
Segmentação Semântica de Incêndios Florestais com U-Net e Encoders Pré-Treinados: Uma Abordagem com Função de Perda Composta e Balanceamento de Classes

Natália de Oliveira Gomes



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

Natália de Oliveira Gomes

**Segmentação Semântica de Incêndios Florestais
com U-Net e Encoders Pré-Treinados: Uma
Abordagem com Função de Perda Composta e
Balanceamento de Classes**

Dissertação de mestrado 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 Mestre em Ciência da
Computação.

Área de concentração: Ciência da Computação

Orientador: Henrique Coelho Fernandes

Uberlândia
2025

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

G633
2025

Segmentação Semântica de Incêndios Florestais com U-Net e Encoders Pré-Treinados [recurso eletrônico] : Uma Abordagem com Função de Perda Composta e Balançamento de Classes / Natália de Oliveira Gomes. - 2025.

Orientador: Henrique Coelho Fernandes.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
DOI <http://doi.org/10.14393/ufu.di.2025.538>
Inclui bibliografia.

1. Computação. I. Fernandes, Henrique Coelho, 1986-, (Orient.).
II. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. III. 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



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação, 28/2025, PPGCO				
Data:	25 de Agosto de 2025	Hora de início:	09:00	Hora de encerramento:	10:59
Matrícula do Discente:	12312CCP025				
Nome do Discente:	Natália de Oliveira Gomes				
Título do Trabalho:	Segmentação Semântica de Incêndios Florestais com U-Net e Encoders Pré-Treinados: Uma Abordagem com Função de Perda Composta e Balanceamento de Classes.				
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: Claudiney Ramos Tinoco - FACOM/UFU, João Fernando Mari - UFV e Henrique Coelho Fernandes - Cranfield University, orientador do(a) candidato(a).

Os examinadores participaram desde as seguintes localidades: Henrique Coelho Fernandes - Milton Keynes/Inglaterra, João Fernando Mari - Rio Paranaíba/MG e Claudiney Ramos Tinoco - Uberlândia/MG. O(a) aluno(a) Natália de Oliveira Gomes participou de Recife/PE.

Iniciando os trabalhos o(a) presidente da mesa, Prof. Dr. Henrique Coelho Fernandes, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao(á) Discente a palavra para a exposição do seu trabalho. A duração da apresentação do(a) 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(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato(a):

Aprovado

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

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 **João Fernando Mari, Usuário Externo**, em 25/08/2025, às 15:35, 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 25/08/2025, às 15:45, 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 **Claudiney Ramos Tinoco, Professor(a) do Magistério Superior**, em 27/08/2025, às 19:12, 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, informando o código verificador **6590684** e o código CRC **A055FB02**.

*Dedico este trabalho à minha família, que,
com gestos de carinho e palavras de ânimo,
regou o solo fértil onde germinaram meus sonhos.*

Agradecimentos

Agradeço, de forma especial, ao meu orientador, Prof. Henrique Fernandes, pelas valiosas contribuições, pela generosidade em compartilhar seu tempo, conhecimento e experiência, bem como pela paciência, apoio constante e pelas palavras de incentivo que tanto me motivaram ao longo desta jornada. Sua orientação foi fundamental para a realização deste trabalho.

Registro também minha gratidão ao Prof. Bruno Travençolo, por ter contribuído significativamente nos passos iniciais deste trabalho, oferecendo direcionamentos importantes e incentivando meu desenvolvimento acadêmico.

À minha família, minha base e inspiração, por caminhar ao meu lado durante toda esta trajetória. Sou profundamente grata pelo amor incondicional, pela compreensão e pelo apoio nos momentos mais desafiadores, que tornaram possível a concretização deste sonho.

*“Somos a primeira geração a sentir os efeitos das mudanças climáticas e a última que
pode fazer algo a respeito.”
(Barack Obama)*

Resumo

A segmentação semântica de incêndios florestais em imagens de satélite é um problema relevante para o monitoramento ambiental, especialmente devido ao forte desbalanceamento entre classes e à natureza esparsa dos focos de fogo. Neste trabalho, o estudo é conduzido sobre imagens multiespectrais do satélite Sentinel-2 (resolução de 20 m), com foco na identificação da classe *incêndio*. O objetivo desta dissertação é propor, implementar e avaliar um *pipeline* de segmentação semântica para detectar incêndios florestais, investigando o impacto de diferentes *backbones*, funções de perda e estratégias de mitigação do desbalanceamento. A metodologia foi estruturada em dois ciclos experimentais: no primeiro, de caráter exploratório, avaliou-se a U-Net com *encoders Visual Geometry Group 16* (VGG-16) e EfficientNetB3, utilizando perdas tradicionais; no segundo, foi adotado um pipeline otimizado com *undersampling*, *data augmentation*, perdas sensíveis à classe minoritária (*Dice*, *Focal* e *Dice+Focal*), treinamento em duas fases com *fine-tuning* e calibração do limiar por varredura na validação maximizando *F1-Score*. Os resultados indicam melhora substancial no Ciclo 2, com maior estabilidade e desempenho, destacando-se a U-Net com *backbone* VGG-16, que apresentou métricas elevadas de medida F1 (média harmônica entre Precisão e *Precision/Recall*) (*F1-Score*) e Interseção sobre União (IoU) no conjunto de teste. A análise qualitativa confirmou maior coerência espacial das máscaras preditas. Conclui-se que a combinação de balanceamento, funções de perda adequadas e calibração do limiar de decisão é determinante para a segmentação eficaz de incêndios em imagens multiespectrais do Sentinel-2.

Palavras-chave: Sensoriamento Remoto, Segmentação Semântica, Aprendizado Profundo, Queimadas, Sentinel-2, Incêndios Florestais. .

Abstract

Semantic segmentation of forest fires in satellite imagery is a relevant problem for environmental monitoring, particularly due to severe class imbalance and the sparse spatial distribution of fire events. In this work, the study is conducted using multispectral Sentinel-2 images with 20 m spatial resolution, focusing on the accurate identification of the *fire* class. The objective of this dissertation is to propose, implement, and evaluate a semantic segmentation pipeline for forest fire detection, investigating the impact of different backbones, loss functions, and class imbalance mitigation strategies. The methodology is structured into two experimental cycles. The first, exploratory in nature, evaluates a U-Net architecture with VGG16 and EfficientNetB3 encoders using conventional loss functions. The second cycle adopts an optimized pipeline incorporating selective undersampling, data augmentation, loss functions tailored to imbalanced data (Dice, Focal, and Dice+Focal), a two-stage training strategy with encoder fine-tuning, and decision threshold calibration through validation-based sweeping to maximize the F1-score. The results show substantial improvements in the second cycle, with increased stability and performance, particularly for the U-Net with the VGG16 backbone, which achieved high F1 and IoU scores on the test set. Qualitative analysis further confirmed improved spatial coherence of the predicted masks. These findings indicate that combining class balancing strategies, appropriate loss functions, and threshold calibration is essential for effective forest fire segmentation in multispectral Sentinel-2 imagery.

Keywords: Remote Sensing, Semantic Segmentation, Deep Learning, Forest Fires, Sentinel-2, Wildfires..

Lista de ilustrações

Figura 1 – Relação hierárquica entre IA, Aprendizado de Máquina, Aprendizado Profundo e CNN. Fonte: Elaborado pela autora.	33
Figura 2 – Bandas espectrais e resoluções do MSI do Sentinel-2. Fonte: Elaborado por Chambrelan e team (2023).	35
Figura 3 – Informação registrada em cada banda do satélite Sentinel-2 para um evento de incêndio. Fonte: Elaborado por Pereira (2021).	36
Figura 4 – Combinações de bandas do satélite Sentinel-2 para um evento de incêndio: (a) bandas 4, 3 e 2 (espectro visível); (b) bandas 12, 11 e 8A (infravermelho de ondas curtas e infravermelho próximo); (c) bandas 1, 9 e 10 (infravermelho e termal). Fonte: Elaborado por Pereira (2021).	37
Figura 5 – Fluxograma da metodologia proposta para a segmentação de incêndios florestais.	48
Figura 6 – Exemplo de uma imagem de satélite do Sentinel-2 no formato <i>False Color</i> e sua respectiva máscara binária, indicando as áreas afetadas pelo incêndio.	50
Figura 7 – Curvas de treinamento e métricas para <i>encoders</i> com <i>Binary Crossentropy</i> no Ciclo 1.	57
Figura 8 – Curvas de treinamento e métricas para <i>Dice Loss</i>	58
Figura 9 – Matrizes de confusão normalizadas na validação com limiares ótimos por função de perda (BCE: 0,45; <i>Dice</i> : 0,40). Ambos os modelos mantêm FP muito baixos devido ao fundo dominante; a diferença central está no <i>recall</i> da classe fogo.	60
Figura 10 – Matrizes de confusão normalizadas no conjunto de teste para U-Net com VGG-16 e EfficientNetB3, avaliadas com limiares escolhidos na validação (BCE: 0,45; <i>Dice</i> : 0,40). Em todos os casos, o fundo apresenta FP ≈ 0 (TN $\approx 1,00$), evidenciando o desequilíbrio de classes.	62
Figura 11 – Curvas de perda e métricas de treino/validação da U-Net com <i>backbone</i> VGG-16 no Ciclo 2, considerando as diferentes funções de perda.	65

Figura 12 – Curvas de perda e métricas de treino/validação da U-Net com <i>backbone EfficientNetB3</i> no <i>Ciclo 2</i> , considerando as diferentes funções de perda.	67
Figura 13 – Matrizes de confusão normalizadas da U-Net com <i>backbone</i> VGG-16 no conjunto de validação do <i>Ciclo 2</i> , considerando as diferentes funções de perda.	70
Figura 14 – Matrizes de confusão normalizadas da U-Net com <i>backbone</i> VGG-16 no conjunto de teste do <i>Ciclo 2</i> , considerando as diferentes funções de perda.	71
Figura 15 – Matrizes de confusão normalizadas da U-Net com <i>backbone</i> EfficientNetB3 no conjunto de validação do <i>Ciclo 2</i> , considerando as diferentes funções de perda.	72
Figura 16 – Matrizes de confusão normalizadas da U-Net com <i>backbone</i> EfficientNetB3 no conjunto de teste do <i>Ciclo 2</i> , considerando as diferentes funções de perda.	74

Lista de tabelas

Tabela 1	– Revisão de Estudos sobre Detecção de Incêndios com Deep Learning .	44
Tabela 2	– Métricas obtidas na validação para a U-Net com <i>encoders</i> e funções de perda (ciclo 1).	55
Tabela 3	– Métricas obtidas no <i>teste</i> para a U-Net com diferentes <i>encoders</i> e funções de perda (ciclo 1). Resultados calculados no limiar ótimo selecionado na validação.	56
Tabela 4	– Métricas na validação (limiares oriundos da validação).	63
Tabela 5	– Métricas no teste (usando os limiares da validação).	63

Lista de siglas

AP Aprendizado Profundo

AVHRR Advanced Very High Resolution Radiometer

BCE Entropia Cruzada Binária

CNN Redes Neurais Convolucionais

CNNs Redes Neurais Convolucionais

CIAR Chaîne Image Autonome et Réactive

DL *Deep Learning*

FP Falsos Positivos

FN Falsos Negativos

F1-Score medida F1 (média harmônica entre Precisão e *Precision/Recall*)

GPUs *Graphics Processing Units*

GANs Redes Adversariais Generativas

HSB *Hue, Saturation, Brightness*

IA Inteligência Artificial

IoU Interseção sobre União

ML *Machine Learning*

MSI *Multi-Spectral Instrument*

NLP *Natural Language Processing*

OLI Operational Land Imager

PR *Precision–Recall* (curva Precisão–Sensibilidade)

SEEG Sistema de Estimativas de Emissões de Gases de Efeito Estufa

SR Sensoriamento Remoto

SAR Radar de Abertura Sintética

SWIR Infravermelho de Ondas Curtas

TMA Anastigmático de Três Espelhos

TP Verdadeiros Positivos

TN Verdadeiros Negativos

VNIR Visível e Infravermelho Próximo

VANTs Veículo Aéreo Não Tripulado

VGG-16 *Visual Geometry Group 16*

Sumário

1	INTRODUÇÃO	25
1.1	Motivação	27
1.2	Objetivos e Desafios da Pesquisa	27
1.3	Hipótese	28
1.4	Contribuições	29
1.5	Organização da Dissertação	30
2	FUNDAMENTAÇÃO TEÓRICA	31
2.1	Métodos de Aprendizado Profundo em Imagens de Sensoria- mento Remoto	31
2.2	Redes Neurais Convolucionais	32
2.3	Aumento de Dados	34
2.4	Sentinel-2	34
2.5	Trabalhos Correlatos	38
3	METODOLOGIA	45
3.1	Aquisição de Dados	49
3.2	Segmentação	50
3.2.1	U-Net (Arquitetura CNN <i>Encoder-Decoder</i>)	50
3.2.2	VGG16 com <i>Transfer Learning</i>	51
3.2.3	EfficientNetB3	51
3.2.4	Métricas	52
4	RESULTADOS	55
4.0.1	Primeiro Ciclo Experimental: Avaliação com <i>Binary Crossentropy</i> e <i>Dice Loss</i>	55
4.0.2	Ciclo 2: objetivos, desenho experimental e comparativo com o Ciclo 1 .	64
5	CONCLUSÃO	75

5.1 Contribuições em Produção Bibliográfica 76

REFERÊNCIAS 79

Introdução

O monitoramento global de incêndios é viabilizado pelo emprego de dados de sensoriamento remoto (SR) (BARMPOUTIS et al., 2020), campo que se refere a uma técnica que envolve a coleta de informações sobre objeto, área ou evento na superfície da Terra sem contato físico. O uso de imagens de satélite é fundamental para a gestão de políticas de conservação ambiental (CHUVIECO et al., 2019). A detecção de focos de incêndio ativo constitui uma área de pesquisa que busca extrair informações críticas dessas imagens para embasar decisões.

