAE	17,819	<u>0,611</u>	121,756	0,379	8,81	39,61
CAB + CABRes + Huber	17,938	<u>0,611</u>	117,715	0,375	8,81	39,61
CAB + MSE	17,766	0,609	112,287	0,374	<b>4,09</b>	<b>29,94</b>
CAB + MAE	17,969	<u>0,614</u>	113,839	0,384	<b>4,09</b>	<b>29,94</b>
CAB + Huber	<u>17,978</u>	<u>0,614</u>	<u>109,562</u>	<u>0,368</u>	<b>4,09</b>	<b>29,94</b>

De acordo com a configuração proposta, avaliou-se os experimentos que se mostraram mais promissores em reduzir o custo computacional para cada métrica, considerando o

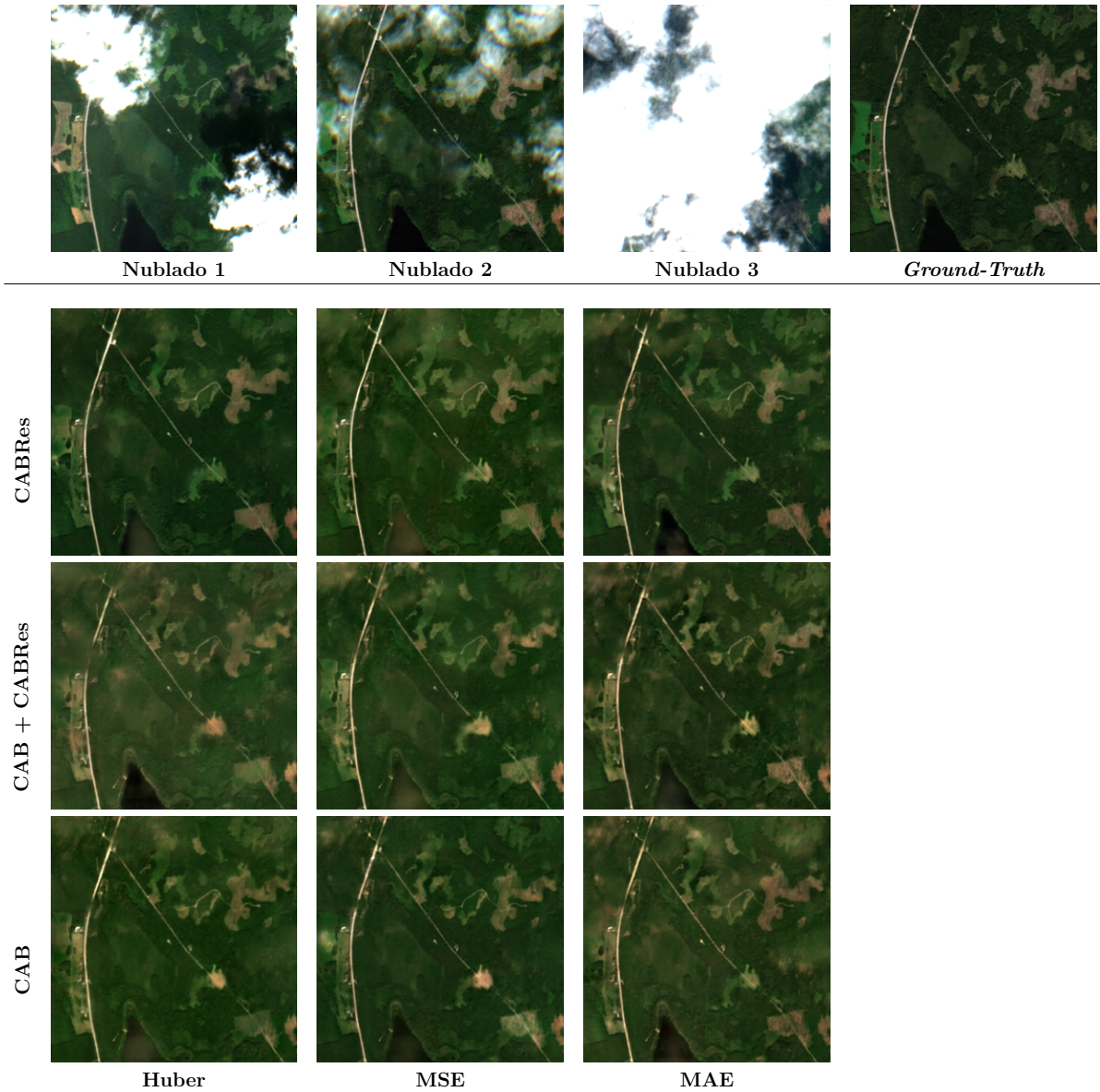


Figura 27 – Resultados qualitativos considerando áreas florestais, estradas e lagos.

DiffCR como *baseline*. Os resultados do DiffCR que motivaram este estudo estão sintetizados na Tabela 7. A partir da aplicação da Fronteira de Pareto é possível definir quais são esses modelos mais promissores e eles estão descritos na Tabela 8. Os modelos que atingiram a fronteira para as métricas baseadas em similaridade estrutura e diferença de intensidade (PSNR e SSIM) e as métricas perceptuais (FID e LPIPS) podem ser observados nas Figuras 30 e 31. É importante reforçar que o objetivo do primeiro conjunto de métrica é maximizá-las, enquanto para o segundo conjunto objetiva-se minimizá-las. Aqui, destacam-se os modelos CABRes + MSE e CAB + Huber que atingiram a fronteira em três das quatro métricas consideradas.

A partir da avaliação pela Fronteira de Pareto realizou-se estudos com objetivo de melhorar a capacidade de generalização através de ajustes de refino para que os modelos pudessem ser explorados em sua capacidade máxima. O pós-processamento possibilita,



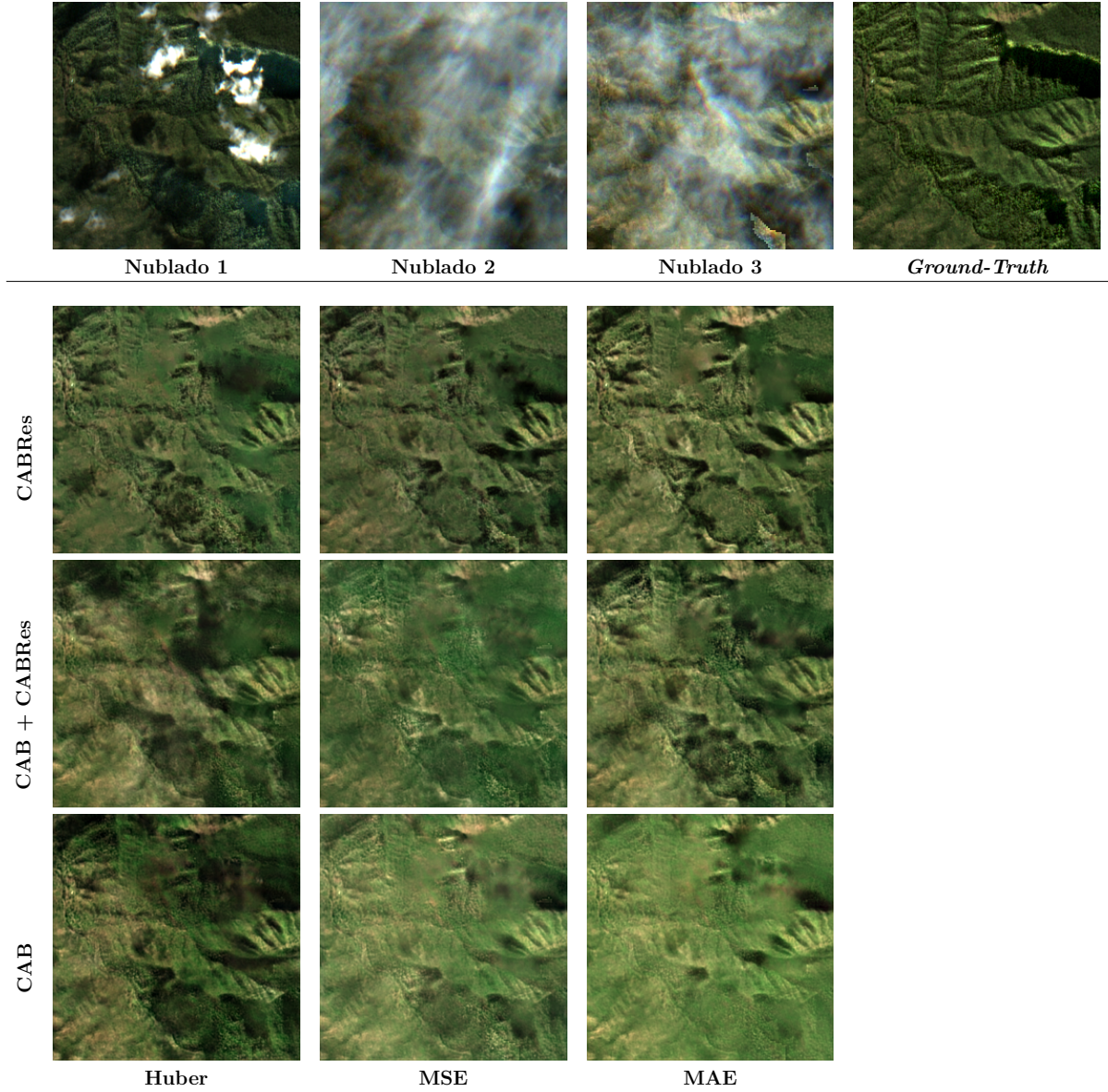


Figura 28 – Resultados qualitativos considerando regiões montanhosas e áreas de floresta.

portanto, melhorar a qualidade visual das imagens para que tenham uma transição mais natural de partes nubladas para regiões reconstruídas sem nuvens. O comportamento dos modelos em termos quantitativos após a otimização baseada na Fronteira de Pareto pode ser observado na Tabela 9. Na Figura 32, através de um exemplo de *patch* aleatório, é possível visualizar o antes e o depois para cada modelo otimizado através da estratégia da Fronteira de Pareto. Por fim, uma análise comparativa entre os resultados dos experimentos otimizados e a literatura, destacando o desempenho das abordagens propostas, principalmente em termos de custo computacional, pode ser observada na Tabela 10.

Embora o DiffCR apresente um desempenho superior em PSNR e SSIM, ele possui um custo computacional significativamente maior. Em contraponto, as abordagens propostas apresentam um equilíbrio melhor entre qualidade e eficiência e, portanto, mais adequadas em cenários com restrições de recursos computacionais. Além disso, as abordagens pro-