Anomalias térmicas identificadas através de dados de satélite são frequentemente ligadas ao uso do solo, como incêndios para manutenção e conversão de terrenos, queimadas florestais, atividades industriais, e outros fenômenos naturais, incluindo erupções vulcânicas. A utilização do fogo é uma prática comum para limpeza de terrenos pós-desmatamento. Essa prática, quando executada sem monitoramento adequado, pode levar a consequências ambientais severas, incluindo a degradação do solo por meio da perda de nutrientes vitais para o crescimento das plantas e impactos adversos na biodiversidade e nos ecossistemas, bem como a deterioração da qualidade do ar (ÖZEL; ALAM; KHAN, 2024).

No Brasil, o problema das queimadas atingiu níveis alarmantes em 2024, de acordo com MapBiomas (2024), iniciativa que faz parte do Sistema de Estimativas de Emissões de Gases de Efeito Estufa (SEEG) do Observatório do Clima. Apenas no mês de agosto, a área total queimada alcançou 5,65 milhões de hectares, representando quase metade (49%) do total queimado desde o início do ano. Este valor corresponde a uma área equivalente ao estado da Paraíba e representa um aumento de 149% em relação ao mesmo período de 2023. Os incêndios impactaram severamente os biomas, com destaque para o Cerrado, que sofreu uma devastação de 2,4 milhões de hectares em agosto de 2024, ou 43% da área total queimada no Brasil no mesmo mês. A situação é agravada pelo fato de que quase dois terços (65%) das queimadas em agosto ocorreram em vegetação nativa, indicando uma relação direta com o desmatamento e a degradação ambiental. A Amazônia, outro bioma vital, também foi amplamente afetada, com 2 milhões de hectares queimados. A

combinação de atividades humanas e condições climáticas desfavoráveis, como a seca, intensifica a ocorrência dessas queimadas e prejudica a biodiversidade.

De acordo com dados obtidos por meio do MapBiomass (2024), o aumento expressivo das queimadas não se limitou a um único bioma. No Pantanal, por exemplo, a área queimada entre janeiro e agosto de 2024 aumentou 249% em comparação com a média dos cinco anos anteriores, totalizando 1,22 milhão de hectares. Na Mata Atlântica, houve um crescimento de 683% na área queimada em agosto em relação à média histórica, com 500 mil hectares consumidos pelo fogo, dos quais 72% eram áreas agropecuárias. No estado de São Paulo, 86% da área queimada em 2024 ocorreu apenas em agosto, afetando predominantemente plantações de cana-de-açúcar. Esses dados apontam para uma tendência preocupante: a expansão das queimadas em áreas de vegetação nativa e em biomas frágeis, exacerbada pela ação humana e pelas condições climáticas adversas. A necessidade de medidas efetivas de prevenção e controle é urgente para mitigar os impactos ambientais, climáticos e sociais decorrentes das queimadas em grande escala (MapBiomass, 2024).

Ao redor do mundo, a situação com queimadas não é diferente, apesar de incêndios florestais no Canadá sejam relativamente comuns (BUMA et al., 2022) as mudanças climáticas estão levando a incêndios maiores e mais frequentes do que nas décadas passadas (HANES et al., 2019). Dados da indicam que o aquecimento global está contribuindo para um aumento de 30–50% na área global queimada até o final do século XXI (UNEP, 2022).

Neste trabalho, o foco recai sobre a detecção ativa de incêndios em imagens multiespectrais, por meio de métodos baseados em Aprendizado Profundo (AP) no campo de SR. Embora existam outras abordagens relevantes — como o uso de sensores térmicos, câmeras de infravermelho ou dados meteorológicos —, a escolha por imagens multiespectrais reflete a proposta metodológica adotada nesta pesquisa. Recentemente, métodos baseados em AP no campo de SR ganharam destaque devido ao aumento nas possibilidades de aplicação e ao bom desempenho apresentado. Métodos baseados em AP conseguem aprender automaticamente a partir de dados complexos em imagens de SR, utilizando diversas camadas em redes neurais. Em sua revisão, eles destacam um crescimento significativo, desde 2015, no número de publicações científicas que aplicam Aprendizado Profundo na análise de imagens de SR, acompanhado por um aumento expressivo nas citações desses trabalhos. Esse crescimento reforça o rápido interesse e evolução dos estudos de AP em imagens de SR Khelifi e Mignotte (2020a).

Este estudo tem como objetivo aprimorar técnicas de detecção de incêndios baseadas em inteligência artificial (IA), com ênfase no aumento da Precisão e na redução de Falsos Positivos em dados obtidos a partir de imagens multiespectrais do satélite Sentinel-2. Para isso, são empregadas redes neurais convolucionais profundas do tipo U-Net, combinadas a *encoders* pré-treinados — VGG-16 e EfficientNet —, explorando

o potencial do aprendizado transferido na extração de padrões espectrais e espaciais relevantes. Os objetivos específicos incluem a aplicação de técnicas de aumento de dados com a biblioteca *Albumentations* (como rotações, inversões, ajustes de brilho e contraste), o balanceamento do conjunto por *undersampling* seletivo e a otimização dos parâmetros de treinamento dos modelos, utilizando uma função de perda composta (*Dice Loss + Binary Focal Loss*). Essas estratégias visam aumentar a sensibilidade dos modelos à classe minoritária (fogo), garantindo maior robustez e generalização na detecção automática de incêndios.

1.1 Motivação

O progresso contínuo na associação de IA para detecção de incêndios reflete uma combinação de avanços tecnológicos e uma necessidade premente de inovação. As técnicas de AP, particularmente, têm mostrado grande potencial na interpretação de dados complexos de sensoriamento remoto, possibilitando a identificação de sinais precoces de incêndios florestais. No entanto, apesar dos avanços significativos, existem ainda lacunas e desafios a serem superados, especialmente no que se refere à precisão e à redução de falsos positivos, que são cruciais para a confiabilidade dos sistemas de detecção.

A otimização dos resultados obtidos em pesquisas anteriores não é apenas uma questão de progresso técnico, mas também um imperativo para a segurança pública e a gestão ambiental. A eficácia na detecção de incêndios tem um impacto direto na capacidade de resposta das equipes de emergência, na minimização de danos ao meio ambiente e na proteção da vida humana e da fauna. Portanto, o aprimoramento contínuo dos sistemas de detecção baseados em IA, por meio da pesquisa e desenvolvimento, é vital para enfrentar os desafios emergentes e dinâmicos associados aos incêndios florestais.

A vegetação possui importância vital para a humanidade, atuando como um recurso natural essencial que oferece benefícios diretos e indiretos, como a regulação climática, a manutenção da biodiversidade e a proteção do solo. No entanto, desastres naturais, especialmente os incêndios florestais, comprometem esses benefícios e contribuem significativamente para o aquecimento global. Por isso, a identificação automática e precoce de incêndios florestais é fundamental, pois permite que os responsáveis pela gestão de desastres planejem com antecedência estratégias de mitigação e táticas de combate ao fogo, minimizando os danos ambientais e socioeconômicos (SCHROEDER et al., 2016).

1.2 Objetivos e Desafios da Pesquisa

O objetivo principal desta pesquisa consiste em propor, implementar e avaliar um *pipeline* de segmentação semântica para a detecção de incêndios florestais em imagens multiespectrais do satélite Sentinel-2, com ênfase na mitigação dos efeitos do forte des-

balanceamento entre classes e na obtenção de modelos robustos para a classe de interesse (*incêndio*). Para isso, foram investigadas variantes da arquitetura U-Net com *encoders* pré-treinados e estratégias complementares de treinamento, balanceamento e calibração do limiar de decisão.

A pesquisa foi conduzida a partir dos seguintes objetivos específicos:

- ❑ avaliar o desempenho da U-Net com diferentes *backbones* (VGG-16 e EfficientNetB3) na segmentação da classe *incêndio*, considerando estabilidade do treinamento e desempenho no conjunto de teste;
- ❑ analisar o impacto da escolha da função de perda em cenários desbalanceados, comparando *Binary Cross-Entropy* e *Dice Loss* com abordagens mais sensíveis à classe minoritária, como *Focal Loss* e a perda composta *Dice+Focal*;
- ❑ investigar os efeitos do *undersampling* de amostras negativas e do *data augmentation* (Albumentations) como mecanismos para aumentar a exposição do modelo a padrões relevantes e reduzir vieses induzidos pelo desbalanceamento;
- ❑ avaliar uma estratégia de treinamento em duas fases, com congelamento inicial do *encoder* e posterior *fine-tuning*, visando maior estabilidade de convergência e melhor generalização;
- ❑ incorporar e avaliar uma etapa de calibração do *threshold* por varredura na validação, maximizando o *F1-Score*, e analisar seu efeito nas métricas finais e nas matrizes de confusão;
- ❑ comparar sistematicamente as combinações de arquitetura e escolhas metodológicas por meio de métricas orientadas à classe positiva (IoU, *F1-Score*, *precision* e *recall*), reportadas para validação e teste, complementadas por análise qualitativa das máscaras preditas.

Do ponto de vista dos desafios, destaca-se que a classe *incêndio* tende a ocupar regiões pequenas e esparsas nas imagens, o que torna a segmentação particularmente sensível ao desbalanceamento e à escolha do limiar de decisão. Além disso, a presença de pares imagem/máscara com inconsistências ou regiões ausentes pode introduzir ruído no processo de treinamento e avaliação. Assim, este estudo concentra-se na construção de um arranjo experimental capaz de isolar e quantificar o impacto das principais decisões metodológicas sobre o desempenho final dos modelos.

1.3 Hipótese

Considerando o forte desbalanceamento entre classes e a natureza esparsa da classe *incêndio* nas imagens Sentinel-2, este trabalho parte da hipótese central de que:

- a combinação de (i) funções de perda sensíveis à classe minoritária (*Dice*, *Focal* e a perda composta *Dice+Focal*), (ii) estratégias de mitigação do desbalanceamento no treinamento (e.g., *undersampling* de amostras negativas e *data augmentation*), e (iii) calibração do limiar de decisão por varredura na validação (maximizando o *F1-Score*) resulta em modelos de segmentação mais robustos e com melhor equilíbrio entre precisão e *recall*, refletido em valores superiores de IoU e *F1-Score* no conjunto de teste, quando comparados a abordagens tradicionais baseadas apenas em *Binary Cross-Entropy* e limiar fixo.

Adicionalmente, assume-se como hipótese complementar que o desempenho e a estabilidade do modelo dependem do *backbone* empregado, sendo esperado que arquiteturas com *encoders* mais estáveis apresentem menor sensibilidade à escolha da função de perda e do limiar de decisão, especialmente em cenários de desbalanceamento extremo.

1.4 Contribuições

Este trabalho apresenta uma abordagem reproduzível para segmentação semântica de áreas queimadas em imagens multiespectrais Sentinel-2, com foco no desafio de desbalanceamento extremo entre classes (fundo vs. incêndio). Considerando a metodologia adotada e os resultados experimentais obtidos, as principais contribuições são:

- Definição e implementação de um *pipeline* de preparação de dados para Sentinel-2, contemplando organização do conjunto multiespectral, padronização do particionamento (*train/validation/test*) e geração de amostras adequadas ao treinamento de redes convolucionais para segmentação;
- Avaliação comparativa de arquiteturas U-Net com *encoders* pré-treinados, incluindo VGG-16 e EfficientNetB3, analisando o impacto do *backbone* na qualidade de segmentação sob o mesmo protocolo experimental e conjunto de métricas;
- Proposição e validação de uma estratégia explícita para lidar com o desbalanceamento de classes por meio de *undersampling* orientado ao conteúdo (priorizando amostras com presença da classe minoritária) e *data augmentation* para ampliar a variabilidade das amostras, reduzindo viés do modelo e melhorando a capacidade de generalização;
- Investigação sistemática de funções de perda voltadas à classe minoritária, comparando perdas clássicas (p.ex., *Binary Cross-Entropy* e *Dice Loss*) com uma função de perda composta (combinação de termos para sobrepor desbalanceamento e refinar bordas), evidenciando ganhos de desempenho (especialmente em *F1-score* e IoU) na segmentação da classe de incêndio;

- ❑ Demonstração da eficácia do treinamento em duas fases (*warm-up* do *decoder* com congelamento do *encoder*, seguido de *fine-tuning* do *encoder*), contribuindo para estabilização do aprendizado, melhor convergência e incremento de desempenho nos conjuntos de validação e teste;
- ❑ Disponibilização do código-fonte, configurações experimentais e artefatos de avaliação (métricas, visualizações qualitativas e resultados por experimento) em repositório público, promovendo reprodutibilidade e facilitando o reuso por pesquisadores e profissionais de monitoramento ambiental.

1.5 Organização da Dissertação

Esta dissertação está organizada em quatro capítulos, além desta introdução.

O Capítulo 2 apresenta a fundamentação teórica que sustenta o trabalho, contemplando conceitos de segmentação semântica e redes neurais convolucionais, a arquitetura U-Net e *encoders* pré-treinados, bem como aspectos do sensoriamento remoto relevantes ao estudo, incluindo características do satélite Sentinel-2 e discussões sobre composições espectrais empregadas na detecção de queimadas. Também são revisados trabalhos correlatos e abordagens recentes aplicadas à segmentação em imagens orbitais.

O Capítulo 3 descreve o desenho experimental adotado, incluindo o *dataset* utilizado, o pré-processamento das imagens e a construção do *pipeline* de treinamento e avaliação. São detalhadas as variações de arquitetura (U-Net com VGG-16 e EfficientNetB3), as estratégias de mitigação do desbalanceamento (como *undersampling* e *data augmentation*), as funções de perda avaliadas (incluindo perdas compostas) e o procedimento de calibração do limiar de decisão por varredura na validação.

O Capítulo 4 apresenta e analisa os resultados obtidos nos dois ciclos experimentais, comparando as combinações de *backbones* e funções de perda com base em métricas orientadas à classe *incêndio*, como IoU, *F1-Score*, precisão e *recall*. Além da avaliação quantitativa, são discutidos o comportamento das curvas de aprendizado, as matrizes de confusão e a análise qualitativa das máscaras preditas, destacando os efeitos das melhorias introduzidas no Ciclo 2.

Por fim, o Capítulo 5 sintetiza os principais achados do estudo, revisita a hipótese de pesquisa à luz dos resultados e apresenta limitações e direções para trabalhos futuros, incluindo a exploração de arquiteturas baseadas em atenção, estratégias alternativas de balanceamento, avaliação em múltiplos sensores e generalização para diferentes biomas e configurações espectrais.

Fundamentação Teórica

Nesta seção serão abordados os conceitos fundamentais que embasam o desenvolvimento desta pesquisa e os principais trabalhos correlatos.

2.1 Métodos de Aprendizado Profundo em Imagens de Sensoriamento Remoto

AP ou *Deep Learning (DL)* é uma subdisciplina do *Machine Learning (ML)* focada no desenvolvimento de modelos capazes de aprender a partir de dados fornecidos, com o objetivo de produzir os resultados esperados. O diferencial dessas técnicas está na utilização de diversas camadas de processamento sobrepostas, o que possibilita a construção de representações hierárquicas dos dados. Cada camada subsequente refina as informações processadas pela anterior, permitindo a extração de características progressivamente mais complexas e abstratas.

No contexto do reconhecimento de imagens, por exemplo, o modelo inicia a análise a partir de informações básicas, como os píxeis. Em camadas mais profundas, essas informações são combinadas para identificar padrões simples, como linhas e contornos, até chegar a uma interpretação mais detalhada e complexa da imagem como um todo.

Nos últimos anos, o uso de técnicas de DL no campo do SR tem atraído atenção devido ao seu desempenho robusto e à ampliação das possibilidades de aplicação (KHELIFI; MIGNOTTE, 2020b). Essas técnicas são capazes de aprender automaticamente representações complexas de dados em imagens de SR por meio de redes neurais compostas por diversas camadas. Segundo esses autores, desde 2015, observa-se um crescimento expressivo no número de publicações científicas relacionadas ao uso de DL para análise de imagens de SR, acompanhado por um aumento significativo nas citações dessas publicações, indicando o crescente interesse da comunidade científica.

Hoeser e Kuenzer (2020) também notam que, desde que o DL começou a ser aplicado no SR por volta de 2015, o número de artigos publicados nessa área mais que dobrou a

cada ano. Esse aumento é impulsionado pela maior disponibilidade de dados e pelo avanço no poder computacional, especialmente com o uso de *Graphics Processing Units* (GPUs), que facilitam o processamento em grande escala. As aplicações incluem diversas áreas, como super-resolução de imagens, fusão de dados, remoção de ruído, previsão climática e detecção de objetos utilizando sensores ópticos, multiespectrais, hiperspectrais e Radar de Abertura Sintética (SAR).

O crescimento na quantidade de dados provenientes de diferentes sensores e fontes heterogêneas tem possibilitado avanços no estudo de processos ambientais. Para esses autores, as técnicas de DL são especialmente adequadas para lidar com os desafios relacionados à fusão de dados de múltiplos sensores no SR.

2.2 Redes Neurais Convolucionais

O ML é um subconjunto da IA que envolve o treinamento de máquinas em grandes conjuntos de dados para aprender padrões e fazer previsões. O AP, que envolve o treinamento de redes neurais com camadas, é uma subclasse poderosa do ML que demonstrou notável sucesso em várias aplicações, como reconhecimento de imagens e fala.

As Redes Neurais Convolucionais (CNN) são um tipo popular de modelo de ML utilizado para reconhecimento de imagens, reconhecimento de fala e *Natural Language Processing* (NLP). As Redes Neurais Convolucionais (CNNs) foram bem-sucedidas em tarefas de reconhecimento de imagens, alcançando resultados de ponta em vários *benchmarks*. Seu sucesso se deve à capacidade de capturar características e padrões espaciais em imagens, utilizando uma arquitetura hierárquica de camadas que realizam operações de convolução e extraem características em diferentes níveis de abstração (XIONG et al., 2021). A Figura 1 ilustra o fluxo.

A relação hierárquica entre os principais campos da inteligência computacional. A IA constitui o domínio mais amplo, englobando métodos e sistemas capazes de simular aspectos do raciocínio humano. Dentro dela, o ML representa um subconjunto voltado ao desenvolvimento de algoritmos que aprendem a partir de dados. O AP, por sua vez, é uma vertente do aprendizado de máquina que utiliza redes neurais profundas para modelar representações complexas. As CNN integram essa categoria, destacando-se como uma das arquiteturas mais utilizadas em tarefas de reconhecimento de padrões e análise de imagens.

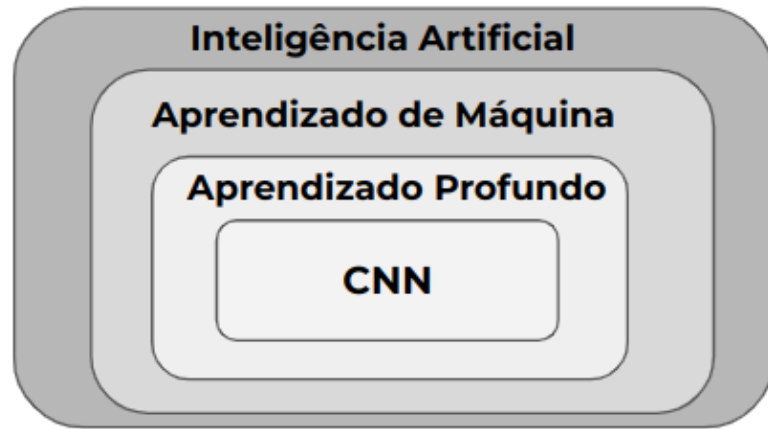


Figura 1 – Relação hierárquica entre IA, Aprendizado de Máquina, Aprendizado Profundo e CNN. Fonte: Elaborado pela autora.

A U-Net é uma arquitetura de CNN projetada especificamente para tarefas de segmentação de imagens biomédicas, mas que também tem sido aplicada em outras áreas de segmentação de imagem. Caracteriza-se por sua forma em “U”, composta por duas partes principais: um caminho de contração (*encoder*) e um caminho de expansão (*decoder*). O caminho de contração segue a arquitetura típica de uma rede convolucional, na qual as imagens são progressivamente reduzidas em tamanho através de camadas convolucionais e de *pooling*, capturando características contextuais. O caminho de expansão consiste em camadas de *upsampling* e convoluções que aumentam a resolução das características extraídas, permitindo a recuperação precisa da localização espacial. Além disso, a U-Net usa conexões de atalho (*skip connections*) que transferem mapas de características de alta resolução do *encoder* para o *decoder*, preservando informações detalhadas essenciais para uma segmentação precisa. Esta combinação de características permite que a U-Net obtenha resultados de segmentação com assertividade notável, mesmo com uma quantidade limitada de dados de treinamento Ronneberger, Fischer e Brox (2015).

No estudo de Pereira (2021), foi utilizada uma U-Net, essa arquitetura foi aplicada nas imagens do Landsat-8, alcançando um desempenho destacado, com métricas de *F-Score* acima de 90%. Quando aplicados em imagens do Sentinel-2, esses modelos mantiveram uma performance robusta, com *F-Score* ultrapassando os 80%.

O uso de CNNs já é bem estabelecido para tarefas de visão computacional devido à sua capacidade de extrair automaticamente características hierárquicas complexas de imagens. No contexto da detecção de incêndios, as arquiteturas VGG-16, InceptionV3 e Xception foram adaptadas, demonstrando a versatilidade e eficácia das CNNs em reconhecer padrões de fogo e fumaça em imagens (SATHISHKUMAR et al., 2023). A escolha dessas arquiteturas específicas se deve às suas características únicas, como a profundidade

das redes e a eficiência no tratamento de parâmetros, que são cruciais para o desempenho na classificação de imagens complexas relacionadas a incêndios.

2.3 Aumento de Dados

O aumento de dados pode ser utilizado para melhorar o desempenho de modelos de AP, especialmente em cenários na qual há uma quantidade limitada de dados rotulados disponíveis para treinamento. Essa técnica consiste na aplicação de transformações nos dados de entrada, como rotações, espelhamentos, distorções e ajustes de cor, com o objetivo de expandir o conjunto de dados de treinamento, aumentando sua variabilidade e, consequentemente, a capacidade do modelo de generalizar para novos dados.

Entre as bibliotecas disponíveis, a *Albumentations* destaca-se por sua eficiência e flexibilidade na aplicação de transformações aleatórias em imagens, permitindo ampliar a diversidade dos dados de entrada sem comprometer sua coerência semântica. Essa biblioteca oferece um conjunto abrangente de operações, como rotações, espelhamentos, translações, variações de brilho e contraste, distorções geométricas e ajustes de cor, que podem ser combinadas dinamicamente durante o treinamento. Dessa forma, o uso da *Albumentations* contribui para reduzir o sobreajuste e melhorar a capacidade de generalização dos modelos, promovendo um desempenho mais robusto em tarefas de segmentação e classificação de imagens Buslaev et al. (2020).

2.4 Sentinel-2

Os satélites Sentinel-2 são instrumentos importantes para a observação da Terra, usados especialmente para monitoramento ambiental, incluindo a detecção de incêndios. Eles capturam imagens multiespectrais, que são imagens tiradas em várias faixas do espectro de luz, permitindo a análise de diferentes características da superfície terrestre.

O Copernicus Sentinel-2A e Sentinel-2B são satélites de alta resolução espacial projetados para fornecer dados ópticos multiespectrais. Eles foram lançados em uma órbita a 786 km de altitude, garantindo um tempo de revisita de 5 dias ao operar em conjunto. Esses satélites são equipados com um único instrumento de imageamento denominado *Multi-Spectral Instrument* (MSI), que cobre uma faixa de 295 km.

O MSI utiliza um Telescópio Anastigmático de Três Espelhos (TMA), tecnologia que permite capturar imagens em 13 diferentes bandas espectrais, desde o Visível e Infravermelho Próximo (VNIR) até o infravermelho de Ondas Curtas Infravermelho de Ondas Curtas (SWIR). Essas bandas funcionam como filtros de cor, ajudando a analisar vários aspectos da superfície terrestre e da atmosfera, com resoluções espaciais variando de 10 a 60 metros. Os nanômetros(nm) são usados para medir os comprimentos de onda da

luz, onde diferentes comprimentos de onda correspondem a diferentes cores ou faixas do espectro eletromagnético.

O espectro VNIR cobre comprimentos de onda de aproximadamente 400 a 1100 nanômetros. Este intervalo inclui tanto a luz visível quanto a infravermelha próxima, que são utilizadas na análise da vegetação, qualidade da água e outros componentes da superfície terrestre. O espectro SWIR abrange comprimentos de onda de aproximadamente 1100 a 3000 nanômetros. Esta faixa é utilizada para detectar umidade, neve, gelo e outras propriedades minerais. As bandas SWIR são menos afetadas por dispersão atmosférica e são especialmente úteis para diferenciar materiais na superfície terrestre que têm respostas espectrais distintas nesta faixa de comprimento de onda (OMER et al., 2020; ALPARONE; GARZELLI; ZOPPETTI, 2023).

As bandas espectrais são divididas em três grupos principais. O primeiro grupo consiste em 4 bandas com resolução de 10 metros, capturando detalhes finos da superfície terrestre: azul (490 nm); verde (560 nm); vermelho (665 nm); e infravermelho próximo (842 nm). O segundo grupo inclui 6 bandas com resolução de 20 metros, usadas para análises mais específicas: quatro bandas estreitas (705 nm, 740 nm, 783 nm e 865 nm) para estudar a vegetação e duas bandas SWIR (1610 nm e 2190 nm). O terceiro grupo contém 3 bandas com resolução de 60 metros, que capturam áreas maiores e são usadas principalmente para correções atmosféricas: 443 nm (aerossóis), usada para medir partículas no ar; 945 nm; e 1375 nm (cirrus). Com essas diferentes bandas espectrais, o MSI do Sentinel-2 pode observar e analisar diferentes aspectos da superfície terrestre e da atmosfera (CHAMBRÉLAN; TEAM, 2023).