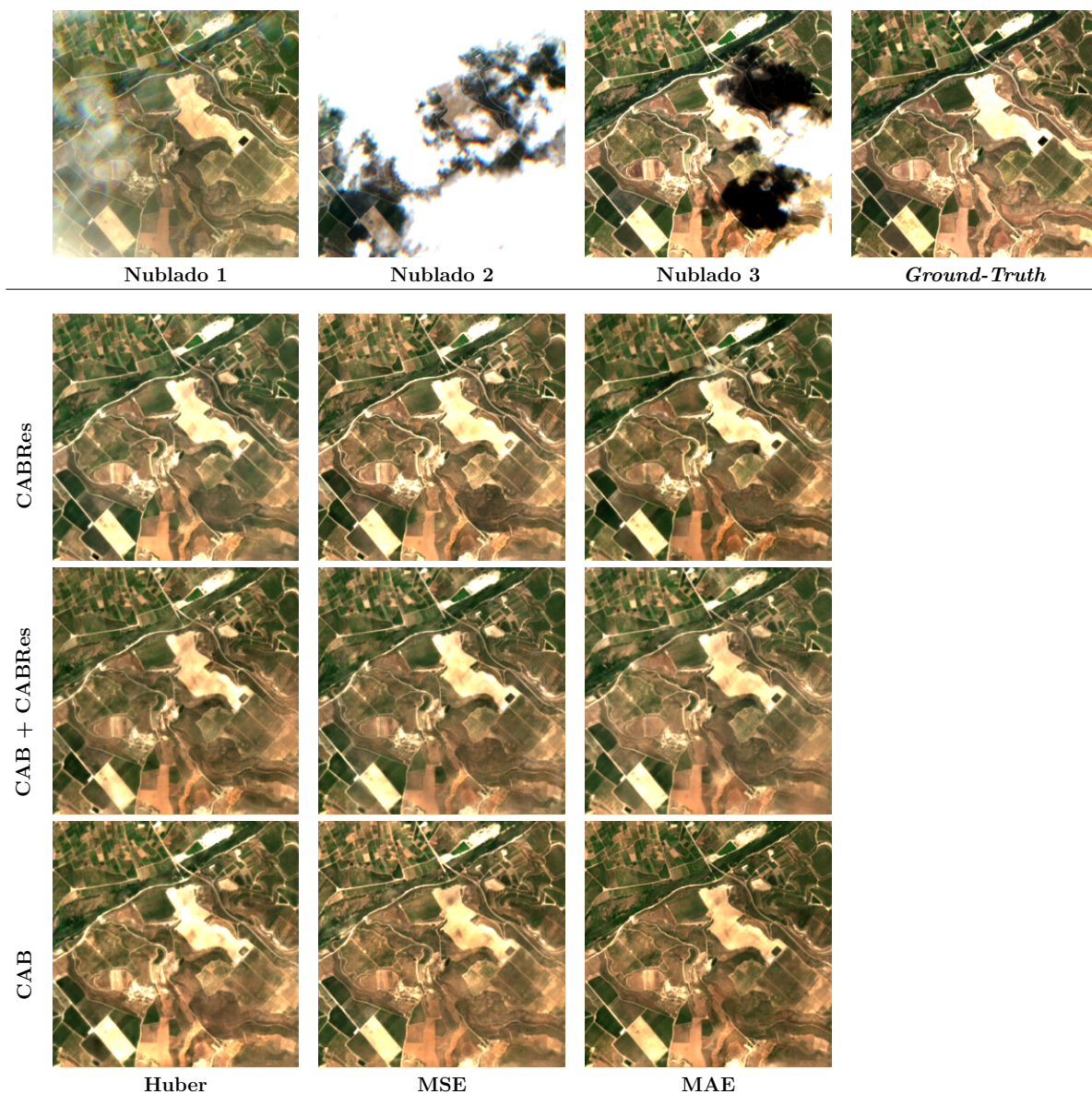


Figura 29 – Resultados qualitativos considerando áreas com terra, áreas plantadas e estradas.

Tabela 7 – Resultados do DiffCR.

	PSNR $\uparrow$	SSIM $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$	# Params (M) $\downarrow$
DiffCR (ZOU et al., 2024)	19,150	0,671	83,162	0,291	22,91

Tabela 8 – Resumo dos modelos presentes na Fronteira de Pareto para as métricas consideradas.

Métrica	Modelos na Fronteira de Pareto
<b>PSNR</b>	CABRes + Huber, CAB + CABRes + MSE, CAB + Huber
<b>SSIM</b>	CABRes + MSE, CAB + MAE
<b>LPIPS</b>	CABRes + MSE, CAB + Huber
<b>FID</b>	CABRes + MSE, CAB + Huber

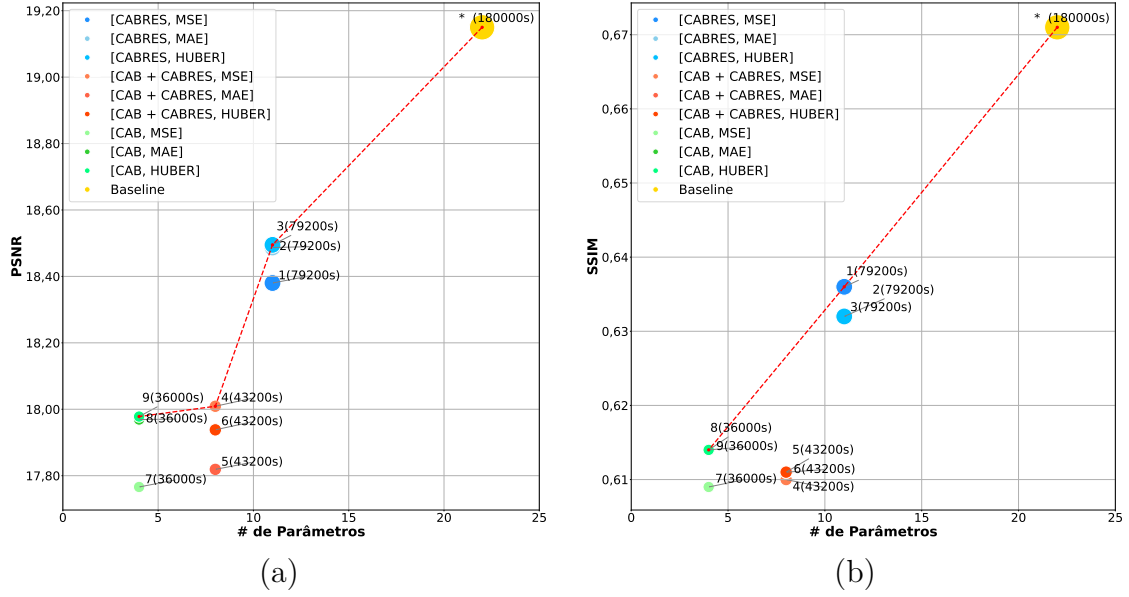


Figura 30 – Resultados da avaliação com a Fronteira de Pareto usando métricas (a) PSNR e (b) SSIM. A linha tracejada vermelha representa a Fronteira de Pareto. O tamanho do círculo para cada experimento no gráfico se refere aproximadamente ao tempo de treinamento de cada modelo para 3000 épocas. Em amarelo, está representado o DiffCR.

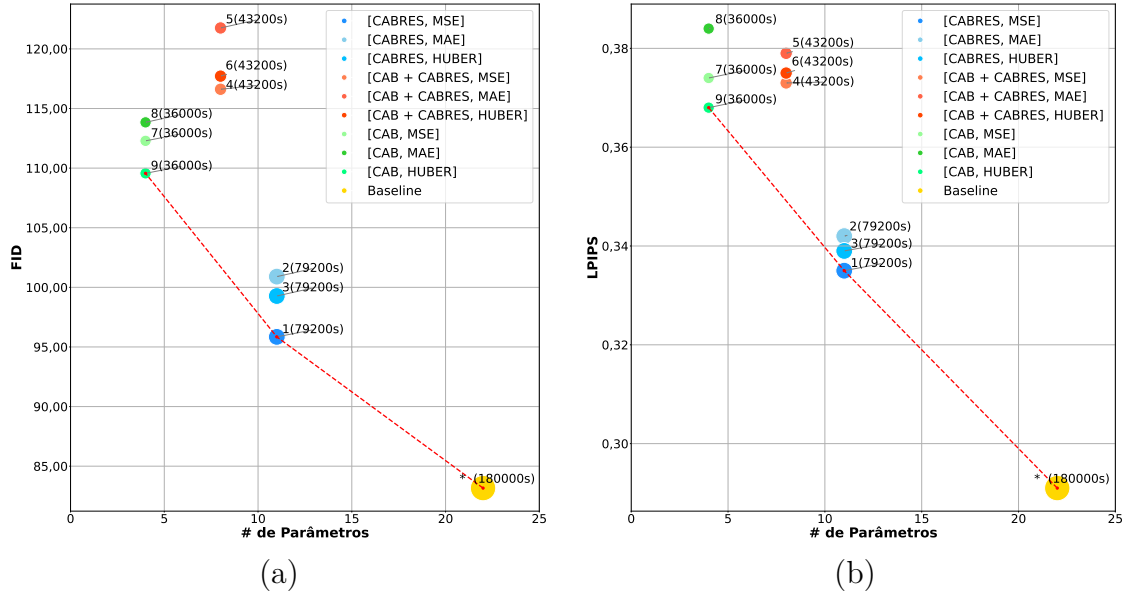


Figura 31 – Resultados da avaliação com a Fronteira de Pareto usando métricas (a) FID e (b) LPIPS. A linha tracejada vermelha representa a Fronteira de Pareto. O tamanho do círculo para cada experimento no gráfico se refere aproximadamente ao tempo de treinamento de cada modelo para 3000 épocas. Em amarelo, está representado o DiffCR.



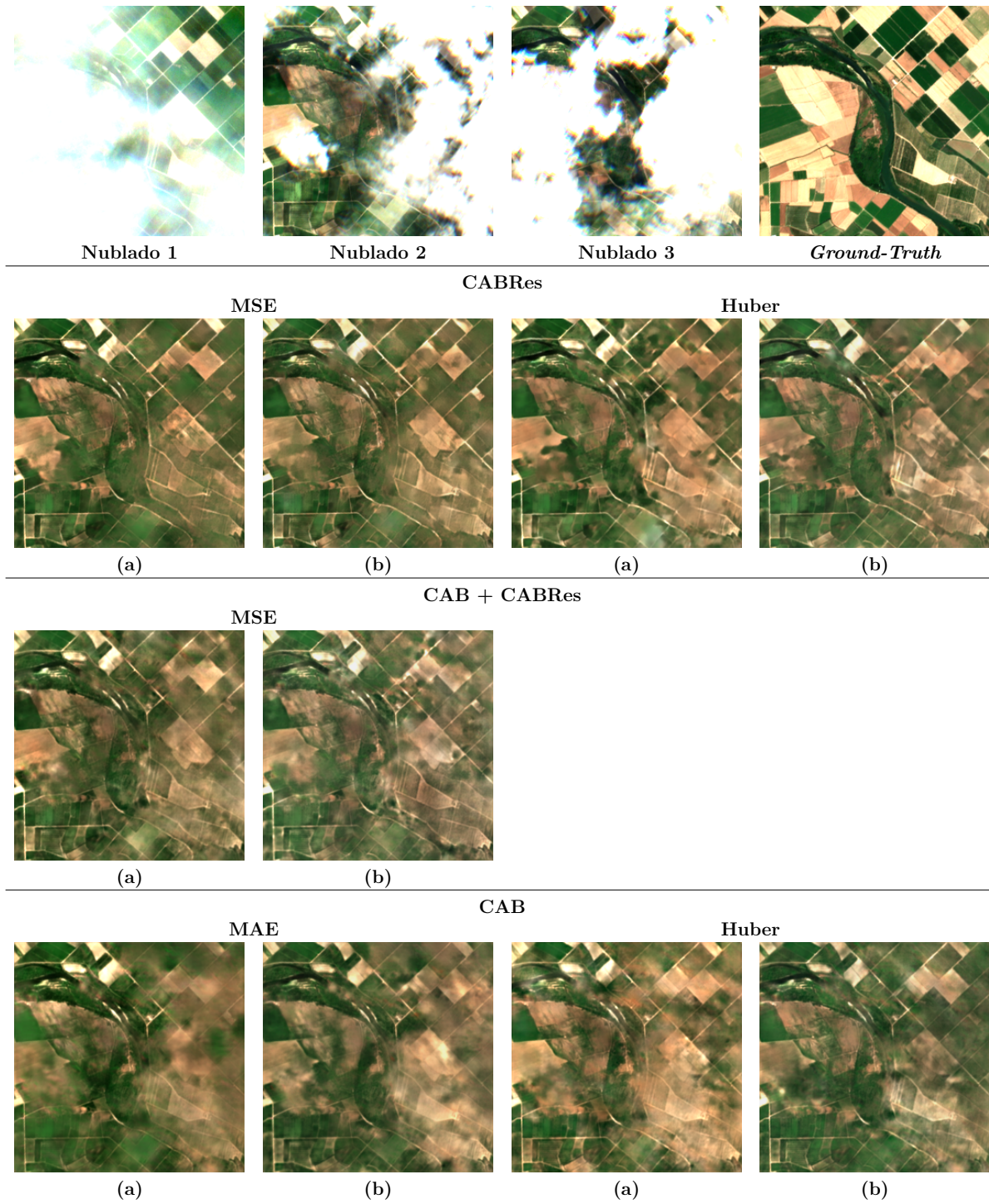


Figura 32 – Comparação de métodos propostos para remoção de nuvens. A primeira linha mostra as imagens nubladas e o *Ground-Truth*. As linhas subsequentes exibem resultados para diferentes abordagens: CABRes, CAB + CABRes e CAB, onde (a) representa os resultados antes da otimização e (b) representa os resultados após a otimização pela Fronteira de Pareto.



Tabela 9 – Comparação dos resultados entre as abordagens propostas após a otimização baseada na Fronteira de Pareto. Os valores entre parênteses representam a diferença em relação aos resultados originais.