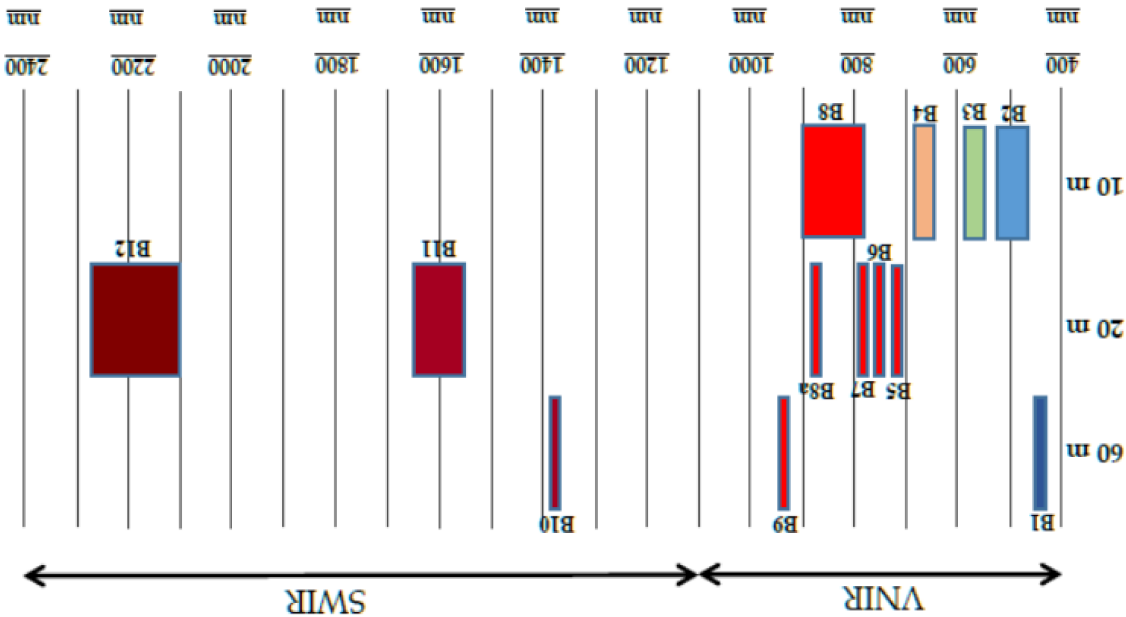


Figura 2 – Bandas espectrais e resoluções do MSI do Sentinel-2. Fonte: Elaborado por Chambrélan e team (2023).

A Figura 2 ilustra as diversas bandas espectrais do MSI do Sentinel-2, destacando como cada grupo de bandas é utilizado para diferentes tipos de análise ambiental. Essas bandas permitem uma observação detalhada de fenômenos naturais, facilitando a detecção de mudanças sutis na superfície terrestre e na atmosfera.

Já a Figura 3 mostra um exemplo prático da aplicação dessas bandas para a detecção de incêndios. As combinações específicas de bandas, como as de infravermelho próximo e SWIR, são cruciais para identificar e monitorar a progressão dos incêndios, permitindo uma resposta mais rápida e eficiente aos desastres naturais.

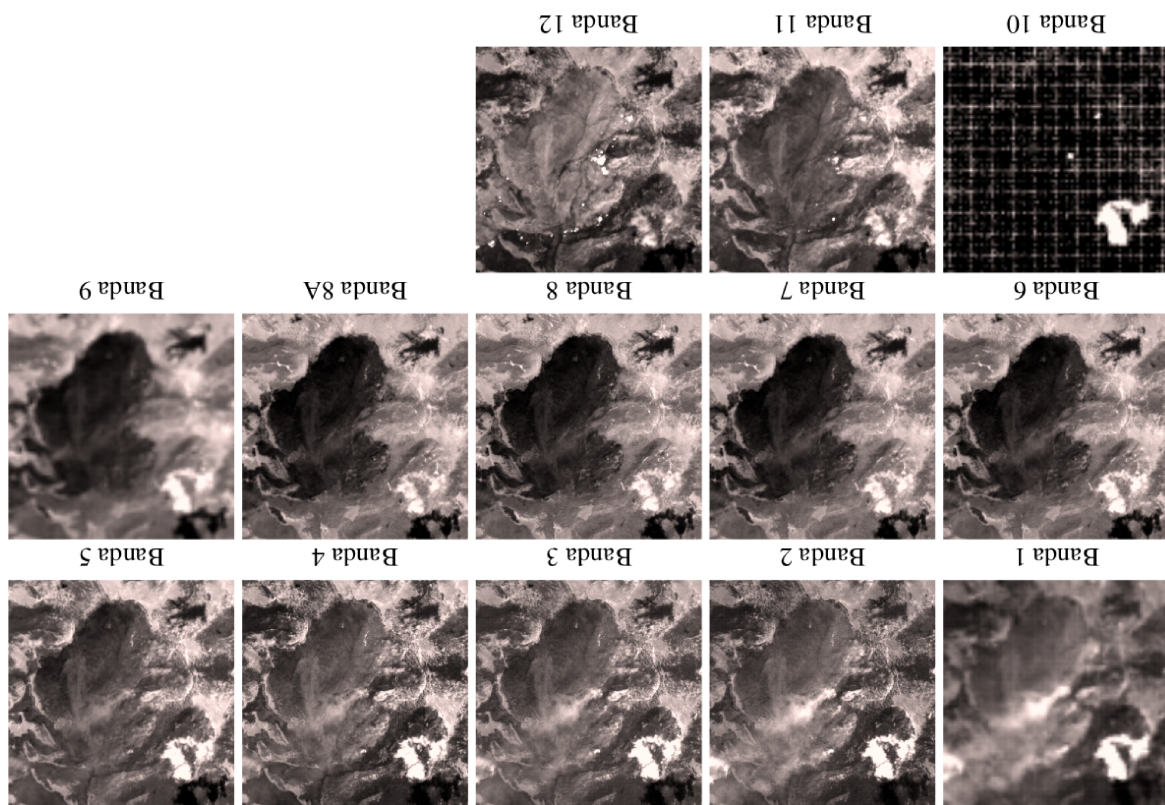


Figura 3 – Informação registrada em cada banda do satélite Sentinel-2 para um evento de incêndio. Fonte: Elaborado por Pereira (2021).

A Figura 4 apresenta diferentes combinações de bandas do satélite Sentinel-2 utilizadas para realçar informações associadas a eventos de incêndio. Por fim, a Figura 4(c) apresenta uma combinação com bandas do infravermelho e termal (1, 9 e 10), que permite identificar variações de temperatura e a extensão das áreas afetadas. Essas diferentes combinações são fundamentais para o monitoramento de incêndios florestais, cada uma oferecendo vantagens específicas em termos de realce e interpretação dos fenômenos.

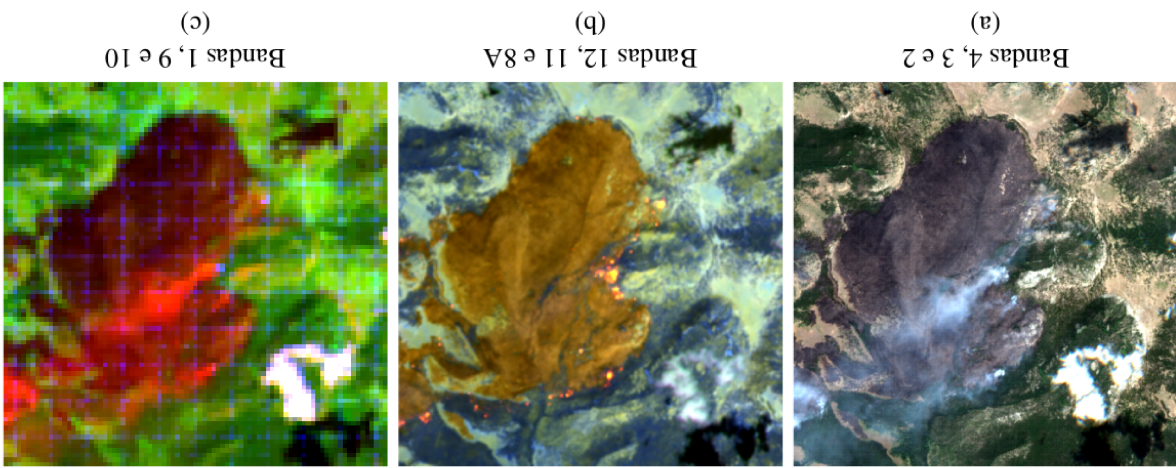


Figura 4 – Combinações de bandas do satélite Sentinel-2 para um evento de incêndio: (a) bandas 4, 3 e 2 (espectro visível); (b) bandas 12, 11 e 8A (infravermelho de ondas curtas e infravermelho próximo); (c) bandas 1, 9 e 10 (infravermelho e termal). Fonte: Elaborado por Pereira (2021).

2.5 Trabalhos Correlatos

Tecnologias de sensoriamento remoto e inteligência artificial possuem potencial para apoiar o planejamento de conservação, abrangendo a identificação de habitats críticos, a avaliação de riscos ecossistêmicos e a análise de paisagens para monitoramento em larga escala Christin, Hervet e Lecomte (2019). Em 2012, ocorreu uma revolução na classificação de imagens com a introdução da AlexNet, uma arquitetura de rede neural convolucional que representou um marco no uso de aprendizado profundo (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). Desde então, o aprendizado profundo tem sido aplicado no sensoriamento remoto para aprimorar a classificação automática de características em imagens de satélite. Redes convolucionais profundas supervisionadas aprendem a identificar objetos em imagens utilizando dados brutos, com mínima necessidade de conhecimento prévio sobre as características em análise. Na segmentação semântica, por exemplo, o modelo requer apenas uma máscara com rótulos para reconhecimento e treinamento eficazes (WAGNER et al., 2019).

Pesquisas recentes têm explorado diferentes abordagens para aprimorar a eficácia das CNNs no domínio de SR. Estudos têm demonstrado que a integração de dados multiespectrais e hiperespectrais com CNNs pode melhorar de maneira considerável a detecção de incêndios em diversas condições ambientais. Além disso, técnicas como transferência de aprendizado e aumento de dados têm sido utilizadas para superar desafios relacionados à escassez de dados rotulados, comuns nesse tipo de aplicação.

A pesquisa no campo da detecção de incêndios ativos concentra-se em fornecer métodos apropriados para determinar se um determinado pixel em uma imagem multiespectral corresponde a um incêndio ativo. Devido às diferentes altitudes orbitais dos satélites e aos diversos sensores equipados com diferentes comprimentos de onda, esses métodos geralmente precisam ser projetados especificamente para cada sistema. As imagens do satélite Landsat-8 foram processadas por meio de condições propostas por Schroeder et al. (2014), Murphy et al. (2016) e Kumar e Roy (2018), ajustadas especialmente para o sensor Operational Land Imager (OLI) sobre métodos automáticos para detecção de regiões com incêndios ativos datam da década de 1970, quando foram analisados os intervalos espectrais mais adequados para detectar incêndios florestais (CHEN; MORTON; RANDERSON, 2024). Matson e Holben (1987) argumentaram que o sensoriamento remoto era a única alternativa viável para monitorar incêndios ativos em regiões isoladas, utilizando o sensor Advanced Very High Resolution Radiometer (AVHRR) para detectar atividade de fogo na Amazônia.

A classificação de imagens de SR continua sendo uma tarefa complexa e multifacetada, enfrentando desafios devido ao grande volume de dados, a diversidade das características da paisagem e a presença frequente de conjuntos de treinamento limitados e desbalanceados (BAN et al., 2020). Além disso, a eficácia da classificação e o custo computacional associado são afetados por uma variedade de fatores, incluindo a escolha dos algoritmos

de classificação, o tipo de sensores de SR utilizados, técnicas de pré-processamento e pós-processamento, a natureza das classes alvo e o nível de desempenho desejado para o produto final, e a seleção e qualidade das amostras de treinamento.

(de Almeida Pereira et al., 2021) introduziram um novo dataset público contendo mais de 150.000 patches de imagem, extraídos de imagens multiespectrais do Landsat-8 capturadas em agosto de 2020, com incêndios ativos em diversas localidades ao redor do mundo. Este dataset foi dividido em duas partes: a primeira com saídas geradas por três algoritmos manuais conhecidos para detecção de incêndios, e a segunda contendo máscaras de fogo anotadas manualmente. O estudo demonstrou que redes neurais convolucionais podem ser usadas para aproximar esses algoritmos manuais, alcançando uma precisão de 87,2% e um *recall* de 92,4% em comparação com as anotações manuais, superando os algoritmos manuais em termos de robustez na detecção de incêndios ativos.

Os métodos utilizados no estudo incluem o treinamento de CNNs com diferentes arquiteturas derivadas da *U-Net* para segmentação de imagens, explorando a combinação de múltiplos *outputs* de detecção para melhorar os resultados. Os modelos foram treinados para replicar os algoritmos manuais de (SCHROEDER et al., 2014), (MURPHY et al., 2016) e (KUMAR; ROY, 2018), que utilizam condições específicas baseadas em bandas espectrais para detectar incêndios. As CNNs foram capazes de capturar as complexidades dessas o a votação entre modelos, que mostrou um equilíbrio superior entre precisão e *recall*.

(PEREIRA et al., 2023) utilizou a arquitetura de redes neurais convolucionais U-Net para segmentação e mapeamento de ecossistemas de canga. Nesse caso, a U-Net foi adaptada para a segmentação de superfícies naturais utilizando imagens do Sentinel-2, ajustando a arquitetura para lidar com a menor resolução espacial dessas imagens. A abordagem combinou máscaras vetoriais derivadas de imagens com baixa cobertura de nuvens e dados de campo, assegurando maior representatividade das classes segmentadas.

A metodologia empregou 2.061 tiles, sendo 927 representando o objeto de interesse (e.g., vegetação) e 1.134 contendo apenas fundo. *Tiles* consistem em subdivisões ou recortes de uma cena de sensoriamento remoto em blocos menores, permitindo que o treinamento de redes neurais seja realizado de forma mais eficiente, com menor demanda de memória e maior controle sobre a variabilidade espacial presente no conjunto de dados.