Modelo	PSNR $\uparrow$	SSIM $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$
CABRes + MSE	18,506 (+0,126)	0,632 (-0,004)	90,386 (-5,459)	0,327 (-0,008)
CABRes + Huber	18,691 (+0,196)	0,641 (+0,009)	94,522 (-4,751)	0,328 (-0,011)
CAB + CABRes + MSE	18,049 (+0,040)	0,609 (-0,001)	114,912 (-1,702)	0,369 (-0,004)
CAB + MAE	18,042 (+0,276)	0,614 (+0,005)	107,192 (-5,095)	0,379 (+0,005)
CAB + Huber	18,058 (+0,089)	0,611 (-0,003)	103,529 (-6,033)	0,365 (-0,019)

Tabela 10 – Comparação de resultados entre experimentos da literatura e abordagens selecionadas dos experimentos após a otimização baseada na Fronteira de Pareto. Os métodos com um asterisco (\*) não exploraram diretamente os resultados para o conjunto de dados SEN2\_MTC\_New, mas Zou et al. (2024) reportam tais abordagens para esse conjunto de dados.

Método	PSNR $\uparrow$	SSIM $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$	# Params (M) $\downarrow$	MACs (G) $\downarrow$
Métodos da Literatura						
STGAN (SARUKKAI et al., 2020)	18,152	0,587	182,150	0,513	231,93	1094,94
CTGAN (HUANG; WU, 2022)	18,308	0,609	128,704	0,384	642,92	632,05
CR-TS-Net* (EBEL et al., 2022)	18,585	0,615	96,364	0,342	38,68	7602,97
PMAA (ZOU et al., 2023)	18,369	0,614	118,214	0,392	3,45	92,34
UnCRtainTS* (EBEL et al., 2023)	18,770	0,631	93,509	0,333	<b>0,56</b>	<b>37,16</b>
DDPM-CR* (JING et al., 2023)	18,742	0,614	94,401	0,329	445,45	852,37
DiffCR (ZOU et al., 2024)	<b>19,150</b>	<b>0,671</b>	<b>83,162</b>	<b>0,291</b>	22,91	45,86
Abordagens Propostas						
CABRes + MSE	18,506	0,632	<b>90,386</b>	<b>0,327</b>	11,90	97,65
CABRes + Huber	<b>18,691</b>	<b>0,641</b>	94,522	0,328	11,90	97,65
CAB + CABRes + MSE	18,049	0,609	114,912	0,369	8,81	39,61
CAB + MAE	18,042	0,614	107,192	0,379	<b>4,09</b>	<b>29,94</b>
CAB + Huber	18,058	0,611	103,529	0,365	<b>4,09</b>	<b>29,94</b>

postas demonstram um desempenho semelhante ao reportado na literatura, com valores de PSNR e SSIM superiores a outros métodos, como STGAN, CTGAN e CR-TS-Net. Especificamente para a abordagem CABRes + Huber, alcançou-se um PSNR de 18,691 e um SSIM de 0,641, valores próximos aos reportados pelo DiffCR (PSNR de 19,150 e SSIM de 0,671), porém com um custo computacional substancialmente menor. As abordagens propostas também se destacam por apresentarem baixos valores de FID e LPIPS, especialmente quando consideramos CABRes + MSE. Essa abordagem obteve o segundo menor FID (90,386) e LPIPS (0,327 – praticamente empatado com o DDPM-CR), mas com um custo computacional relativamente menor.

Assim, a principal novidade das abordagens está na considerável redução do custo computacional, já que as abordagens CABRes + MSE e CABRes + Huber possuem apenas 11,90M de parâmetros, o que representa uma redução de 48% em relação ao DiffCR, mas mantendo um desempenho altamente competitivo. Embora essas abordagens apresentem um valor de MACs superior ao DiffCR, isso apenas indica um maior número de operações computacionais e não implica em maior uso de memória, o que reforça ainda

mais a aplicabilidade das abordagens propostas em ambientes com recursos computacionais limitados. As abordagens CAB + MAE e CAB + Huber reduzem ainda mais a complexidade, com apenas 4,09M de parâmetros e 29,94G MACs, o que representa uma redução de 82% no número de parâmetros e 35% no número de operações computacionais, quando comparado ao DiffCR. Ao considerar especificamente o UnCRtainTS, o modelo mais eficiente em termos computacionais na literatura, as abordagens CAB + MAE e CAB + Huber oferecem uma redução adicional no valor de MACs, além de utilizar da difusão latente, que ainda pode ser otimizada para alcançar melhores resultados. Em termos de latência (com  $\text{step} = 1$ ), considerando a configuração da máquina destes experimentos, a média de inferência por imagem foi de, respectivamente, 0,1049s, 0,1051s e 0,1083s para CAB, CAB + CABRes e CABRes.

Os resultados demonstram que as abordagens propostas apresentam uma alternativa viável a outros modelos da literatura, especialmente em situações onde o custo computacional é um fator crítico, como no monitoramento ambiental, onde a tarefa de reconstrução de imagens de sensoriamento remoto, combinada com a análise meteorológica, pode reduzir consideravelmente o custo computacional sem perdas substanciais na qualidade da reconstrução, o que é um mecanismo mais adequado para implementações em larga escala ou dispositivos embarcados. Além disso, é essencial destacar que as funções de perda MSE e Huber se destacaram no treinamento das abordagens, com o MSE influenciando fortemente as métricas perceptuais, ou seja, FID e LPIPS, e a função Huber favorecendo as métricas PSNR e SSIM. Dependendo da cobertura do solo, a função de perda impacta diretamente a qualidade da reconstrução. O MSE tende a ser mais eficiente para reconstrução de superfícies homogêneas, como áreas urbanas e vegetação densa. No entanto, por penalizar erros significativos, pode levar a um refino excessivo dos detalhes, comprometendo a reconstrução de bordas e transições sutis, especialmente em áreas com grande variação, como lagos e regiões de vegetação fragmentada. Devido às suas características, o MAE evita o refino excessivo observado no MSE, preservando áreas muito escuras, como lagos e rios. Assim, embora o MAE tenha dificuldades em maximizar a similaridade estrutural global, ela apresenta um bom desempenho para classes específicas e superfícies de baixa reflectância. Por fim, a função Huber combina características de ambas as funções, favorecendo a reconstrução de áreas com transições abruptas, como as bordas entre vegetação e solo exposto. Além disso, a combinação das características das duas funções pode ser útil para extrair de padrões geométricos em cenários onde há edifícios e estradas.

Com intuito verificar o comportamento das abordagens propostas em termos de mais e menos recursos computacionais, conduziu-se experimentos onde há um aumento de canais (filtros) de 64 para 128 e também uma redução para 32. Essa análise comparativa das variações dos modelos sob a perspectiva das mesmas métricas avaliadas pode ser observada na Tabela 11.

Tabela 11 – Resultados experimentais para 32 e 128 canais para os modelos otimizados com a Fronteira de Pareto. Os sublinhados são os melhores resultados para os respectivos grupos de canais considerados e em negrito são os melhores resultados gerais.

Modelos	Canais	PSNR $\uparrow$	SSIM $\uparrow$	FID $\downarrow$	LPIPS $\downarrow$	# Params (M) $\downarrow$	MACs (G) $\downarrow$
CABRes + MSE	32	<u>17,861</u>	<u>0,618</u>	<u>102,882</u>	<u>0,372</u>	3,04	24,63
CABRes + Huber	32	17,856	0,612	105,106	<u>0,369</u>	3,04	24,63
CAB + CABRes + MSE	32	17,354	0,587	121,742	<u>0,405</u>	2,26	10,11
CAB + MAE	32	17,605	0,597	115,370	0,393	<b>1,08</b>	<b>7,69</b>
CAB + Huber	32	17,677	0,596	111,287	0,389	<b>1,08</b>	<b>7,69</b>
CABRes + MSE	128	18,830	0,631	96,143	0,317	47,39	388,43
CABRes + Huber	128	<b>19,018</b>	<b>0,643</b>	<b>92,962</b>	<b>0,314</b>	47,39	388,43
CAB + CABRes + MSE	128	<u>18,393</u>	<u>0,612</u>	<u>113,454</u>	<u>0,352</u>	35,00	156,38
CAB + MAE	128	18,341	0,624	102,414	0,346	<u>16,12</u>	<u>117,27</u>
CAB + Huber	128	18,518	0,624	99,781	0,342	<u>16,12</u>	<u>117,27</u>

A princípio, observa-se que esse número de canais impacta na qualidade da reconstrução. A função de perda Huber também se mostrou adequada nesse cenário experimentado, visto que apresenta os melhores valores para os dois grupos considerados. Um destaque a considerar é especificamente para o modelo CABRes + Huber, que apresenta resultados ainda mais próximos ao DiffCR, o que evidencia a qualidade da abordagem proposta que pode praticamente igualar a *baseline*. Outro ponto a se destacar é que esses experimentos demonstraram que os conjuntos de experimentos CAB + CABRes apresentaram desempenho ainda inferior aos modelos CAB + Huber tanto para 32 quanto para 128 canais, mesmo com aproximadamente o dobro de parâmetros treináveis.

Desse modo, essa análise indica que (i) aumentar o número de canais melhora as métricas de reconstrução e (ii) a função de perda Huber se destaca, principalmente para as métricas perceptuais.

## 6.4 Considerações

Este capítulo apresentou diferentes abordagens para remoção de nuvens através de difusão latente combinada com canais de atenção e blocos residuais. Explorou-se diferentes funções de perda para avaliar o impacto no treinamento das abordagens e na qualidade da reconstrução. Essas abordagens demonstraram-se eficientes para equilibrar desempenho e custo computacional, potencializando-as para aplicações onde há restrições de *hardware*.

Os experimentos indicam que, embora o DiffCR com o TCFBlock ainda apresente desempenho qualitativamente um pouco superior, há uma demanda muito maior por custo computacional, se comparado com as abordagens propostas. Essas abordagens com os respectivos blocos propostos apresentaram uma considerável redução de custo computacional, com valores 48% a 82% menores de parâmetros se comparado à *baseline*, sem uma perda considerável de qualidade.

Por fim, outra contribuição deste estudo é sobre a utilização da função de perda para

o treinamento, visto que ela impacta diretamente à reconstrução em virtude da cobertura de solo considerada. O destaque é a função Huber, que combina aspectos do MAE e MSE, e tende a ter desempenho mais abrangente, principalmente quando se aumenta a quantidade de filtros utilizados para as abordagens propostas, o que sugere que essa função gera um equilíbrio em um cenário de dados escassos.

---

## Conclusão

Este trabalho avaliou diferentes aspectos da segmentação e remoção de nuvens em imagens, considerando a influência de arquiteturas, *encoders*, funções de perda e otimizadores. Os experimentos conduzidos demonstraram que a escolha de cada um desses componentes impacta diretamente o desempenho dos modelos, sendo essencial uma seleção criteriosa para obter resultados satisfatórios.

Na tarefa de segmentação de nuvens, observou-se que os *encoders* mais profundos, como ResNet-50, DenseNet-121 e VGG-16, resultam em segmentações mais precisas independentemente da arquitetura adotada. Entre as arquiteturas testadas, a MA-Net destacou-se por sua capacidade de incorporar mecanismos de atenção, de modo a melhorar a segmentação. Especificamente, a combinação MA-Net com *encoder* VGG-16 reduziu a ocorrência de falsos positivos e falsos negativos, o que é relevante para a identificação de nuvens finas com transparência de solo.

Para remoção de nuvens utilizando modelos de difusão latente, verificou-se que a seleção do otimizador e da função de perda influencia significativamente a qualidade da reconstrução. A combinação do otimizador AdamW com a função de perda Huber mostrou-se particularmente eficaz, permitindo um treinamento mais estável e enfatizando regiões com maior oclusão de nuvens. Além disso, as novas abordagens propostas neste trabalho reduziram o custo computacional entre 48% a 82% para, respectivamente, as abordagens CABRes e CAB, sem perdas consideráveis na qualidade da reconstrução de imagens com nuvens quando comparadas ao DiffCR.

Um fator a se considerar neste trabalho é acerca do impacto da função de perda, principalmente quando considera-se diversas coberturas de solo ocluídas por diferentes tipos de nuvens. Dentre as tendências reportadas, pode-se mencionar: a função MSE sendo mais aderente para coberturas de solo que envolvem estradas e vegetação; a função MAE consegue capturar melhor corpos d'água; e a função Huber demonstra robustez onde há cenários mais complexos, onde consideram-se diferentes bodas e padrões geométricos, como é o caso de áreas urbanas. Além disso, a função MSE tende a aprimorar avaliações baseadas em métricas perceptuais, enquanto a função Huber tende a favorecer as métricas

baseadas em correlação e ruído. Dessa forma, este trabalho demonstra a viabilidade de arquiteturas mais eficientes, que preservam a relação de equilíbrio entre desempenho e complexidade computacional, o que é particularmente útil em monitoramento ambiental e análise meteorológica, além de poder inspirar treinamento de modelos semelhantes, mas com diferentes bandas espectrais e coberturas de solo consideradas. Por fim, é importante mencionar que, apesar dos bons resultados reportados, ainda há espaço para melhoria na reconstrução, principalmente com combinação adaptativa de função de perda e modelos treinados a partir de regras de decisão, que considerem prioritariamente a cobertura de solo para escolha da função de perda no treinamento.

## 7.1 Principais Contribuições

Artigos publicados:

- ❑ SILVA, L. H. F. P.; JUNIOR, J. D. D.; MARI, J. F.; ESCARPINATI, M. C.; BACKES, A. R. Non-linear co-registration in UAVs' images using deep learning. In: **SIBGRAPI - CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES**, 35., 2022, Natal: IEEE, 2022. v. 1, p. 1-6.
- ❑ DA SILVA, M. V.; SILVA, L. H. F. P.; JUNIOR, J. D. D.; ESCARPINATI, M. C.; BACKES, A. R.; MARI, J. F. Generating synthetic multispectral images using neural style transfer: A study with application in channel alignment. *Computers and Electronics in Agriculture*, v. 206, p. 107668, 2023.
- ❑ ARAKAKI, L. G.; SILVA, L. H. F. P.; DA SILVA, M. V.; MELO, B. M.; BACKES, A. R.; ESCARPINATI, M. C.; MARI, J. F. Evaluation of U-Net Backbones for Cloud Segmentation in Satellite Images. In: **VISIGRAPP - INTERNATIONAL JOINT CONFERENCE ON COMPUTER VISION, IMAGING AND COMPUTER GRAPHICS THEORY AND APPLICATIONS**, 18., 2023, Lisboa: SCITEPRESS, 2023. v. 4, VISAPP, p. 452-458.
- ❑ FERREIRA, J. L. R.; SILVA, L. H. F. P.; BACKES, A. R.; ESCARPINATI, M. C.; MARI, J. F. Evaluating Multiple Combinations of Models and Encoders to Segment Clouds in Satellite Images. In: **VISIGRAPP - INTERNATIONAL JOINT CONFERENCE ON COMPUTER VISION, IMAGING AND COMPUTER GRAPHICS THEORY AND APPLICATIONS**, 19., 2024, Roma: SCITEPRESS, 2024. p. 233-241.
- ❑ SILVA, L. H. F. P.; MARI, J. F.; ESCARPINATI, M. C.; BACKES, A. R. Evaluating Combinations of Optimizers and Loss Functions for Cloud Removal Using Diffusion Models. In: **VISIGRAPP - INTERNATIONAL JOINT CONFERENCE ON COMPUTER VISION, IMAGING AND COMPUTER**

**GRAPHICS THEORY AND APPLICATIONS**, 20., 2025, Porto: SCITEPRESS, 2025. p. 648-656.

Artigo aceito para publicação:

- ❑ SILVA, L. H. F. P.; ESCARPINATI, M. C.; BACKES, A. R.; MARI, J. F.. Evaluating architecture and encoder combinations for cloud segmentation in satellite images. *Communications in Computer and Information Science*, [S. l.].

Artigos submetidos para publicação:

- ❑ SILVA, L. H. F. P.; MARI, J. F.; ESCARPINATI, M. C.; BACKES, A. R. A survey on cloud removal in satellite images using deep learning. *Pattern Recognition*, [S. l.]. submetido em 09 dez. 2024. Em revisão.
- ❑ SILVA, L. H. F. P.; MARI, J. F.; ESCARPINATI, M. C.; BACKES, A. R. Cloud Removal with Compact Diffusion Models: A Residual Block-Based Approach. *Remote Sensing Applications: Society and Environment*, [S. l.]. submetido em 25 mar. 2025. Em revisão.

## 7.2 Trabalhos Futuros

Como trabalhos futuros, pretende-se:

- ❑ Avaliar as abordagens propostas para reconstrução em outras bases de dados, principalmente para contemplar outras coberturas de solo e bandas espectrais.
- ❑ Integrar um *framework* que seja capaz de realizar a segmentação e, posteriormente a reconstrução com as abordagens propostas, a fim de aprimorar a qualidade das imagens geradas.





---

## Referências

- ABBASI, R.; MARTINEZ, P.; AHMAD, R. The digitization of agricultural industry—a systematic literature review on agriculture 4.0. **Smart Agricultural Technology**, Elsevier, p. 100042, 2022. Disponível em: <<https://doi.org/10.1016/j.atech.2022.100042>>.
- AHMED, K. M.; DHUBAIB, B. E. A. Zotero: A bibliographic assistant to researcher. **Journal of Pharmacology and Pharmacotherapeutics**, SAGE Publications Sage India: New Delhi, India, v. 2, n. 4, p. 304–305, 2011. Disponível em: <<https://doi.org/10.4103/0976-500X.85940>>.
- AJAYI, O. G.; OLAIDE, D. M. Assessment of generative adversarial networks for cloud occlusion removal on remotely sensed images. **Arabian Journal of Geosciences**, Springer, v. 17, n. 4, p. 141, 2024. Disponível em: <<https://doi.org/10.1007/s12517-024-11939-y>>.
- AKHTER, R.; SOFI, S. A. Precision agriculture using iot data analytics and machine learning. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, v. 34, n. 8, p. 5602–5618, 2022. Disponível em: <<https://doi.org/10.1016/j.jksuci.2021.05.013>>.
- ANGEL, Y.; HOUBORG, R.; MCCABE, M. F. Reconstructing cloud contaminated pixels using spatiotemporal covariance functions and multitemporal hyperspectral imagery. **Remote Sensing**, MDPI, v. 11, n. 10, p. 1145, 2019. Disponível em: <<https://doi.org/10.3390/rs11101145>>.
- ARAKAKI, L. G. et al. Evaluation of u-net backbones for cloud segmentation in satellite images. In: **VISIGRAPP (4: VISAPP)**. [s.n.], 2023. p. 452–458. Disponível em: <<https://doi.org/10.5220/0011787200003417>>.
- ARMSTRONG, R. et al. ‘scoping the scope’ of a cochrane review. **Journal of public health**, Oxford University Press, v. 33, n. 1, p. 147–150, 2011. Disponível em: <<https://doi.org/10.1093/pubmed/fdr015>>.
- ARNAUDO, E. et al. Robust burned area delineation through multitask learning. In: SPRINGER. **Joint European Conference on Machine Learning and Knowledge Discovery in Databases**. 2023. p. 436–447. Disponível em: <[https://doi.org/10.1007/978-3-031-74633-8\\_32](https://doi.org/10.1007/978-3-031-74633-8_32)>.

- ARP, L. et al. Training-free thick cloud removal for sentinel-2 imagery using value propagation interpolation. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 216, p. 168–184, 2024. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2024.07.030>>.
- ASTOLA, H. et al. Comparison of sentinel-2 and landsat 8 imagery for forest variable prediction in boreal region. **Remote Sensing of Environment**, Elsevier, v. 223, p. 257–273, 2019. Disponível em: <<https://doi.org/10.1016/j.rse.2019.01.019>>.
- AYBAR, C. et al. Cloudsen12, a global dataset for semantic understanding of cloud and cloud shadow in sentinel-2. **Scientific data**, Nature Publishing Group UK London, v. 9, n. 1, p. 782, 2022. Disponível em: <<https://doi.org/10.1038/s41597-022-01878-2>>.
- BAI, T. et al. Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion. **Remote Sensing**, MDPI, v. 8, n. 9, p. 715, 2016. Disponível em: <<https://doi.org/10.3390/rs8090715>>.
- BARBOSA, G. et al. Optimization and learning rate influence on breast cancer image classification. In: INSTICC. **Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP**. SciTePress, 2024. p. 792–799. ISBN 978-989-758-679-8. Disponível em: <<https://doi.org/10.5220/0012507100003660>>.
- BASKAR, D.; PARAMBALATH, N. K.; KRISHNANUNNI, S. O. Variational mode decomposition-enhanced thin cloud removal using unet vision transformer cycle-consistent generative adversarial network. **Journal of Applied Remote Sensing**, Society of Photo-Optical Instrumentation Engineers, v. 18, n. 2, p. 026504–026504, 2024. Disponível em: <<https://doi.org/10.1117/1.JRS.18.026504>>.
- BERMUDEZ, J. D. et al. Synthesis of multispectral optical images from sar/optical multitemporal data using conditional generative adversarial networks. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 16, n. 8, p. 1220–1224, 2019. Disponível em: <<https://doi.org/10.1109/LGRS.2019.2894734>>.
- CAI, W.; WEI, Z. Remote sensing image classification based on a cross-attention mechanism and graph convolution. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 19, p. 1–5, 2020. Disponível em: <<https://doi.org/10.1109/LGRS.2020.3026587>>.
- CAO, R. et al. Thick cloud removal in landsat images based on autoregression of landsat time-series data. **Remote Sensing of Environment**, Elsevier, v. 249, p. 112001, 2020. Disponível em: <<https://doi.org/10.1016/j.rse.2020.112001>>.
- CHASTAIN, R. et al. Empirical cross sensor comparison of sentinel-2a and 2b msi, landsat-8 oli, and landsat-7 etm+ top of atmosphere spectral characteristics over the conterminous united states. **Remote sensing of environment**, Elsevier, v. 221, p. 274–285, 2019. Disponível em: <<https://doi.org/10.1016/j.rse.2018.11.012>>.
- CHAURASIA, A.; CULURCIELLO, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: IEEE. **2017 IEEE visual communications and image processing (VCIP)**. 2017. p. 1–4. Disponível em: <<https://doi.org/10.1109/VCIP.2017.8305148>>.

- CHEN, S. et al. Cloud removal with sar-optical data fusion and graph-based feature aggregation network. **Remote Sensing**, MDPI, v. 14, n. 14, p. 3374, 2022. Disponível em: <<https://doi.org/10.3390/rs14143374>>.
- CHEN, Y. et al. Thick cloud removal in multitemporal remote sensing images via low-rank regularized self-supervised network. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 62, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3358493>>.
- \_\_\_\_\_. Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 157, p. 93–107, 2019. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2019.09.003>>.
- \_\_\_\_\_. Thick clouds removal from multitemporal zy-3 satellite images using deep learning. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 13, p. 143–153, 2019. Disponível em: <<https://doi.org/10.1109/JSTARS.2019.2954130>>.
- \_\_\_\_\_. Thick clouds removing from multitemporal landsat images using spatiotemporal neural networks. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–14, 2020. Disponível em: <<https://doi.org/10.1109/TGRS.2020.3043980>>.
- COUTTS, A. M. et al. Thermal infrared remote sensing of urban heat: Hotspots, vegetation, and an assessment of techniques for use in urban planning. **Remote sensing of environment**, Elsevier, v. 186, p. 637–651, 2016. Disponível em: <<https://doi.org/10.1016/j.rse.2016.09.007>>.
- CROITORU, F.-A. et al. Diffusion models in vision: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2023. Disponível em: <<https://doi.org/10.1109/TPAMI.2023.3261988>>.
- CZERKAWSKI, M. et al. Satellitecloudgenerator: controllable cloud and shadow synthesis for multi-spectral optical satellite images. **Remote Sensing**, MDPI, v. 15, n. 17, p. 4138, 2023. Disponível em: <<https://doi.org/10.3390/rs15174138>>.
- \_\_\_\_\_. Deep internal learning for inpainting of cloud-affected regions in satellite imagery. **Remote Sensing**, MDPI, v. 14, n. 6, p. 1342, 2022. Disponível em: <<https://doi.org/10.3390/rs14061342>>.
- DAI, P.; JI, S.; ZHANG, Y. Gated convolutional networks for cloud removal from bi-temporal remote sensing images. **Remote Sensing**, MDPI, v. 12, n. 20, p. 3427, 2020. Disponível em: <<https://doi.org/10.3390/rs12203427>>.
- DARBAGHSHAHI, F. N.; MOHAMMADI, M. R.; SORYANI, M. Cloud removal in remote sensing images using generative adversarial networks and sar-to-optical image translation. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–9, 2021. Disponível em: <<https://doi.org/10.1109/TGRS.2021.3131035>>.
- DEB, K. et al. A fast and elitist multiobjective genetic algorithm: Nsga-ii. **IEEE transactions on evolutionary computation**, Ieee, v. 6, n. 2, p. 182–197, 2002. Disponível em: <<https://doi.org/10.1109/4235.996017>>.