Para mitigar o *overfitting* e aumentar a robustez do modelo, foram aplicadas técnicas de aumento de dados, como rotações, ajustes de brilho, saturação e matiz, além da conversão para o espaço de cor *Hue, Saturation, Brightness* (HSB). Esse modelo de cor armazena informações em três componentes: matiz (responsável pela variação de cor), saturação (intensidade ou pureza da cor) e brilho (luminosidade do pixel). A possibilidade de manipular cada componente separadamente permite gerar variações controladas e realistas nas imagens, ampliando a diversidade do conjunto de treinamento e tornando o modelo mais capaz de generalizar para diferentes condições de iluminação e coloração.

Em uma área de 30.000 ha no Vale do Rio Peixe Bravo, o algoritmo de aprendizado profundo identificou e segmentou 762 manchas de canga com uma precisão geral de 98,5%, evidenciando o potencial preditivo das técnicas adotadas.

Enquanto (BOUGUETTAYA et al., 2022) concentram-se em dados coletados por Veículo Aéreo Não Tripulado (VANTs), que têm maior flexibilidade e podem voar em altitudes variadas, o trabalho aqui proposto considera imagens multiespectrais provenientes de satélites para detecção em larga escala. Os satélites, apesar de não oferecerem a flexibilidade de operação dos VANTs, são essenciais para o monitoramento contínuo e de grandes áreas.

(BOUGUETTAYA et al., 2022) ressaltam a importância de algoritmos avançados de CNN, mas enfrentam desafios relacionados à escassez de dados de treinamento de alta qualidade para otimizar esses modelos em detecção de incêndios. Em contraste, a presente pesquisa busca abordar diretamente essa lacuna por meio da aplicação de técnicas de aumento de dados, que visam ampliar artificialmente o conjunto de dados disponível para o treinamento de modelos de aprendizado profundo. Esta técnica é particularmente relevante no contexto de imagens de satélite, onde as variações climáticas e de cobertura do terreno podem limitar a quantidade de dados rotulados adequados. O aumento de dados tem o potencial de melhorar a generalização dos modelos e, assim, aumentar a acurácia e reduzir os falsos positivos, um desafio crítico na detecção de incêndios, como confirmado através do trabalho de (SANTOS et al., 2023).

Os (ROSTAMI et al., 2022), no campo da detecção de incêndios utilizando aprendizado profundo aplicou a técnica de aumento de dados denominada como inversão, especificamente a técnica de inversão horizontal, vertical e horizontal-vertical. Como resultado, as amostras de treinamento quadruplicaram para 400 patches de imagem.

O RCAug se diferencia de métodos tradicionais ao aplicar transformações de maneira aleatória e dinâmica durante o treinamento, gerando um número virtualmente ilimitado de novas imagens com alta variabilidade. Isso proporciona ao modelo a capacidade de lidar melhor com variações complexas de iluminação, escala e ângulo. Em estudos anteriores, essa técnica demonstrou melhorias significativas em métricas como o IoU e a F1-score, como evidenciado em Santos et al. (2023), que reportaram um aumento de 0,51 para 0,81 na métrica de IoU em imagens histológicas de tumores.

Por outro lado, técnicas tradicionais de aumento de dados, como o uso de *flips* horizontais e verticais, têm sido amplamente empregadas em diversos trabalhos. No estudo de (ROSTAMI et al., 2022), por exemplo, essas técnicas foram aplicadas ao treinamento do modelo MultiScale-Net para detecção de incêndios em imagens do Landsat-8. O aumento de dados foi utilizado principalmente para melhorar a capacidade do modelo de lidar com diferentes orientações das imagens, quadruplicando o conjunto de treinamento original. Embora eficaz em cenários com grande quantidade de dados e em problemas relativamente simples de detecção, essa abordagem apresenta limitações na capacidade

de generalizar para variações mais complexas, como mudanças dinâmicas de iluminação, forma ou escala dos incêndios.

A comparação entre essas abordagens revela que o RCAug oferece uma vantagem em termos de variabilidade e capacidade de generalização. Ao contrário das técnicas tradicionais, que introduzem uma variabilidade limitada nas imagens de treinamento, o RCAug pode expandir a diversidade do conjunto de dados. A aplicação contínua de múltiplas transformações durante o treinamento permite que o modelo lide com a complexidade inerente de cenários naturais, como incêndios de diferentes tamanhos, formas e condições atmosféricas.

Considerando os conhecimentos disponíveis (ANAYA; TRIPATHY; REINHARDT, 2023), o autor também afirma, com base nos seus estudos, que a escolha de um método específico de IA pode não ser tão crucial quanto a qualidade e o balanceamento dos dados utilizados no treinamento, isso porque a importância da proporção entre pixels de fogo e não fogo no dataset de treinamento, pode ser um fator determinante na eficiência das detecções.

(HE; GARCIA, 2009) afirma que é “é importante considerar como é feita a amostragem dos dados de treinamento, especialmente se houver necessidade de mapear classes raras” — como é o caso dos incêndios — assim como forma de analisar o comportamento de amostras balanceadas ou desbalanceadas para o treinamento. (ANAYA; TRIPATHY; REINHARDT, 2023) empregou duas abordagens distintas para o balanceamento de dados em sua análise. Primeiramente, adotou um formato balanceado, no qual a seleção de pixels foi realizada de maneira aleatória para assegurar uma proporção equitativa de 1:1 entre pixels de não fogo e fogo. Adicionalmente, explorou um formato desbalanceado, onde a proporção foi estabelecida em 1:100, significando que para cada pixel indicativo de incêndio, cem pixels representativos de ausência de fogo foram selecionados aleatoriamente, destacando uma metodologia intencionalmente inclinada para investigar os impactos de disparidades acentuadas na distribuição dos dados.

A importância de adotar uma abordagem diversificada na formação de amostras reside na natureza variável dos incêndios e na complexidade dos ambientes em que ocorrem. Incêndios podem variar em termos de intensidade, área afetada, e contexto ecológico, tornando a tarefa de detecção um desafio multidimensional. Além disso, condições atmosféricas, variações sazonais e a presença de elementos que possam ser erroneamente interpretados como fogo (como reflexos solares ou iluminação artificial) introduzem complexidade adicional na análise. Portanto, testar novas composições de amostras é vital para desenvolver modelos que sejam capazes de discernir entre verdadeiros incêndios e falsos positivos com maior acurácia, ajustando-se efetivamente à diversidade de dados encontrada em cenários reais de sensoriamento remoto (ANAYA; TRIPATHY; REINHARDT, 2023).

Já (SATHISHKUMAR et al., 2023) investigaram a detecção de incêndios florestais

e fumaça em imagens com o uso de modelos de aprendizagem profunda pré-treinados, dentre eles VGG-16, InceptionV3 e *Xception*. Para contornar o problema de esquecimento catastrófico, os autores empregaram a técnica *Learning without Forgetting* (LwF), de modo a manter o desempenho da rede em tarefas previamente aprendidas ao mesmo tempo em que inserem novos conjuntos de dados. O estudo obteve resultados promissores, destacando-se a arquitetura *Xception*, que alcançou 98,72% de acurácia na classificação do conjunto de dados original e manteve bons resultados para um novo conjunto de imagens quando retreinada com LwF. Esse trabalho evidencia o potencial do uso de *transfer learning* associado a estratégias que evitam o esquecimento de tarefas anteriores, mostrando-se relevante para aplicações de monitoramento e detecção de incêndios em cenários com dados limitados e em constante atualização.

O estudo conduzido por (LEE et al., 2024) propôs uma abordagem automatizada para detecção de áreas queimadas utilizando imagens do satélite Sentinel-2 na República da Coreia. Com foco em 14 casos de incêndios florestais ocorridos entre 2019 e 2023, o trabalho explorou diferentes tipos de cobertura do solo e comparou métodos de aprendizado profundo, aprendizado de máquina de índices espectrais. O objetivo principal foi identificar combinações otimizadas de modelos e dados de entrada para melhorar a precisão na detecção de áreas queimadas.

A metodologia consistiu em seis etapas principais. Primeiro, as imagens do Sentinel-2 foram pré-processadas, incluindo correção atmosférica por meio do *Sen2cor*, *resampleamento* de bandas SWIR para alinhamento de resolução espacial e equalização de histograma para melhorar o contraste das imagens. Em seguida, dados de referência foram gerados com base em inspeções manuais de imagens do Google Earth, rotulando áreas queimadas para uso no treinamento e validação dos modelos.

Os métodos de aprendizado de máquina e profundo incluíram os modelos U-Net e HRNet para DL e o framework AutoML. O U-Net utilizou uma arquitetura *encoder-decoder* com conexões *skip*, enquanto o HRNet combinou múltiplas escalas de resolução em sua estrutura, utilizando a arquitetura HRNetV2-W48 com mecanismo OCR para otimizar a segmentação. No aprendizado de máquina, o AutoML automatizou processos como pré-processamento, seleção de algoritmos e ajuste de hiperparâmetros, testando modelos como *Random Forest* e *Gradient Boosting*.

A otimização de hiperparâmetros foi uma etapa essencial no estudo. Nos modelos de DL, foram realizados testes iterativos para ajustar tamanho do lote, taxa de aprendizado e tamanho dos recortes das imagens. Valores ótimos incluíram tamanhos de lote de 8 para U-Net e 4 para HRNet, taxas de aprendizado de 5×10^{-6} e 5×10^{-3} , respectivamente, e recortes de 256×256 pixels. Para o AutoML, a busca em grade identificou configurações ideais para cada algoritmo testado.

Os índices espectrais NDVI e NBR também foram analisados utilizando o método de threshold automático (Otsu) para distinguir áreas queimadas de não queimadas. Essa

abordagem foi comparada ao desempenho dos modelos de DL e ML, considerando as combinações de bandas RGB e falsa cor (B4, B8, B3) como entradas.

Os resultados indicaram que o HRNet apresentou a maior precisão geral, com IoU de 89,40% em uma das áreas de teste, enquanto o U-Net demonstrou maior robustez em cenários com dados desbalanceados, apesar de apresentar falsos positivos em áreas agrícolas. O AutoML teve desempenho inferior ao DL, mas destacou-se pela automação do processo. A inclusão de índices espectrais melhorou ligeiramente os resultados em cenários específicos. Imagens em falsa cor foram identificadas como a melhor combinação de entrada, destacando sua eficácia na detecção de áreas queimadas.

Este estudo contribui significativamente ao explorar técnicas avançadas de aprendizado de máquina e profundo combinadas com análise espectral, aplicadas à detecção de incêndios florestais em ambientes complexos. A otimização detalhada dos hiperparâmetros garantiu que os modelos fossem ajustados às características específicas das imagens Sentinel-2 e dos terrenos analisados. Os resultados reforçam o potencial de sistemas automáticos para suporte em estratégias de mitigação e resposta a desastres ambientais.

Uma revisão sistemática conduzida por (MACHADO et al., 2023) analisou 37 estudos publicados entre 2018 e 2023, revelando resultados promissores na aplicação dessas técnicas. Os autores identificaram que a maioria das implementações alcançou taxas de precisão superiores a 90%, demonstrando a viabilidade e eficiência desses métodos para o monitoramento florestal.

A relevância desses sistemas automatizados torna-se ainda mais evidente quando consideramos que, entre 2002 e 2016, aproximadamente 420 milhões de hectares de floresta foram consumidos por incêndios anualmente em escala global. Este cenário tende a se agravar com as mudanças climáticas, que potencialmente aumentarão a frequência e intensidade dos incêndios florestais.

Os estudos analisados foram categorizados em cinco principais abordagens: *classificação pura*, *detecção*, *detecção combinada com classificação*, *segmentação* e *segmentação com classificação*. Entretanto, apesar dos resultados promissores, persistem desafios significativos na implementação desses sistemas, principalmente relacionados à escassez de dados de treinamento e ao desbalanceamento dos conjuntos de dados disponíveis.

Para superar essas limitações, a literatura sugere algumas direções promissoras para pesquisas futuras: a *otimização de hiperparâmetros* através de técnicas bio-inspiradas, a *integração de múltiplas fontes de dados* (incluindo imagens de satélite, VANTs e drones), a *hibridização de diferentes modelos de DL*, e a implementação de técnicas avançadas de aumento de dados, como *Neural Style Transfer* e Redes Adversariais Generativas (GANs). Além disso, a exploração de diferentes estratégias de particionamento de dados pode contribuir para melhorar o desempenho dos modelos.