DELLOYE, C.; WEISS, M.; DEFOURNY, P. Retrieval of the canopy chlorophyll content from sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. **Remote Sensing of Environment**, Elsevier, v. 216, p. 245–261, 2018. Disponível em: <<https://doi.org/10.1016/j.rse.2018.06.037>>.

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. **2009 IEEE conference on computer vision and pattern recognition**. 2009. p. 248–255. Disponível em: <<https://doi.org/10.1109/CVPR.2009.5206848>>.

DESPRÉS, P.; JIA, X. A review of gpu-based medical image reconstruction. **Physica Medica**, Elsevier, v. 42, p. 76–92, 2017. Disponível em: <<https://doi.org/10.1016/j.ejmp.2017.07.024>>.

DING, H. et al. Robust haze and thin cloud removal via conditional variational autoencoders. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3349779>>.

DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. **Computer Science Review**, Elsevier, v. 40, p. 100379, 2021. Disponível em: <<https://doi.org/10.1016/j.cosrev.2021.100379>>.

DOU, A. et al. Remote sensing image cloud removal based on multi-scale spatial information perception. **Multimedia Systems**, Springer, v. 30, n. 5, p. 249, 2024. Disponível em: <<https://doi.org/10.1007/s00530-024-01442-5>>.

DU, W. et al. An efficient approach to remove thick cloud in vnir bands of multi-temporal remote sensing images. **Remote Sensing**, MDPI, v. 11, n. 11, p. 1284, 2019. Disponível em: <<https://doi.org/10.3390/rs11111284>>.

DUAN, C.; BELGIU, M.; STEIN, A. Efficient cloud removal network for satellite images using sar-optical image fusion. **IEEE Geoscience and Remote Sensing Letters**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/LGRS.2024.3397875>>.

\_\_\_\_\_. Feature enhancement network for cloud removal in optical images by fusing with sar images. **International Journal of Remote Sensing**, Taylor & Francis, v. 45, n. 1, p. 51–67, 2024. Disponível em: <<https://doi.org/10.1080/01431161.2023.2292014>>.

\_\_\_\_\_. Multi-feature fusion network for efficient cloud removal using sar-optical image fusion. In: IEEE. **IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium**. 2024. p. 9062–9065. Disponível em: <<https://doi.org/10.1109/IGARSS53475.2024.10641774>>.

DUAN, C.; PAN, J.; LI, R. Thick cloud removal of remote sensing images using temporal smoothness and sparsity regularized tensor optimization. **Remote Sensing**, MDPI, v. 12, n. 20, p. 3446, 2020. Disponível em: <<https://doi.org/10.3390/rs12203446>>.

EBEL, P. et al. Uncertainties: Uncertainty quantification for cloud removal in optical satellite time series. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [s.n.], 2023. p. 2085–2095. Disponível em: <<https://doi.org/10.1109/CVPRW59228.2023.00202>>.

\_\_\_\_\_. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 59, n. 7, p. 5866–5878, 2020. Disponível em: <<https://doi.org/10.1109/TGRS.2020.3024744>>.

\_\_\_\_\_. Multisensor Data Fusion for Cloud Removal in Global and All-Season Sentinel-2 Imagery. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2020. Disponível em: <<https://doi.org/10.1109/TGRS.2020.3024744>>.

EBEL, P.; SCHMITT, M.; ZHU, X. X. Cloud removal in unpaired sentinel-2 imagery using cycle-consistent gan and sar-optical data fusion. In: IEEE. **IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium**. 2020. p. 2065–2068. Disponível em: <<https://doi.org/10.1109/IGARSS39084.2020.9324060>>.

\_\_\_\_\_. Internal learning for sequence-to-sequence cloud removal via synthetic aperture radar prior information. In: IEEE. **2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS**. 2021. p. 2691–2694. Disponível em: <<https://doi.org/10.1109/IGARSS47720.2021.9554268>>.

EBEL, P. et al. Sen12ms-cr-ts: A remote-sensing data set for multimodal multitemporal cloud removal. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–14, 2022. Disponível em: <<https://doi.org/10.1109/TGRS.2022.3146246>>.

EPPENHOF, K. A. et al. Progressively trained convolutional neural networks for deformable image registration. **IEEE Transactions on Medical Imaging**, IEEE, v. 39, n. 5, p. 1594–1604, 2019. Disponível em: <<https://doi.org/10.1109/TMI.2019.2953788>>.

FAN, T. et al. Ma-net: A multi-scale attention network for liver and tumor segmentation. **IEEE Access**, IEEE, v. 8, p. 179656–179665, 2020. Disponível em: <<https://doi.org/10.1109/ACCESS.2020.3025372>>.

FARIA, L. C. de et al. Paving the way for automatic mapping of rural roads in the amazon rainforest. In: IEEE. **2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. 2023. p. 1–6. Disponível em: <<https://doi.org/10.1109/SIBGRAPI59091.2023.10347153>>.

FERREIRA, J. R. et al. Evaluating multiple combinations of models and encoders to segment clouds in satellite images. In: INSTICC. **Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP**. SciTePress, 2024. p. 233–241. ISBN 978-989-758-679-8. Disponível em: <<https://doi.org/10.5220/0012506700003660>>.

GAO, J. et al. Sentinel-2 cloud removal considering ground changes by fusing multitemporal sar and optical images. **Remote Sensing**, MDPI, v. 13, n. 19, p. 3998, 2021. Disponível em: <<https://doi.org/10.3390/rs13193998>>.

\_\_\_\_\_. Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. **Remote Sensing**, MDPI, v. 12, n. 1, p. 191, 2020. Disponível em: <<https://doi.org/10.3390/rs12010191>>.

GHILDIYAL, S. et al. Ssgan: Cloud removal in satellite images using spatio-spectral generative adversarial network. **European Journal of Agronomy**, Elsevier, v. 161, p. 127333, 2024. Disponível em: <<https://doi.org/10.1016/j.eja.2024.127333>>.

GHOZATLOU, O.; DATCU, M. Hybrid gan and spectral angular distance for cloud removal. In: IEEE. **2021 IEEE International Geoscience and**

- Remote Sensing Symposium IGARSS**. 2021. p. 2695–2698. Disponível em: <<https://doi.org/10.1109/IGARSS47720.2021.9554891>>.
- GONZALEZ, R. C. **Digital image processing**. [S.l.]: Pearson education, 2009.
- GOODFELLOW, I. et al. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1.
- GUO, F. et al. Haze removal for single image: A comprehensive review. **Neurocomputing**, Elsevier, 2023. Disponível em: <<https://doi.org/10.1016/j.neucom.2023.03.061>>.
- GUO, H. et al. Mambair: A simple baseline for image restoration with state-space model. In: SPRINGER. **European Conference on Computer Vision**. 2024. p. 222–241. Disponível em: <[https://doi.org/10.1007/978-3-031-72649-1\\_13](https://doi.org/10.1007/978-3-031-72649-1_13)>.
- GUO, J. et al. Rsdehazenet: Dehazing network with channel refinement for multispectral remote sensing images. **IEEE Transactions on geoscience and remote sensing**, IEEE, v. 59, n. 3, p. 2535–2549, 2020. Disponível em: <<https://doi.org/10.1109/TGRS.2020.3004556>>.
- GUO, Y. et al. Cloud detection for satellite imagery using attention-based u-net convolutional neural network. **Symmetry**, MDPI, v. 12, n. 6, p. 1056, 2020. Disponível em: <<https://doi.org/10.3390/sym12061056>>.
- \_\_\_\_\_. Deep learning for visual understanding: A review. **Neurocomputing**, v. 187, p. 27–48, 2016. Disponível em: <<https://doi.org/10.1016/j.neucom.2015.09.116>>.
- HAN, K. et al. A survey on vision transformer. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 45, n. 1, p. 87–110, 2022. Disponível em: <<https://doi.org/10.1109/TPAMI.2022.3152247>>.
- HAN, S.; WANG, J.; ZHANG, S. Former-cr: A transformer-based thick cloud removal method with optical and sar imagery. **Remote Sensing**, MDPI, v. 15, n. 5, p. 1196, 2023. Disponível em: <<https://doi.org/10.3390/rs15051196>>.
- HASAN, C. et al. Cloud removal from satellite imagery using multispectral edge-filtered conditional generative adversarial networks. **International Journal of Remote Sensing**, Taylor & Francis, v. 43, n. 5, p. 1881–1893, 2022. Disponível em: <<https://doi.org/10.1080/01431161.2022.2048915>>.
- HAYNES, D.; CORNS, S.; VENAYAGAMOORTHY, G. K. An exponential moving average algorithm. In: **2012 IEEE Congress on Evolutionary Computation**. [s.n.], 2012. p. 1–8. Disponível em: <<https://doi.org/10.1109/CEC.2012.6252962>>.
- HE, K.; SUN, J.; TANG, X. Single image haze removal using dark channel prior. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 33, n. 12, p. 2341–2353, 2010. Disponível em: <<https://doi.org/10.1109/TPAMI.2010.168>>.
- HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2016. p. 770–778. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>.
- HE, W. et al. Remote sensing image reconstruction using tensor ring completion and total variation. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 57, n. 11, p. 8998–9009, 2019. Disponível em: <<https://doi.org/10.1109/TGRS.2019.2924017>>.

- HEDLEY, J. D. et al. Coral reef applications of sentinel-2: Coverage, characteristics, bathymetry and benthic mapping with comparison to landsat 8. **Remote sensing of environment**, Elsevier, v. 216, p. 598–614, 2018. Disponível em: <<https://doi.org/10.1016/j.rse.2018.07.014>>.
- HELBERT, P. et al. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 12, n. 7, p. 2217–2226, 2019. Disponível em: <<https://doi.org/10.1109/JSTARS.2019.2918242>>.
- HIRAHARA, N. et al. Denoising and inpainting of sea surface temperature image with adversarial physical model loss. In: SPRINGER. **Pattern Recognition: 5th Asian Conference, ACPR 2019, Auckland, New Zealand, November 26–29, 2019, Revised Selected Papers, Part I 5**. 2020. p. 339–352. Disponível em: <[https://doi.org/10.1007/978-3-030-41404-7\\_24](https://doi.org/10.1007/978-3-030-41404-7_24)>.
- HO, J.; JAIN, A.; ABBEEL, P. Denoising diffusion probabilistic models. **Advances in neural information processing systems**, v. 33, p. 6840–6851, 2020. Disponível em: <<https://doi.org/10.48550/arXiv.2006.11239>>.
- HORE, A.; ZIOU, D. Image quality metrics: Psnr vs. ssim. In: IEEE. **2010 20th international conference on pattern recognition**. 2010. p. 2366–2369. Disponível em: <<https://doi.org/10.1109/ICPR.2010.579>>.
- HOSPEDALES, T. et al. Meta-learning in neural networks: A survey. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 44, n. 9, p. 5149–5169, 2021. Disponível em: <<https://doi.org/10.1109/TPAMI.2021.3079209>>.
- HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: **2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [s.n.], 2018. p. 7132–7141. Disponível em: <<https://doi.org/10.1109/CVPR.2018.00745>>.
- HUANG, G. et al. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2017. p. 4700–4708. Disponível em: <<https://doi.org/10.1109/CVPR.2017.243>>.
- HUANG, G.-L.; WU, P.-Y. Ctgan: Cloud transformer generative adversarial network. In: IEEE. **2022 IEEE International Conference on Image Processing (ICIP)**. 2022. p. 511–515. Disponível em: <<https://doi.org/10.1109/ICIP46576.2022.9897229>>.
- HUANG, S.-C. et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. **NPJ Digital Medicine**, Nature Publishing Group UK London, v. 6, n. 1, p. 74, 2023. Disponível em: <<https://doi.org/10.1038/s41746-023-00811-0>>.
- JEPPESEN, J. H. et al. A cloud detection algorithm for satellite imagery based on deep learning. **Remote sensing of environment**, Elsevier, v. 229, p. 247–259, 2019. Disponível em: <<https://doi.org/10.1016/j.rse.2019.03.039>>.
- JI, S. et al. Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 59, n. 1, p. 732–748, 2020. Disponível em: <<https://doi.org/10.1109/TGRS.2020.2994349>>.

JI, T.-Y. et al. A unified framework of cloud detection and removal based on low-rank and group sparse regularizations for multitemporal multispectral images. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–15, 2022. Disponível em: <<https://doi.org/10.1109/TGRS.2022.3152630>>.