A integração de múltiplas fontes de dados merece destaque particular, pois permite uma abordagem mais abrangente e robusta. Por exemplo, enquanto as imagens de satélite

podem fornecer uma visão ampla para identificação inicial de incêndios, os VANTs e drones podem complementar com dados detalhados em tempo real das áreas afetadas, facilitando análises mais precisas de danos e permitindo acesso a regiões de difícil alcance, com custos operacionais reduzidos em comparação com métodos convencionais. A tabela abaixo tras um resumo dos trabalhos explanados e comparativo com o atual trabalho:

Tabela 1 – Revisão de Estudos sobre Detecção de Incêndios com Deep Learning

Autor	Objetivo	Metodologia	Presente trabalho
Krizhevsky et al. (2017)	Introdução da AlexNet para classificação de imagens.	CNN com camadas convolucionais e <i>pooling</i> para extração de características.	Utilizo redes U-Net, VGG-16 e Xception para detecção de incêndios.
Schroeder et al. (2016), Murphy et al. (2016), Kumar et al. (2018)	Detecção de incêndios em imagens do satélite Landsat-8 , usando regras baseadas em bandas espectrais.	Métodos heurísticos aplicados às imagens Landsat-8, com ajuste para o sensor OLI.	No presente trabalho há a aplicação de CNN para aprender padrões automaticamente, sem necessidade de regras manuais.
Pereira et al. (2023)	Segmentação de ecossistemas de canga com U-Net e Sentinel-2.	U-Net treinada para segmentação de superfícies naturais, com ajuste para baixa resolução espacial.	Aplicação específica para incêndios , utilizando dados multiespectrais e otimização de hiperparâmetros .
Abdelmalek et al. (2022)	Detecção de incêndios com VANTs (drones) e CNNs.	Uso de imagens aéreas de drones, otimizadas para baixa altitude.	Utilizo imagens do satélite Sentinel-2 , que ainda são menos exploradas em comparação ao Landsat , para detecção de incêndios em larga escala .
Santos et al. (2023)	Aplicação de Data Augmentation (aumento de dados) para melhorar a generalização de CNNs.	Técnicas de aumento de dados, como inversão horizontal/vertical e RCAug, aplicadas a incêndios.	Implemento aumento de dados, mas com ajuste otimizado para imagens de satélite multiespectrais, e foco na redução de falsos positivos .
Rostami et al. (2022)	Uso de técnicas de Data Augmentation para melhorar a detecção de incêndios em imagens Landsat-8.	CNNs treinadas com aumento de dados para aumentar o número de amostras de fogo.	O presente trabalho inclui otimização de hiperparâmetros , além de imagens Sentinel-2.
Sathishkumar et al. (2023)	Comparação de arquiteturas CNNs (VGG-16, InceptionV3 e Xception) para detecção de incêndios e fumaça.	Aplicação de LwF .	No trabalho em questão, a Técnica de LwF , devido alto custo envolvido no processo de detecção.

Fonte: Elaborado pela autora.

Metodologia

Neste trabalho, foi aplicada a Segmentação Semântica para detecção de focos de incêndios florestais em imagens multiespectrais do satélite Sentinel-2, utilizando composições em falsa cor (*False Color*) para realçar contrastes relevantes à tarefa. A segmentação foi formulada como um problema binário, com os *pixels* classificados como “incêndio” ou “não-incêndio”.

O fluxo metodológico desenvolvido compreendeu dois ciclos experimentais distintos, conforme ilustrado no fluxograma a seguir. No primeiro ciclo, de caráter exploratório, avaliou-se o desempenho da arquitetura U-Net com dois *encoders* diferentes: VGG-16 e EfficientNetB3. Nesse estágio inicial, não foram aplicadas técnicas de balanceamento de classes ou aumento de dados (*data augmentation*), e os modelos foram treinados com as funções de perda Entropia Cruzada Binária (BCE) e *Dice Loss*. Essa etapa foi essencial para identificar as limitações enfrentadas na detecção da classe minoritária (incêndio), refletidas principalmente nos baixos valores de IoU e *F1-Score*.

Com base nas deficiências observadas, o segundo ciclo experimental incorporou um pipeline otimizado. A arquitetura Xception foi substituída pela EfficientNetB3 como *encoder* da U-Net, e foram integradas múltiplas melhorias metodológicas: aplicação de *undersampling* para reduzir amostras sem ocorrência de incêndio, aumento de dados com a biblioteca *Albumentations* (incluindo *flip*, rotação, brilho e contraste), além do uso de uma função de perda composta (*Dice Loss* + *Binary Focal Loss*). O treinamento também passou a ser conduzido em duas fases — inicialmente com o *encoder* congelado, seguido de *fine-tuning* com taxa de aprendizado reduzida —, promovendo um aprendizado mais estável e eficiente. Para facilitar a reprodutibilidade e a comparação entre os ciclos, recomenda-se que a tabela comparativa entre Ciclo 1 e Ciclo 2 (com *backbones*, perdas, balanceamento, aumento de dados e estratégia de *threshold*) seja apresentada nesta seção de Metodologia (no tópico de desenho experimental), e não apenas no Capítulo de Resultados.

A avaliação foi conduzida com foco nas métricas mais sensíveis ao desbalanceamento, como IoU, F1-Score, precisão e revocação (*recall*), todas aplicadas exclusivamente à classe

“incêndio”. Adicionalmente, realizou-se uma análise qualitativa das predições por meio da inspeção visual das máscaras geradas, permitindo avaliar não apenas a exatidão, mas também a coerência morfológica das regiões segmentadas.

O *threshold* de decisão empregado nas avaliações foi definido por varredura de *thresholds* maximizando o F1-Score na validação, fixando-se em 0,45 para BCE e 0,40 para *Dice Loss*; os mesmos valores foram utilizados para computar as matrizes de confusão e as métricas resumidas nas tabelas apresentadas no Capítulo de Resultados.

Resgatando o objetivo central deste trabalho — aplicar e avaliar modelos de segmentação semântica para a detecção de incêndios florestais em imagens multiespectrais do satélite Sentinel-2 — foram conduzidos dois ciclos experimentais principais.

O primeiro ciclo experimental teve caráter exploratório e avaliou o desempenho da U-Net com *encoders* VGG16 e Xception, utilizando, primeiramente, a função de perda BCE e, posteriormente, a *Dice Loss*. Esse ciclo permitiu evidenciar as dificuldades enfrentadas na segmentação da classe “incêndio” nas imagens, com métricas de IoU e F1-Score bastante reduzidas. Mesmo com a aplicação da *Dice Loss*, os ganhos foram modestos, revelando a necessidade de ajustes metodológicos diante do desbalanceamento observado.

O segundo ciclo experimental foi concebido a partir das limitações observadas no Ciclo 1, mantendo a mesma arquitetura de referência (U-Net com codificadores pré-treinados VGG16 e EfficientNetB3), porém incorporando melhorias metodológicas orientadas ao cenário de forte desbalanceamento entre as classes “fogo” e “não-fogo”. Assim, o objetivo do Ciclo 2 não é substituir a linha de base, mas refiná-la, permitindo uma comparação controlada entre um treinamento convencional e uma configuração mais adequada à segmentação de eventos raros em imagens Sentinel-2.

Em termos de manipulação de dados, o Ciclo 2 mantém os mesmos arquivos de divisão utilizados no Ciclo 1, garantindo comparabilidade completa entre os experimentos. A principal diferença consiste na ativação explícita de um esquema de *undersampling* sobre o conjunto de treino, no qual a classe majoritária (ausência de fogo) é reduzida para se igualar, em número de amostras, à classe minoritária (presença de fogo). Esse balanceamento visa atenuar o viés do modelo em favor da classe negativa e aumentar sua sensibilidade à detecção de áreas queimadas.

Outro aprimoramento diz respeito à estratégia de *data augmentation*. Enquanto o Ciclo 1 utilizou apenas redimensionamento ou transformações simples, o Ciclo 2 adota um pacote mais robusto, combinando operações geométricas (reflexões, translações e rotações leves) e fotométricas (variação de brilho, contraste, matiz e desfoque). Essas transformações visam simular variações de iluminação, condições atmosféricas e heterogeneidade espectral da vegetação, ampliando a capacidade de generalização dos modelos.

Além das melhorias de pré-processamento, o Ciclo 2 amplia o conjunto de funções de perda avaliadas. Além da BCE e da *Dice Loss*, são introduzidas a *Focal Loss* e a perda composta *Dice+Focal*, mais adequadas ao tratamento de classes raras e cenários

desbalanceados. O *sweep* do limiar de decisão também foi refinado, passando a considerar um conjunto mais denso de valores no intervalo $[0,10,0,90]$, permitindo avaliar com maior granularidade o compromisso entre *precision*, *recall* e IoU. Além disso, o Ciclo 2 investigou sistematicamente o efeito de quatro funções de perda: **BCE**, *Dice*, *Focal* e a combinação *Dice Focal*. Em conjunto, esse arranjo experimental permitiu analisar o impacto isolado e combinado das estratégias de balanceamento e das funções de perda.

Entre todas as combinações testadas, o modelo VGG-16 com *Focal Loss* obteve o melhor desempenho global no Ciclo 2, alcançando valores de *F1-score* superiores a 0,96 tanto em validação quanto em teste. Esse comportamento superou significativamente o observado no Ciclo 1, em que a ausência de mecanismos explícitos de balanceamento resultava em oscilações acentuadas de *precision* e dificuldade de segmentar regiões menores ou esparsas.

Ambos os experimentos foram gerados com base no conjunto integral de imagens do FireDataset_20m — contendo 5.167 imagens para treinamento, 1.226 para validação e 1.858 para teste. Além disso, foram realizados em ambiente computacional com processador AMD Ryzen 5 2500U, 24 GB de memória DDR4 e GPU AMD Radeon 535 com 2 GB de VRAM, sob sistema operacional Windows 10. O desenvolvimento foi conduzido em *Python* 3.9, com suporte das bibliotecas *TensorFlow*, *NumPy*, *Matplotlib*, *Pillow*, *scikit-image* e *tqdm*.

O treinamento foi executado com o otimizador Adam, utilizando taxa de aprendizado inicial de 1×10^{-4} e *batches* de quatro amostras. A fase de *fine-tuning* adotou uma taxa reduzida (1×10^{-5}) para permitir ajustes graduais nos pesos do *encoder*. Mecanismos de *Early Stopping* e *Model Checkpoint* foram empregados para prevenir sobreajuste e armazenar os melhores pesos com base na métrica de IoU na validação.

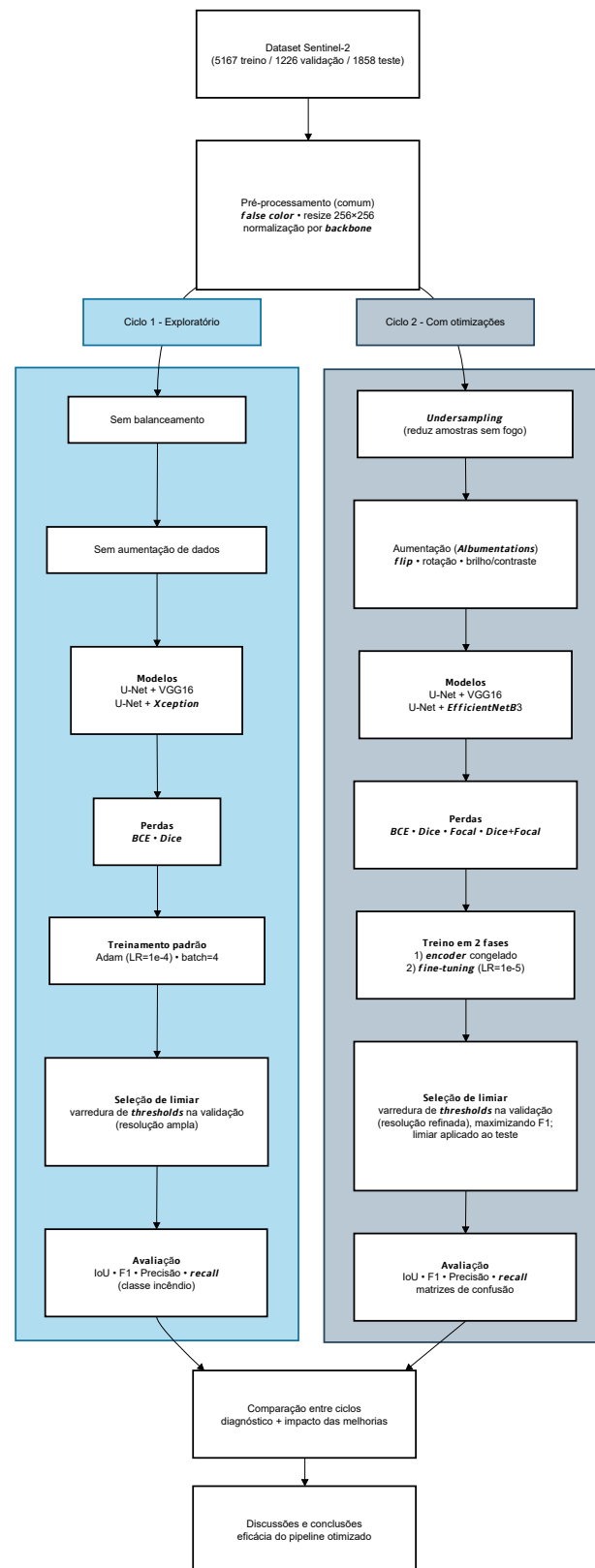


Figura 5 – Fluxograma da metodologia proposta para a segmentação de incêndios florestais.

3.1 Aquisição de Dados

O primeiro passo deste estudo envolveu a aquisição das imagens necessárias para o treinamento e avaliação dos modelos de segmentação. As imagens utilizadas foram extraídas do *FireDataset_20m*, disponível no site do projeto CIAR ¹. O projeto Chaîne Image Autonome et Réactive (CIAR), desenvolvido pelo IRT Saint Exupéry, foca no desenvolvimento de arquiteturas de Inteligência Artificial voltadas ao processamento embarcado em sistemas aeroespaciais, como satélites e drones. A iniciativa busca otimizar a autonomia desses sistemas por meio da análise de imagens em tempo real, permitindo uma tomada de decisão ágil diretamente no hardware. Para isso, o projeto integra algoritmos de redes neurais a dispositivos de baixo consumo energético, superando os desafios técnicos da transmissão de dados e da latência em missões críticas.

Os dados consistem em imagens de satélite com resolução espacial de 20 metros, capturadas pelo satélite Sentinel-2. Para cada imagem, foram fornecidas pelo Instituto Francês de Pesquisa Tecnológica, máscaras de incêndio, marcando as áreas de interesse que contêm incêndios. As imagens foram coletadas utilizando o formato *False Color* (B12, B11, B04), uma combinação de bandas espectrais que melhor capturam as características relevantes para a detecção de incêndios. Como exemplo, a Figura 6 mostra uma das imagens de satélite utilizadas no estudo, onde as áreas de incêndio são destacadas pela combinação das bandas B12, B11 e B04. A imagem é apresentada no formato de *False Color*, otimizada para destacar as regiões afetadas por incêndios.