JIANG, B. et al. Fdt-net: Deep-learning network for thin-cloud removal in remote sensing image using frequency domain training strategy. **IEEE Geoscience and Remote Sensing Letters**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/LGRS.2024.3389995>>.

JIN, M.; WANG, P.; LI, Y. Hya-gan: remote sensing image cloud removal based on hybrid attention generation adversarial network. **International Journal of Remote Sensing**, Taylor & Francis, v. 45, n. 6, p. 1755–1773, 2024. Disponível em: <<https://doi.org/10.1080/01431161.2024.2316673>>.

JING, R. et al. Cloud removal for optical remote sensing imagery using the spacyclegan network. **Journal of Applied Remote Sensing**, Society of Photo-Optical Instrumentation Engineers, v. 16, n. 3, p. 034520–034520, 2022. Disponível em: <<https://doi.org/10.1117/1.JRS.16.034520>>.

\_\_\_\_\_. Denoising diffusion probabilistic feature-based network for cloud removal in sentinel-2 imagery. **Remote Sensing**, MDPI, v. 15, n. 9, p. 2217, 2023. Disponível em: <<https://doi.org/10.3390/rs15092217>>.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. **Computers and electronics in agriculture**, Elsevier, v. 147, p. 70–90, 2018. Disponível em: <<https://doi.org/10.1016/j.compag.2018.02.016>>.

KAWAR, B. et al. Denoising diffusion restoration models. **Advances in Neural Information Processing Systems**, v. 35, p. 23593–23606, 2022. Disponível em: <<https://doi.org/10.48550/arXiv.2201.11793>>.

KHANAL, S.; FULTON, J.; SHEARER, S. An overview of current and potential applications of thermal remote sensing in precision agriculture. **Computers and electronics in agriculture**, Elsevier, v. 139, p. 22–32, 2017. Disponível em: <<https://doi.org/10.1016/j.compag.2017.05.001>>.

KHANDELWAL, P. et al. Cloudnet: A deep learning approach for mitigating occlusions in landsat-8 imagery using data coalescence. In: IEEE. **2022 IEEE 18th International Conference on e-Science (e-Science)**. 2022. p. 117–127. Disponível em: <<https://doi.org/10.1109/eScience55777.2022.00026>>.

KINGMA, D. P. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014. Disponível em: <<https://doi.org/10.48550/arXiv.1412.6980>>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, 2012. Disponível em: <<https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>>.

\_\_\_\_\_. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, ACM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. Disponível em: <<https://doi.org/10.1145/3065386>>.



KUMAR, B. P. et al. Geo-environmental monitoring and assessment of land degradation and desertification in the semi-arid regions using landsat 8 oli/tirs, lst, and ndvi approach. **Environmental Challenges**, Elsevier, v. 8, p. 100578, 2022. Disponível em: <<https://doi.org/10.1016/j.envc.2022.100578>>.

LACROIX, P. et al. Use of sentinel-2 images for the detection of precursory motions before landslide failures. **Remote Sensing of Environment**, Elsevier, v. 215, p. 507–516, 2018. Disponível em: <<https://doi.org/10.1016/j.rse.2018.03.042>>.

LECUN, Y.; BENGIO, Y.; HINTON, G. E. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>.

LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. **Neural computation**, MIT Press, v. 1, n. 4, p. 541–551, 1989. Disponível em: <<https://doi.org/10.1162/neco.1989.1.4.541>>.

\_\_\_\_\_. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: <<https://doi.org/10.1109/5.726791>>.

LEE, K.-Y.; SIM, J.-Y. Cloud removal of satellite images using convolutional neural network with reliable cloudy image synthesis model. In: IEEE. **2019 IEEE International Conference on Image Processing (ICIP)**. 2019. p. 3581–3585. Disponível em: <<https://doi.org/10.1109/ICIP.2019.8803666>>.

LI, C.; LIU, X.; LI, S. Transformer meets gan: Cloud-free multispectral image reconstruction via multi-sensor data fusion in satellite images. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2023. Disponível em: <<https://doi.org/10.1109/TGRS.2023.3326545>>.

LI, H. et al. Pyramid attention network for semantic segmentation. **arXiv preprint arXiv:1805.10180**, 2018. Disponível em: <<https://doi.org/10.48550/arXiv.1805.10180>>.

LI, J.; HU, Q.; AI, M. Haze and thin cloud removal via sphere model improved dark channel prior. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 16, n. 3, p. 472–476, 2018. Disponível em: <<https://doi.org/10.1109/LGRS.2018.2874084>>.

LI, J. et al. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 166, p. 373–389, 2020. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2020.06.021>>.

\_\_\_\_\_. Deep learning based thin cloud removal fusing vegetation red edge and short wave infrared spectral information for sentinel-2a imagery. **Remote Sensing**, MDPI, v. 13, n. 1, p. 157, 2021. Disponível em: <<https://doi.org/10.3390/rs13010157>>.

\_\_\_\_\_. Thin cloud removal fusing full spectral and spatial features for sentinel-2 imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 15, p. 8759–8775, 2022. Disponível em: <<https://doi.org/10.1109/JSTARS.2022.3211857>>.

LI, M. et al. Lge-cr: Integrating non-local gated attention and two-flow residual dense learning for cloud removal. **IEEE Geoscience and Remote Sensing Letters**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/LGRS.2024.3422804>>.

\_\_\_\_\_. Decloudnet: Cross-patch consistency is a non-trivial problem for thin cloud removal from wide-swath multi-spectral images. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3427788>>.

LI, R. et al. Multiattention network for semantic segmentation of fine-resolution remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–13, 2021. Disponível em: <<https://doi.org/10.1109/TGRS.2021.3093977>>.

LI, W.; LI, Y.; CHAN, J. C.-W. Thick cloud removal with optical and sar imagery via convolutional-mapping-deconvolutional network. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 58, n. 4, p. 2865–2879, 2019. Disponível em: <<https://doi.org/10.1109/TGRS.2019.2956959>>.

LI, W. et al. Thin cloud removal with residual symmetrical concatenation network. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 153, p. 137–150, 2019. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2019.05.003>>.

LI, X. et al. Generating high-quality and high-resolution seamless satellite imagery for large-scale urban regions. **Remote Sensing**, MDPI, v. 12, n. 1, p. 81, 2019. Disponível em: <<https://doi.org/10.3390/rs12010081>>.

\_\_\_\_\_. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. **ISPRS journal of photogrammetry and remote sensing**, Elsevier, v. 148, p. 103–113, 2019. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2018.12.013>>.

\_\_\_\_\_. Hf-t2cr: High-fidelity thin and thick cloud removal in optical satellite images through sar fusion. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3416128>>.

LI, Y. et al. Hs2p: Hierarchical spectral and structure-preserving fusion network for multimodal remote sensing image cloud and shadow removal. **Information Fusion**, Elsevier, v. 94, p. 215–228, 2023. Disponível em: <<https://doi.org/10.1016/j.inffus.2023.02.002>>.

LI, Z. et al. Thick cloud removal in high-resolution satellite images using stepwise radiometric adjustment and residual correction. **Remote Sensing**, MDPI, v. 11, n. 16, p. 1925, 2019. Disponível em: <<https://doi.org/10.3390/rs11161925>>.

LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. **Remote sensing and image interpretation**. [S.l.]: John Wiley & Sons, 2015.

LIN, D. et al. A remote sensing image dataset for cloud removal. **arXiv preprint arXiv:1901.00600**, 2019. Disponível em: <<https://doi.org/10.48550/arXiv.1901.00600>>.

- LIN, J. et al. Robust thick cloud removal for multitemporal remote sensing images using coupled tensor factorization. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–16, 2022. Disponível em: <<https://doi.org/10.1109/TGRS.2022.3140800>>.
- LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical image analysis**, Elsevier, v. 42, p. 60–88, 2017. Disponível em: <<https://doi.org/10.1016/j.media.2017.07.005>>.
- LIU, H.; HUANG, B.; CAI, J. Thick cloud removal under land cover changes using multisource satellite imagery and a spatiotemporal attention network. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 61, p. 1–18, 2023. Disponível em: <<https://doi.org/10.1109/TGRS.2023.3236106>>.
- LIU, J.; PAN, B.; SHI, Z. Cascaded memory network for optical remote sensing imagery cloud removal. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3376609>>.
- LIU, J. et al. Residual denoising diffusion models. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2024. p. 2773–2783. Disponível em: <<https://doi.org/10.1109/CVPR52733.2024.00268>>.
- LIU, R. et al. Transfusion-cr: Two-phase sar-to-optical translation and deep feature fusion for cloud removal. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 62, p. 1–11, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3439854>>.
- LOSHCHILOV, I.; HUTTER, F. **Decoupled Weight Decay Regularization**. 2019. Disponível em: <<https://doi.org/10.48550/arXiv.1711.05101>>.
- LYU, H.; QIAN, J.; WANG, Y. Study on the unified theory of thin cloud detection and removal based on physical model. In: IEEE. **IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium**. 2024. p. 5549–5552. Disponível em: <<https://doi.org/10.1109/IGARSS53475.2024.10642238>>.
- MA, D. et al. Cloud removal from satellite images using a deep learning model with the cloud-matting method. **Remote Sensing**, MDPI, v. 15, n. 4, p. 904, 2023. Disponível em: <<https://doi.org/10.3390/rs15040904>>.
- MA, J. et al. Sct-cr: A synergistic convolution-transformer modeling method using sar-optical data fusion for cloud removal. **International Journal of Applied Earth Observation and Geoinformation**, Elsevier, v. 130, p. 103909, 2024. Disponível em: <<https://doi.org/10.1016/j.jag.2024.103909>>.
- MARASTONI, N.; GIACOBAZZI, R.; PREDA, M. D. A deep learning approach to program similarity. In: **Proceedings of the 1st international workshop on machine learning and software engineering in symbiosis**. [s.n.], 2018. p. 26–35. Disponível em: <<https://doi.org/10.1145/3243127.3243131>>.
- MARI, J. F. **SIN393: Introdução à Visão Computacional (2023)**. GitHub, 2023. Disponível em: <[https://github.com/joaofmari/SIN393\\_Introduction-to-computer-vision\\_2023](https://github.com/joaofmari/SIN393_Introduction-to-computer-vision_2023)>.

- MERANER, A. et al. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 166, p. 333–346, 2020. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2020.05.013>>.
- MINAEI, S. et al. Image segmentation using deep learning: A survey. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 44, n. 7, p. 3523–3542, 2021. Disponível em: <<https://doi.org/10.1109/TPAMI.2021.3059968>>.
- MOHAJERANI, S.; KRAMMER, T. A.; SAEEDI, P. Cloud detection algorithm for remote sensing images using fully convolutional neural networks. **arXiv preprint arXiv:1810.05782**, 2018. Disponível em: <<https://doi.org/10.48550/arXiv.1810.05782>>.
- MOHAJERANI, S.; SAEEDI, P. Cloud-net: An end-to-end cloud detection algorithm for landsat 8 imagery. In: IEEE. **IGARSS IEEE International Geoscience and Remote Sensing Symposium**. 2019. p. 1029–1032. Disponível em: <<https://doi.org/10.1109/IGARSS.2019.8898776>>.
- MOHER, D. et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. **International journal of surgery**, Elsevier, v. 8, n. 5, p. 336–341, 2010. Disponível em: <<https://doi.org/10.1016/j.ijssu.2010.02.007>>.
- MOREIRA, R. et al. Packet vision: a convolutional neural network approach for network traffic classification. In: IEEE. **2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. 2020. p. 256–263. Disponível em: <<https://doi.org/10.1109/SIBGRAPI51738.2020.00042>>.
- NGATCHOU, P.; ZAREI, A.; EL-SHARKAWI, A. Pareto multi objective optimization. In: IEEE. **Proceedings of the 13th international conference on, intelligent systems application to power systems**. 2005. p. 84–91. Disponível em: <<https://doi.org/10.1109/ISAP.2005.1599245>>.
- NIU, Z.; ZHONG, G.; YU, H. A review on the attention mechanism of deep learning. **Neurocomputing**, Elsevier, v. 452, p. 48–62, 2021. Disponível em: <<https://doi.org/10.1016/j.neucom.2021.03.091>>.
- NOGUEIRA, K.; PENATTI, O. A.; SANTOS, J. A. D. Towards better exploiting convolutional neural networks for remote sensing scene classification. **Pattern Recognition**, Elsevier, v. 61, p. 539–556, 2017. Disponível em: <<https://doi.org/10.1016/j.patcog.2016.07.001>>.
- PAGE, M. J. et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. **bmj**, British Medical Journal Publishing Group, v. 372, 2021. Disponível em: <<https://doi.org/10.1136/bmj.n71>>.
- PAHLEVAN, N. et al. Sentinel-2/landsat-8 product consistency and implications for monitoring aquatic systems. **Remote sensing of Environment**, Elsevier, v. 220, p. 19–29, 2019. Disponível em: <<https://doi.org/10.1016/j.rse.2018.10.027>>.
- \_\_\_\_\_. Simultaneous retrieval of selected optical water quality indicators from landsat-8, sentinel-2, and sentinel-3. **Remote Sensing of Environment**, Elsevier, v. 270, p. 112860, 2022. Disponível em: <<https://doi.org/10.1016/j.rse.2021.112860>>.