As respectivas máscaras de incêndio foram fornecidas para cada imagem, indicando as áreas de interesse que contêm incêndios. Estas máscaras são imagens binárias onde o valor 1 representa as áreas queimadas (incêndio) e 0 representa as áreas não queimadas. As máscaras são fundamentais para o treinamento supervisionado, pois fornecem *ground truth* que o modelo tenta aprender a prever (Figura 6):

O primeiro passo deste estudo envolveu a aquisição das imagens necessárias para o treinamento e avaliação dos modelos de segmentação. As imagens utilizadas foram extraídas do site do projeto CIAR *FireDataset_20m*². Os dados consistem em imagens de satélite com resolução espacial de 20 metros, capturadas pelo satélite Sentinel-2.

¹ O conjunto de dados está disponível em: <https://ciar.irt-saintexupery.com/index.php/s/3RLvFSak6Qt7NnB?path=%2Fwildifre_data_20m#editor>

² <https://ciar.irt-saintexupery.com/index.php/s/3RLvFSak6Qt7NnB?path=%2Fwildifre_data_20m#editor>

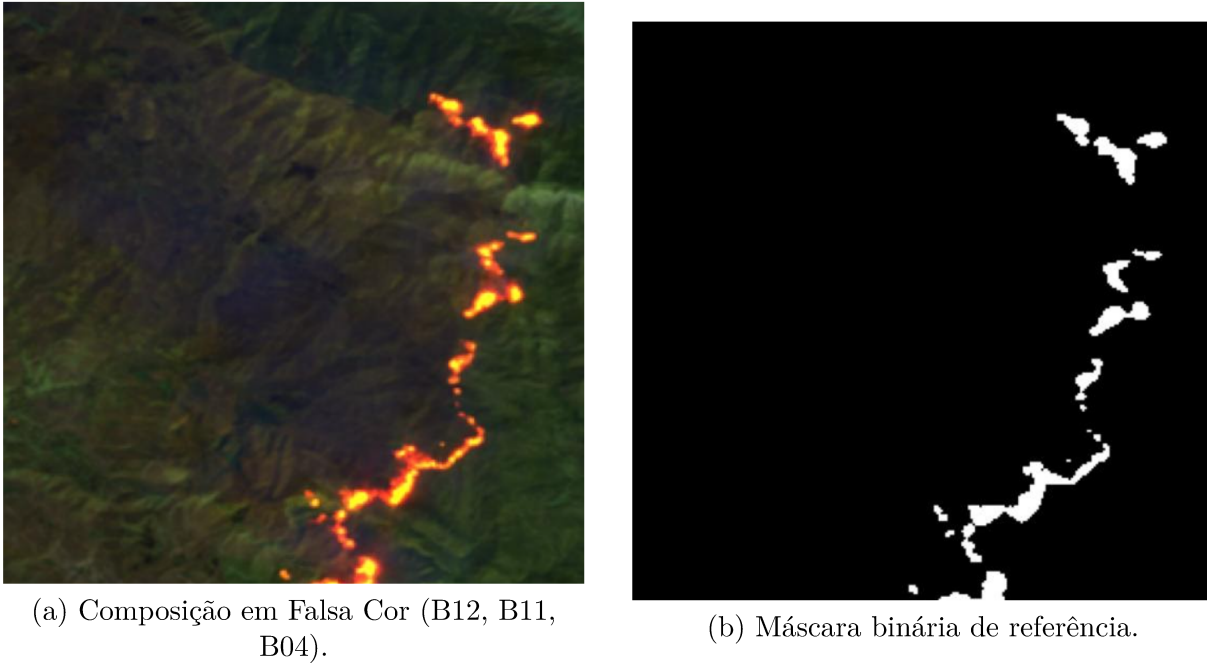


Figura 6 – Exemplo de uma imagem de satélite do Sentinel-2 no formato *False Color* e sua respectiva máscara binária, indicando as áreas afetadas pelo incêndio.

Para garantir a consistência nas entradas dos modelos de segmentação, as imagens e máscaras foram redimensionadas para uma dimensão fixa de 256×256 *pixels*, o que também ajudou a reduzir a complexidade computacional do processo. Além disso, o conjunto de dados foi dividido em lotes de 4 imagens, com as imagens sendo agrupadas para otimizar o treinamento.

3.2 Segmentação

O processo de segmentação foi realizado utilizando as seguintes arquiteturas de modelos, U-Net com *encoder* VGG-16 e EfficientNetB3, cada uma com suas características específicas para lidar com as imagens de satélite.

3.2.1 U-Net (Arquitetura CNN *Encoder-Decoder*)

A U-Net é uma arquitetura de rede neural convolucional especializada em tarefas de segmentação de imagens. A principal característica da U-Net é sua estrutura em forma de “U”, composta por um caminho de contração, que comprime as informações, e um caminho de expansão, que as reconstrói. Essa arquitetura permite uma segmentação precisa das regiões afetadas pelo fogo, preservando a resolução espacial e detalhando as áreas de interesse nas imagens. A U-Net tem sido amplamente utilizada em diversas áreas, desde a segmentação biomédica até o sensoriamento remoto.

A U-Net é particularmente adequada para a segmentação de incêndios florestais, pois combina alta precisão espacial com a capacidade de identificar regiões específicas nas

imagens de satélite, que apresentam anomalias térmicas e sinais de fumaça. Suas skip connections (conexões de atalho) garantem que informações de alta resolução sejam mantidas, o que é fundamental para evitar a perda de detalhes, especialmente em áreas menores ou com pouca visibilidade. A U-Net foi utilizada com sucesso por Ronneberger, Fischer e Brox (2015) em várias tarefas de segmentação de imagens, e sua aplicação no sensoriamento remoto para a detecção de incêndios tem se mostrado promissora.

3.2.2 VGG16 com *Transfer Learning*

O modelo VGG-16 é uma arquitetura de rede neural convolucional caracterizada por sua simplicidade e profundidade. Composto por várias camadas convolucionais seguidas por camadas totalmente conectadas, o VGG-16 é amplamente utilizado para classificação de imagens, mas pode ser adaptado para tarefas de segmentação e detecção de padrões específicos, como incêndios florestais, por meio da utilização de camadas adicionais. Sua profundidade, combinada com a capacidade de aprender representações robustas, torna o **VGG1-6!** (**VGG1-6!**) uma escolha eficaz para detectar padrões complexos em imagens multiespectrais.

A simplicidade da arquitetura do VGG-16, aliada à sua eficácia em segmentação de imagens, facilita sua adaptação para tarefas de detecção de incêndios. A possibilidade de incorporar camadas de segmentação à rede permite a identificação precisa das regiões afetadas por incêndios, ao mesmo tempo em que minimiza o risco de falsos positivos, especialmente quando combinada com técnicas de aprendizado transferido. A arquitetura VGG16 foi explorada por Sathishkumar et al. (2023) para a identificação de padrões de fumaça e fogo em imagens multiespectrais, mostrando resultados consistentes na detecção de incêndios em diferentes tipos de terreno.

3.2.3 EfficientNetB3

Essa escolha foi motivada por seu design escalável, que equilibra profundidade, largura e resolução de forma eficiente, alcançando alto desempenho com menor custo computacional (TAN; LE, 2020).

Na literatura recente, a EfficientNet tem se mostrado eficaz em tarefas de segmentação semântica em imagens multiespectrais de satélite. Por exemplo, Soundararajan et al. Soundararajan et al. (2025) utilizaram a variante EfficientNet-B08 como encoder em uma arquitetura *DeepLabV3+*, alcançando valores de IoU superiores a 0,91 na detecção de áreas queimadas e desmatadas com imagens Sentinel-2.

Outros estudos exploraram variações da U-Net, como RAUNet e AUNet, com *encoders* modernos aplicados a dados do Sentinel-2, obtendo resultados robustos em cenários com forte desbalanceamento de classes (SHIRVANI; ABDI; GOODMAN, 2023).

3.2.4 Métricas

A avaliação dos modelos foi conduzida com o objetivo de quantificar o desempenho das previsões feitas sobre o conjunto de teste, utilizando métricas fundamentais para a tarefa de segmentação de incêndios. As principais métricas computadas durante a avaliação foram *Loss*, *IoU*, *Precision* e *Recall*.

A função de perda utilizada foi a BCE, também conhecida como *Log Loss*. De acordo com (GOODMAN; MILLER; SMYTH, 1991) uma função de perda utilizada para problemas de classificação binária e pode ser aplicada efetivamente em tarefas de segmentação binária. Em problemas de segmentação, o objetivo é classificar cada *pixel* de uma imagem em uma das duas classes: a classe de interesse (no caso, incêndio) ou o fundo (não incêndio). A BCE mede a dissimilaridade entre a probabilidade predita pelo modelo para cada *pixel* e a classe verdadeira (1 ou 0) correspondente a esse *pixel*.

O coeficiente IoU foi calculado para avaliar a qualidade da sobreposição entre as áreas preditas e as áreas reais de incêndio, representadas pelas máscaras de *ground through*. O IoU foi calculado *pixel a pixel*, obtendo-se a razão entre a interseção e a união das áreas identificadas corretamente como incêndio:

$$\text{IoU} = \frac{\text{VerdadeirosPositivos(TP)}}{\text{TP} + \text{FalsosNegativos(FN)} + \text{FalsosPositivos(FP)}} \quad (1)$$

Essa métrica é especialmente útil para problemas de segmentação, pois penaliza de forma severa os falsos positivos e falsos negativos.

A Precisão foi calculada para avaliar a taxa de acerto do modelo nas observações classificadas como positivas. Ou seja, ela quantifica a proporção de *pixels* classificados corretamente como incêndio entre todos os *pixels* classificados como incêndio pelo modelo:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

O *Recall* foi calculado para medir a capacidade do modelo de identificar corretamente as observações positivas, ou seja, as áreas de incêndio. Ele calcula a proporção de *pixels* com incêndio corretamente classificados entre todos os *pixels* que realmente possuem incêndio. A equação utilizada foi:

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

Além das métricas quantitativas, foi efetuada uma análise qualitativa dos resultados. As previsões das máscaras de incêndio foram sobrepostas às imagens RGB originais, permitindo uma inspeção visual dos focos de incêndio detectados. Isso proporcionou uma melhor compreensão das áreas afetadas e do desempenho do modelo em diferentes regiões das imagens. Para quantificar e validar os resultados de maneira mais detalhada, matrizes de confusão *pixel a pixel* foram geradas, com o objetivo de avaliar a distribuição de TP, FP, FN e Verdadeiros Negativos (TN) em cada uma das imagens de teste.

A implementação completa do *pipeline* metodológico descrito neste capítulo, incluindo os dois ciclos experimentais, encontra-se disponível em um repositório público. O código contempla as rotinas de pré-processamento, treinamento, varredura de limiar, avaliação quantitativa e geração das figuras e tabelas apresentadas neste trabalho.³

³ Repositório público: <<https://github.com/Natalia-oli/mestrado-segmentacao-incendios-sentinel2>>.

Resultados

4.0.1 Primeiro Ciclo Experimental: Avaliação com *Binary Cross-entropy* e *Dice Loss*

Observou-se que a U-Net com VGG-16 apresentou desempenho consistentemente superior, enquanto a EfficientNetB3 melhorou de modo perceptível quando treinada com *Dice Loss* em comparação à *Binary Cross-Entropy*, sugerindo maior sensibilidade à classe minoritária fogo com perdas baseadas em sobreposição.

As curvas de treinamento e validação reforçam essa leitura: com VGG-16, as métricas convergem rapidamente a valores altos e estáveis; com EfficientNetB3, a *Dice Loss* acelera a subida e estabilização do *recall* e da métrica IoU frente à BCE. A Tabela 2 apresenta o comportamento das métricas na validação:

Tabela 2 – Métricas obtidas na validação para a U-Net com *encoders* e funções de perda (ciclo 1).

Função de Perda	Encoder	IoU	<i>F1-Score</i>	<i>precision</i>	<i>Recall</i>
Binary Crossentropy	VGG-16	0,9388	0,9684	0,9667	0,9701
Dice Loss	VGG-16	0,9378	0,9679	0,9670	0,9689
Binary Crossentropy	EfficientNetB3	0,7112	0,8312	0,8242	0,8384
Dice Loss	EfficientNetB3	0,7446	0,8536	0,8495	0,8576

Fonte: elaboração própria.

No teste, a liderança do VGG-16 com BCE se mantém, com melhora marginal sobre a validação, indicando boa capacidade de generalização. Ainda sobre VGG-16, a *Dice Loss* exibe *precision* ligeiramente superior (0,9736) às custas de *recall* menor, sinalizando um limiar efetivo mais conservador. Já no EfficientNetB3, a *Dice Loss* volta a ser a melhor escolha, elevando tanto IoU quanto *F1-Score* em relação à BCE, o que reforça o benefício de perdas sensíveis à classe minoritária para esse *encoder*. Veja a Tabela 3:

Tabela 3 – Métricas obtidas no *teste* para a U-Net com diferentes *encoders* e funções de perda (ciclo 1). Resultados calculados no limiar ótimo selecionado na validação.