- PAULA, A. et al. Could spaceborne multispectral data be used as smart proxies' indicators for agroforestry ecosystem services?—a systematic review. **Ecological Indicators**, Elsevier, v. 156, p. 111162, 2023. Disponível em: <<https://doi.org/10.1016/j.ecolind.2023.111162>>.
- PELLEGRINA, H. S. Trade, productivity, and the spatial organization of agriculture: Evidence from brazil. **Journal of Development Economics**, Elsevier, v. 156, p. 102816, 2022. Disponível em: <<https://doi.org/10.1016/j.jdeveco.2021.102816>>.
- PETCHROMPO, S. et al. A review of pareto pruning methods for multi-objective optimization. **Computers & Industrial Engineering**, Elsevier, v. 167, p. 108022, 2022. Disponível em: <<https://doi.org/10.1016/j.cie.2022.108022>>.
- PICCIALLI, F. et al. A survey on deep learning in medicine: Why, how and when? **Information Fusion**, Elsevier, v. 66, p. 111–137, 2021. Disponível em: <<https://doi.org/10.1016/j.inffus.2020.09.006>>.
- PINAYA, W. H. et al. Brain imaging generation with latent diffusion models. In: SPRINGER. **MICCAI Workshop on Deep Generative Models**. 2022. p. 117–126. Disponível em: <[https://doi.org/10.1007/978-3-031-18576-2\\_12](https://doi.org/10.1007/978-3-031-18576-2_12)>.
- PODSIADLO, I.; PARIS, C.; BRUZZONE, L. A study of the robustness of the long short-term memory classifier to cloudy time series of multispectral images. In: SPIE. **Image and Signal Processing for Remote Sensing XXVI**. 2020. v. 11533, p. 335–343. Disponível em: <<https://doi.org/10.1117/12.2574383>>.
- PONTI, M. A. et al. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: **SIBGRAPI Tutorials**. IEEE Computer Society, 2017. p. 17–41. ISBN 978-1-5386-0619-3. Disponível em: <<https://doi.org/10.1109/SIBGRAPI-T.2017.12>>.
- RAHMAN, M. M.; SZABÓ, G. Multi-objective urban land use optimization using spatial data: A systematic review. **Sustainable Cities and Society**, Elsevier, v. 74, p. 103214, 2021. Disponível em: <<https://doi.org/10.1016/j.scs.2021.103214>>.
- ROBBINS, H.; MONRO, S. A stochastic approximation method. **The annals of mathematical statistics**, JSTOR, p. 400–407, 1951. Disponível em: <<https://www.jstor.org/stable/2236626>>.
- ROMBACH, R. et al. High-resolution image synthesis with latent diffusion models. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [s.n.], 2022. p. 10684–10695. Disponível em: <<https://doi.org/10.1109/CVPR52688.2022.01042>>.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. 2015. p. 234–241. Disponível em: <[https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)>.
- RUCHTE, M.; GRABOCKA, J. Scalable pareto front approximation for deep multi-objective learning. In: IEEE. **2021 IEEE international conference on data mining (ICDM)**. 2021. p. 1306–1311. Disponível em: <<https://doi.org/10.1109/ICDM51629.2021.00162>>.

- RUDER, S. **An overview of gradient descent optimization algorithms**. 2017. Disponível em: <<https://doi.org/10.48550/arXiv.1609.04747>>.
- SAHARIA, C. et al. Palette: Image-to-image diffusion models. In: **ACM SIGGRAPH 2022 conference proceedings**. [s.n.], 2022. p. 1–10. Disponível em: <<https://doi.org/10.1145/3528233.3530757>>.
- SANCHES, I. D. et al. Campo verde database: Seeking to improve agricultural remote sensing of tropical areas. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 15, n. 3, p. 369–373, 2018. Disponível em: <<https://doi.org/10.1109/LGRS.2017.2789120>>.
- SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2018. p. 4510–4520. Disponível em: <<https://doi.org/10.1109/CVPR.2018.00474>>.
- SARÈ, S. B. N. et al. Joint cloud removal and classification of sentinel-2 image time series for agricultural land cover mapping in northern benin. In: IEEE. **IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium**. 2024. p. 4824–4827. Disponível em: <<https://doi.org/10.1109/IGARSS53475.2024.10641427>>.
- SARUKKAI, V. et al. Cloud removal from satellite images using spatiotemporal generator networks. In: **Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision**. [s.n.], 2020. p. 1796–1805. Disponível em: <<https://doi.org/10.1109/WACV45572.2020.9093564>>.
- SCHMITT, M. et al. Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, IV-2/W7, p. 153–160, 2019. Disponível em: <<https://doi.org/10.5194/isprs-annals-IV-2-W7-153-2019>>.
- SCHMITT, M.; HUGHES, L. H.; ZHU, X. X. The sen1-2 dataset for deep learning in sar-optical data fusion. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, IV-1, p. 141–146, 2018. Disponível em: <<https://doi.org/10.5194/isprs-annals-IV-1-141-2018>>.
- SEBASTIANELLI, A. et al. Spatio-temporal sar-optical data fusion for cloud removal via a deep hierarchical model. **arXiv preprint arXiv:2106.12226**, 2021. Disponível em: <<https://doi.org/10.48550/arXiv.2106.12226>>.
- \_\_\_\_\_. Plfm: Pixel-level merging of intermediate feature maps by disentangling and fusing spatial and temporal data for cloud removal. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–16, 2022. Disponível em: <<https://doi.org/10.1109/TGRS.2022.3208694>>.
- SEGAL-ROZENHAIMER, M. et al. Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (cnn). **Remote Sensing of Environment**, Elsevier, v. 237, p. 111446, 2020. Disponível em: <<https://doi.org/10.1016/j.rse.2019.111446>>.

- SEYREK, E. C.; UYSAL, M. A comparative analysis of various activation functions and optimizers in a convolutional neural network for hyperspectral image classification. **Multimedia Tools and Applications**, Springer, v. 83, n. 18, p. 53785–53816, 2024. Disponível em: <<https://doi.org/10.1007/s11042-023-17546-5>>.
- SHEN, H. et al. A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 12, n. 3, p. 862–874, 2019. Disponível em: <<https://doi.org/10.1109/JSTARS.2019.2898348>>.
- SHEPP, L. A.; LOGAN, B. F. The fourier reconstruction of a head section. **IEEE Transactions on nuclear science**, IEEE, v. 21, n. 3, p. 21–43, 1974. Disponível em: <<https://doi.org/10.1109/TNS.1974.6499235>>.
- SHI, S. et al. A novel thin cloud removal method based on multiscale dark channel prior (mdcp). **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 19, p. 1–5, 2021. Disponível em: <<https://doi.org/10.1109/LGRS.2021.3062853>>.
- SHI, W. et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2016. p. 1874–1883. Disponível em: <<https://doi.org/10.1109/CVPR.2016.207>>.
- SILVA, L. H. F. P. et al. Non-linear co-registration in uavs' images using deep learning. In: IEEE. **2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. 2022. v. 1, p. 1–6. Disponível em: <<https://doi.org/10.1109/SIBGRAPI55357.2022.9991781>>.
- \_\_\_\_\_. Evaluating combinations of optimizers and loss functions for cloud removal using diffusion models. In: **Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP**. [s.n.], 2025. p. 648–656. ISBN 978-989-758-728-3. ISSN 2184-4321. Disponível em: <<https://doi.org/10.5220/0013252100003912>>.
- SILVA, M. V. da et al. Generating synthetic multispectral images using neural style transfer: A study with application in channel alignment. **Computers and Electronics in Agriculture**, Elsevier, v. 206, p. 107668, 2023. Disponível em: <<https://doi.org/https://doi.org/10.1016/j.compag.2023.107668>>.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Disponível em: <<https://doi.org/10.48550/arXiv.1409.1556>>.
- STOCK, A. et al. Comparison of cloud-filling algorithms for marine satellite data. **Remote Sensing**, MDPI, v. 12, n. 20, p. 3313, 2020. Disponível em: <<https://doi.org/10.3390/rs12203313>>.
- STUBENRAUCH, C. J. et al. Assessment of global cloud datasets from satellites: Project and database initiated by the gewex radiation panel. **Bulletin of the American Meteorological Society**, American Meteorological Society, v. 94, n. 7, p. 1031–1049, 2013. Disponível em: <<https://doi.org/10.1175/BAMS-D-12-00117.1>>.

SUI, J. et al. Diffusion enhancement for cloud removal in ultra-resolution remote sensing imagery. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3411671>>.

\_\_\_\_\_. Semantic distortion-aware network with cloud classification for remote sensing cloud removal. In: IEEE. **IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium**. 2024. p. 7363–7366. Disponível em: <<https://doi.org/10.1109/IGARSS53475.2024.10640875>>.

SUN, D.-L.; JI, T.-Y.; DING, M. A new sparse collaborative low-rank prior knowledge representation for thick cloud removal in remote sensing images. **Remote Sensing**, MDPI, v. 16, n. 9, p. 1518, 2024. Disponível em: <<https://doi.org/10.3390/rs16091518>>.

SUN, L. et al. Cloud-aware generative network: Removing cloud from optical remote sensing images. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 17, n. 4, p. 691–695, 2019. Disponível em: <<https://doi.org/10.1109/LGRS.2019.2928840>>.

SZEGEDY, C. et al. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2015. p. 1–9. Disponível em: <<https://doi.org/10.1109/CVPR.2015.7298594>>.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International conference on machine learning**. 2019. p. 6105–6114. Disponível em: <<https://doi.org/10.48550/arXiv.1905.11946>>.

TAN, Z. C. et al. Unsupervised remote sensing image thin cloud removal method based on contrastive learning. **IET Image Processing**, Wiley Online Library, v. 18, n. 7, p. 1844–1861, 2024. Disponível em: <<https://doi.org/10.1049/ipr2.13067>>.

TIAN, Y. et al. Recent advances on loss functions in deep learning for computer vision. **Neurocomputing**, Elsevier, v. 497, p. 129–158, 2022. Disponível em: <<https://doi.org/10.1016/j.neucom.2022.04.127>>.

TOTH, C.; JÓZKÓW, G. Remote sensing platforms and sensors: A survey. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 115, p. 22–36, 2016. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2015.10.004>>.

TREBING, K.; STANCZYK, T.; MEHRKANOON, S. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. **Pattern Recognition Letters**, Elsevier, v. 145, p. 178–186, 2021. Disponível em: <<https://doi.org/10.1016/j.patrec.2021.01.036>>.

TRICCO, A. C. et al. Prisma extension for scoping reviews (prisma-scr): checklist and explanation. **Annals of internal medicine**, American College of Physicians, v. 169, n. 7, p. 467–473, 2018. Disponível em: <<https://doi.org/10.7326/M18-0850>>.

TU, Z. et al. Fully-connected tensor network decomposition and group sparsity for multitemporal images cloud removal. **Inverse Problems and Imaging**, Inverse Problems and Imaging, p. 59–86, 2024. Disponível em: <<https://doi.org/10.3934/ipi.2024025>>.

WANG, M. et al. Idf-cr: Iterative diffusion process for divide-and-conquer cloud removal in remote-sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3378720>>.



- WANG, Q. et al. A comprehensive survey of loss functions in machine learning. **Annals of Data Science**, Springer, p. 1–26, 2020. Disponível em: <<https://doi.org/10.1007/s40745-020-00253-5>>.
- WANG, S. et al. Linformer: Self-attention with linear complexity. **arXiv preprint arXiv:2006.04768**, 2020. Disponível em: <<https://doi.org/10.48550/arXiv.2006.04768>>.
- WANG, X. et al. Thin and thick cloud removal on remote sensing image by conditional generative adversarial network. In: IEEE. **IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium**. 2019. p. 1426–1429. Disponível em: <<https://doi.org/10.1109/IGARSS.2019.8897958>>.
- WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. **IEEE signal processing magazine**, IEEE, v. 26, n. 1, p. 98–117, 2009. Disponível em: <<https://doi.org/10.1109/MSP.2008.930649>>.
- WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. **IEEE transactions on image processing**, IEEE, v. 13, n. 4, p. 600–612, 2004. Disponível em: <<https://doi.org/10.1109/TIP.2003.819861>>.
- \_\_\_\_\_. Virtual image-based cloud removal for landsat images. **GIScience & Remote Sensing**, Taylor & Francis, v. 60, n. 1, p. 2160411, 2023. Disponível em: <<https://doi.org/10.1080/15481603.2022.2160411>>.
- WEN, X. et al. Generative adversarial learning in yuv color space for thin cloud removal on satellite imagery. **Remote Sensing**, MDPI, v. 13, n. 6, p. 1079, 2021. Disponível em: <<https://doi.org/10.3390/rs13061079>>.
- \_\_\_\_\_. An effective network integrating residual learning and channel attention mechanism for thin cloud removal. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 19, p. 1–5, 2022. Disponível em: <<https://doi.org/10.1109/LGRS.2022.3161062>>.
- WU, C. et al. Multi-stage frequency attention network for progressive optical remote sensing cloud removal. **Remote Sensing**, MDPI, v. 16, n. 15, p. 2867, 2024. Disponível em: <<https://doi.org/10.3390/rs16152867>>.
- WU, P. et al. Cloudformer: A cloud-removal network combining self-attention mechanism and convolution. **Remote Sensing**, MDPI, v. 14, n. 23, p. 6132, 2022. Disponível em: <<https://doi.org/10.3390/rs14236132>>.
- WYLIE, D. et al. Trends in global cloud cover in two decades of hirs observations. **Journal of climate**, v. 18, n. 15, p. 3021–3031, 2005. Disponível em: <<https://doi.org/10.1175/JCLI3461.1>>.
- WYLIE, D. P.; MENZEL, W. Two years of cloud cover statistics using vas. **Journal of Climate**, v. 2, n. 4, p. 380–392, 1989. Disponível em: <[https://doi.org/10.1175/1520-0442\(1989\)002<0380:TYOCCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<0380:TYOCCS>2.0.CO;2)>.
- XIA, B. et al. Diffir: Efficient diffusion model for image restoration. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. [s.n.], 2023. p. 13095–13105. Disponível em: <<https://doi.org/10.1109/ICCV51070.2023.01204>>.

XIA, Y. et al. Cloud removal of optical remote sensing imagery with multitemporal sar-optical data using x-mtgan. In: IEEE. **IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium**. 2019. p. 3396–3399. Disponível em: <<https://doi.org/10.1109/IGARSS.2019.8899105>>.

XIANG, X.; TAN, Y. Unpaired thick cloud removal with multiscale sar-optical data fusion. In: IEEE. **2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)**. 2024. p. 665–668. Disponível em: <<https://doi.org/10.1109/CVIDL62147.2024.10604277>>.

XIANG, X.; TAN, Y.; YAN, L. Cloud-guided fusion with sar-to-optical translation for thick cloud removal. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3431556>>.

XIAO, X.; LU, Y. Cloud removal of optical remote sensing imageries using sar data and deep learning. In: IEEE. **2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)**. 2021. p. 1–5. Disponível em: <<https://doi.org/10.1109/APSAR52370.2021.9688535>>.

XIAO, Y. et al. Ediffr: An efficient diffusion probabilistic model for remote sensing image super-resolution. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2023. Disponível em: <<https://doi.org/10.1109/TGRS.2023.3341437>>.

XIE, Y. et al. Auto-cm: Unsupervised deep learning for satellite imagery composition and cloud masking using spatio-temporal dynamics. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [s.n.], 2023. p. 14575–14583. Disponível em: <<https://doi.org/10.1609/aaai.v37i12.26704>>.

XU, F. et al. Glf-cr: Sar-enhanced cloud removal with global–local fusion. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 192, p. 268–278, 2022. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2022.08.002>>.

XU, M. et al. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 149, p. 215–225, 2019. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2019.01.025>>.

XU, P.; ZHU, X.; CLIFTON, D. A. Multimodal learning with transformers: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2023. Disponível em: <<https://doi.org/10.1109/TPAMI.2023.3275156>>.

XU, Z. et al. Cloudy image arithmetic: A cloudy scene synthesis paradigm with an application to deep-learning-based thin cloud removal. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 60, p. 1–16, 2021. Disponível em: <<https://doi.org/10.1109/TGRS.2021.3122253>>.

YAN, Y. et al. Pnbt-cr: A cloud removal method for ship detection. **IEEE Geoscience and Remote Sensing Letters**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/LGRS.2024.3361150>>.

YANG, L. et al. Diffusion models: A comprehensive survey of methods and applications. **ACM Computing Surveys**, ACM New York, NY, USA, v. 56, n. 4, p. 1–39, 2023. Disponível em: <<https://doi.org/10.1145/3626235>>.

- YANG, Q. et al. Multi-scale deep residual learning for cloud removal. In: IEEE. **IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium**. 2020. p. 4967–4970. Disponível em: <<https://doi.org/10.1109/IGARSS39084.2020.9323261>>.
- YANG, X. et al. Weakly-supervised cloud detection and effective cloud removal for remote sensing images. **Journal of Visual Communication and Image Representation**, Elsevier, v. 98, p. 104006, 2024. Disponível em: <<https://doi.org/10.1016/j.jvcir.2023.104006>>.
- YATES, M. et al. Evaluation of synthetic aerial imagery using unconditional generative adversarial networks. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 190, p. 231–251, 2022. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2022.06.010>>.
- YU, W.; ZHANG, X.; PUN, M.-O. Cloud removal in optical remote sensing imagery using multiscale distortion-aware networks. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 19, p. 1–5, 2022. Disponível em: <<https://doi.org/10.1109/LGRS.2022.3144686>>.
- ZHANG, H. K. et al. Characterization of sentinel-2a and landsat-8 top of atmosphere, surface, and nadir brdf adjusted reflectance and ndvi differences. **Remote sensing of environment**, Elsevier, v. 215, p. 482–494, 2018. Disponível em: <<https://doi.org/10.1016/j.rse.2018.04.031>>.
- ZHANG, Q. et al. U-seqnet: learning spatiotemporal mapping relationships for multimodal multitemporal cloud removal. **GIScience & Remote Sensing**, Taylor & Francis, v. 61, n. 1, p. 2330185, 2024. Disponível em: <<https://doi.org/10.1080/15481603.2024.2330185>>.
- \_\_\_\_\_. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 162, p. 148–160, 2020. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2020.02.008>>.
- \_\_\_\_\_. Combined deep prior with low-rank tensor svd for thick cloud removal in multitemporal images. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 177, p. 161–173, 2021. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2021.04.021>>.
- ZHANG, W. et al. A lightweight machine-learning method for cloud removal in remote sensing images constrained by conditional information. **Remote Sensing**, MDPI, v. 16, n. 17, p. 3134, 2024. Disponível em: <<https://doi.org/10.3390/rs16173134>>.
- ZHANG, X. et al. Removing cloud cover interference from sentinel-2 imagery in google earth engine by fusing sentinel-1 sar data with a cnn model. **International Journal of Remote Sensing**, Taylor & Francis, v. 43, n. 1, p. 132–147, 2022. Disponível em: <<https://doi.org/10.1080/01431161.2021.2012295>>.
- ZHANG, Y. et al. A flexible spatiotemporal thick cloud removal method with low requirements for reference images. **Remote Sensing**, MDPI, v. 15, n. 17, p. 4306, 2023. Disponível em: <<https://doi.org/10.3390/rs15174306>>.

- \_\_\_\_\_. Mr image super-resolution with squeeze and excitation reasoning attention network. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [s.n.], 2021. p. 13425–13434. Disponível em: <<https://doi.org/10.1109/CVPR46437.2021.01322>>.
- \_\_\_\_\_. A coarse-to-fine framework for cloud removal in remote sensing image sequence. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 57, n. 8, p. 5963–5974, 2019. Disponível em: <<https://doi.org/10.1109/TGRS.2019.2903594>>.
- \_\_\_\_\_. Double rank-one prior: Thin cloud removal by visible bands. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3396874>>.
- \_\_\_\_\_. Remote sensing image cloud removal by deep image prior with a multitemporal constraint. **Optics Continuum**, Optica Publishing Group, v. 1, n. 2, p. 215–226, 2022. Disponível em: <<https://doi.org/10.1364/OPTCON.439671>>.
- ZHAO, B. et al. Pm-lsmn: A physical-model-based lightweight self-attention multiscale net for thin cloud removal. **IEEE Geoscience and Remote Sensing Letters**, IEEE, 2024. Disponível em: <<https://doi.org/10.1109/LGRS.2024.3403674>>.
- ZHAO, H. et al. Pyramid scene parsing network. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2017. p. 2881–2890. Disponível em: <<https://doi.org/10.1109/CVPR.2017.660>>.
- ZHAO, M.; OLSEN, P.; CHANDRA, R. Seeing through clouds in satellite images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 61, p. 1–16, 2023. Disponível em: <<https://doi.org/10.1109/TGRS.2023.3239592>>.
- ZHAO, X.; JIA, K. Cloud removal in remote sensing using sequential-based diffusion models. **Remote Sensing**, MDPI, v. 15, n. 11, p. 2861, 2023. Disponível em: <<https://doi.org/10.3390/rs15112861>>.
- ZHAO, Y. et al. Cloud removal using multimodal gan with adversarial consistency loss. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 19, p. 1–5, 2021. Disponível em: <<https://doi.org/10.1109/LGRS.2021.3093887>>.
- ZHENG, J.; LIU, X.-Y.; WANG, X. Single image cloud removal using u-net and generative adversarial networks. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 59, n. 8, p. 6371–6385, 2020. Disponível em: <<https://doi.org/10.1109/TGRS.2020.3027819>>.
- ZHENG, W.-J. et al. Feature-domain fidelity and tensor low-rank regularization for cloud removal in remote sensing images. In: **IEEE. IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium**. 2024. p. 8274–8277. Disponível em: <<https://doi.org/10.1109/IGARSS53475.2024.10641683>>.
- \_\_\_\_\_. Spatial-spectral-temporal connective tensor network decomposition for thick cloud removal. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 199, p. 182–194, 2023. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2023.04.006>>.

ZHOU, B.; WANG, Y. A thin-cloud removal approach combining the cirrus band and rtm-based algorithm for landsat-8 oli data. In: IEEE. **IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium**. 2019. p. 1434–1437. Disponível em: <<https://doi.org/10.1109/IGARSS.2019.8898644>>.

ZHOU, J. et al. Cloud removal for optical remote sensing imagery using distortion coding network combined with compound loss functions. **Remote Sensing**, MDPI, v. 14, n. 14, p. 3452, 2022. Disponível em: <<https://doi.org/10.3390/rs14143452>>.

ZHU, X. X. et al. Deep learning in remote sensing: A comprehensive review and list of resources. **IEEE geoscience and remote sensing magazine**, IEEE, v. 5, n. 4, p. 8–36, 2017. Disponível em: <<https://doi.org/10.1109/MGRS.2017.2762307>>.

ZI, Y. et al. Thick cloud removal in multitemporal remote sensing images using a coarse-to-fine framework. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 21, p. 1–5, 2024. Disponível em: <<https://doi.org/10.1109/LGRS.2024.3378691>>.

\_\_\_\_\_. Thin cloud removal for remote sensing images using a physical-model-based cyclegan with unpaired data. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 19, p. 1–5, 2021. Disponível em: <<https://doi.org/10.1109/LGRS.2021.3140033>>.

\_\_\_\_\_. Thin cloud removal for multispectral remote sensing images using convolutional neural networks combined with an imaging model. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 14, p. 3811–3823, 2021. Disponível em: <<https://doi.org/10.1109/JSTARS.2021.3068166>>.

ZOU, X. et al. Pmaa: A progressive multi-scale attention autoencoder model for high-performance cloud removal from multi-temporal satellite imagery. **European Conference on Artificial Intelligence (ECAI)**, 2023. Disponível em: <<https://doi.org/10.3233/FAIA230636>>.

\_\_\_\_\_. Diffcr: A fast conditional diffusion framework for cloud removal from optical satellite images. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 62, p. 1–14, 2024. Disponível em: <<https://doi.org/10.1109/TGRS.2024.3365806>>.