Função de Perda	<i>Encoder</i>	IoU	<i>F1-Score</i>	<i>precision</i>	<i>Recall</i>
Binary Crossentropy	VGG-16	0,9406	0,9694	0,9683	0,9705
Dice Loss	VGG-16	0,9370	0,9675	0,9736	0,9615
Binary Crossentropy	EfficientNetB3	0,6702	0,8025	0,7910	0,8143
Dice Loss	EfficientNetB3	0,7014	0,8245	0,8079	0,8418

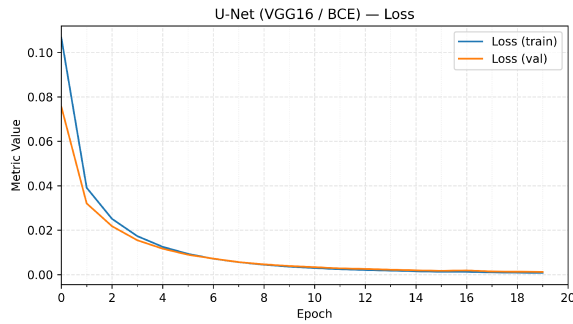
Fonte: elaboração própria.

Em relação a curva de perda da VGG-16 com BCE apresenta queda acentuada até aproximadamente a 4^a–6^a época e, depois, estabilização em patamar baixo, com distância mínima entre treino e validação. Esse comportamento sugere boa capacidade de generalização no regime atual (mesmo sem fortes *augmentations*), além de uma otimização eficiente da cabeça de segmentação. A proximidade entre as curvas mitiga a hipótese de *underfitting* e corrobora os resultados de validação reportados acima.

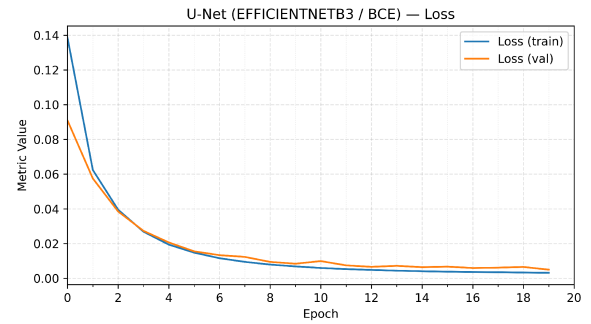
Com a EfficientNetB3, a perda também decai de forma consistente, porém a separação entre as curvas de treino e validação é mais evidente, sobretudo no início do treinamento. Isso indica maior volatilidade na fase inicial e possível dependência mais forte de hiperparâmetros (taxa de aprendizado, regularização) e da preparação dos dados (normalização e proporção classe minoritária). Embora a perda final também atinja patamar baixo, a diferença relativa para a VGG-16 antecipa as métricas de validação inferiores observadas a seguir, sugerindo que, sob BCE e nas condições atuais do conjunto, a arquitetura VGG-16 está mais alinhada com o padrão de heterogeneidade das imagens e com o grau de desbalanceamento presente (Figura 7).

Ainda na Figura 7, a trajetória de *F1-Score*/IoU confirma a leitura da perda: há ganho expressivo nas primeiras épocas e saturação precoce. O pareamento de *precision* e *recall* indica que o limiar global adotado está adequado, evitando tanto a supersegmentação quanto a omissão de focos. Em contextos de fogo esparsos, esse equilíbrio é particularmente relevante, pois o custo de FN pode ser alto, enquanto excesso de FP compromete a utilidade operacional. O perfil observado sugere que, para VGG-16+BCE, a capacidade representacional do *encoder* é suficiente para delinear fronteiras de queima com boa consistência intra-época.

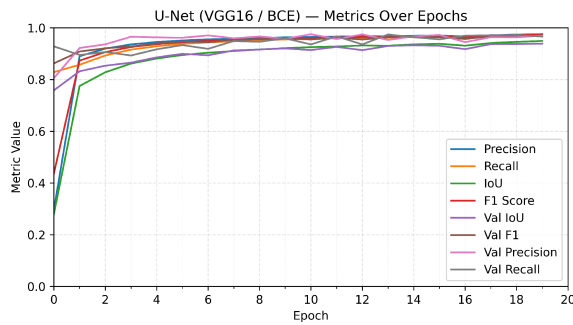
Apesar de convergir, a EfficientNetB3 atinge níveis de F1/IoU visivelmente menores, com flutuações iniciais mais marcadas entre *precision* e *recall*, a Figura 7 representa esse comportamento. Esse padrão é compatível com uma fronteira de decisão mais sensível ao desbalanceamento espacial, quando treinada apenas com BCE. Portanto, a penalização uniforme da BCE, sem reponderação, pode não enfatizar suficientemente a classe minoritária para esse *encoder*, que tende a exigir *losses* ou *schemes* de treino mais robustos.



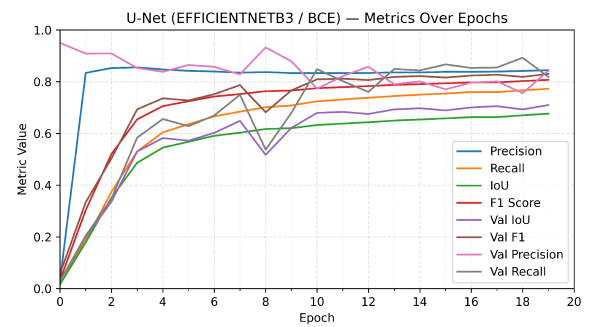
(a) VGG-16 + BCE — perda converge rapidamente.



(b) EfficientNetB3 + BCE — perda reduz, mas exibe maior distância treino-validação.



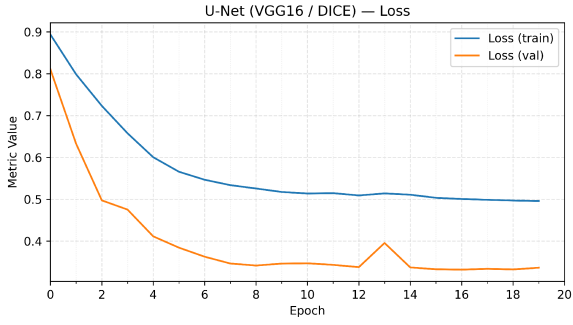
(c) VGG-16 + BCE — $F1$ -Score e IoU saturam cedo ($\approx 0,96$ – $0,94$), com $precision/recall$ balanceados.



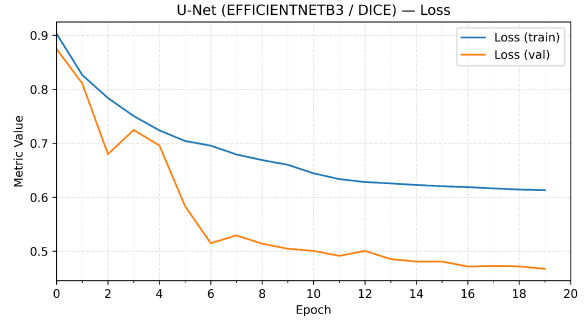
EfficientNetB3 + BCE — $F1$ -Score/ IoU inferiores ($\approx 0,83/0,71$) e maior variabilidade inicial.

Figura 7 – Curvas de treinamento e métricas para *encoders* com *Binary Crossentropy* no Ciclo 1.

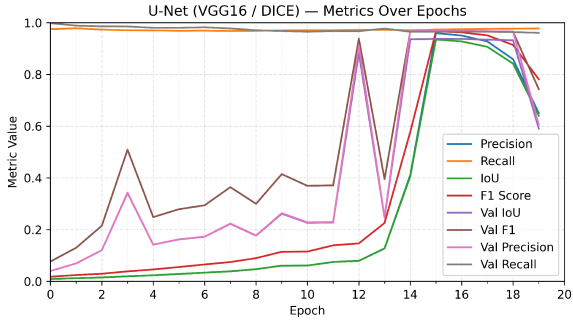
Ao trocar *BCE* por *Dice* (Figura 8), o VGG-16 mantém desempenho alto e estável, enquanto o EfficientNetB3 apresenta ganho substancial em $F1$ -Score/ IoU . Essa melhora é coerente com a própria definição da *Dice Loss*. Em síntese, a *Dice* reduz a lacuna entre os *encoders*, embora o VGG-16 permaneça superior neste ciclo.



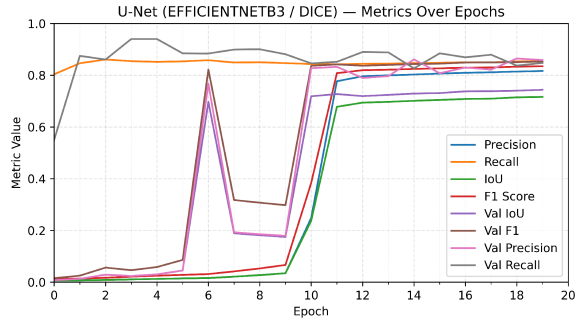
(a) VGG-16 + *Dice* — *val. loss* cai de forma monotônica e estabiliza, sem sinal claro de *overfitting*.



(b) EfficientNetB3 + *Dice* — *loss* reduz, mas mantém maior distância treino–validação até ~época 10 (convergência mais lenta).



(c) VGG-16 + *Dice* — *F1-Score*/IoU sobem rapidamente e saturam em $\approx 0,968/0,937$; *precision/recall* $> 0,96$.



(d) EfficientNetB3 + *Dice* — *F1-Score*/IoU atingem $\approx 0,853/0,745$ na validação, com maior variabilidade nas primeiras épocas.

Figura 8 – Curvas de treinamento e métricas para *Dice Loss*.

Nas Figuras 8(a)–(b), observa-se que a *val. loss* do VGG-16 decresce de forma monotônica, aproximando-se rapidamente da curva de treino e sugerindo boa generalização já nas primeiras épocas. Em contraste, o EfficientNetB3 mantém maior distância treino–validação até aproximadamente a época 10, indicando convergência mais lenta e maior sensibilidade às primeiras atualizações dos pesos. Esse comportamento é compatível com *encoders* mais profundos e com maior regularização implícita, que tendem a exigir mais épocas para estabilizar a otimização quando treinados com *Dice*.

As Figuras 8(c)–(d) explicitam o efeito da *Dice*: o VGG-16 atinge e sustenta métricas elevadas (*F1-Score* $\approx 0,968$, IoU $\approx 0,938$, com *precision/recall* superiores a 0,96), enquanto o EfficientNetB3 melhora substancialmente em relação ao BCE, estabilizando em *F1-Score* $\approx 0,854$ e IoU $\approx 0,745$. Notam-se oscilações iniciais no EfficientNetB3 (variação entre épocas ~6–10), que se atenuam após a redução adaptativa da taxa de aprendizado (*ReduceLRonPlateau*), produzindo trajetórias mais suaves e consistente crescimento das métricas. Em termos práticos, a *Dice* diminui o viés para o fundo e aumenta a sobreposição nas regiões minoritárias; ainda assim, neste ciclo, o VGG-16 permanece superior e mais estável ao longo de todo o treinamento.

Esses achados motivam, para o ciclo 2, (i) testar perda composta, (ii) reforçar *data*

augmentation espacial/fotométrica para reduzir a variância nas épocas iniciais do EfficientNetB3, e (iii) calibrar o limiar de decisão com base na curva *precision–recall* do conjunto de validação. A expectativa é reduzir o *gap* treino–validação do EfficientNetB3 nas primeiras épocas e aumentar sua IoU sem degradar *precision*, aproximando seu patamar do observado no VGG-16.

As matrizes de confusão normalizadas da validação (Figura 9) mostram um padrão consistente com o desbalanceamento do conjunto: a classe de fundo apresenta *false positives* praticamente nulos em todos os cenários (especificidade muito alta), enquanto as diferenças entre arquiteturas e funções de perda emergem sobretudo no *recall* da classe fogo (*true positives*). Em termos práticos, o erro dominante é de *false negative*, isto é, regiões com fogo que deixam de ser segmentadas.

Comparando as combinações com *Binary Crossentropy*, observa-se que o VGG-16 atinge $TP \approx 0,97$ e $FN \approx 0,03$ com limiar ótimo 0,45 [Figura 9(a)], ao passo que o EfficientNetB3, no mesmo limiar, retém $TP \approx 0,84$ e $FN \approx 0,16$ [Figura 9(b)]. Esse hiato de ~ 13 p.p. em *recall* explica a superioridade do VGG-16 nas métricas agregadas (*F1-Score* e IoU) sob *BCE*: ambos os modelos mantêm *precision* elevada graças ao fundo dominante, mas o VGG-16 erra menos por omissão de fogo.

Ao substituir *BCE* por *Dice Loss* e ajustar o limiar ótimo para 0,40, o VGG-16 virtualmente preserva o desempenho ($TP \approx 0,97$; $FN \approx 0,03$) [Figura 9(c)], sugerindo saídas já bem calibradas em torno da fronteira de decisão. O EfficientNetB3, por sua vez, apresenta ganho mensurável de *recall* ($TP \approx 0,86$; $FN \approx 0,14$) sem penalizar os *FP* [Figura 9(d)]. Esse comportamento é compatível com a natureza da *Dice*, que recompensa diretamente a sobreposição entre máscaras e, portanto, reduz o viés para a classe majoritária típico da *BCE* em cenários altamente desbalanceados.

Esses resultados reforçam que a escolha do ponto de operação (função de perda + limiar) deve ser orientada pelo custo do erro: para *monitoramento* e *alerta* de incêndios, costuma-se privilegiar *recall* elevado (redução de *FN*), aceitando um pequeno incremento de *FP* desde que operacionalmente manejável. No presente ciclo, o VGG-16 já opera próximo do teto de *recall* com *FP* mínimos; o EfficientNetB3 responde melhor a ajustes sensíveis ao desbalanceamento (p. ex., *Dice*) e ao limiar, indicando margem adicional para otimização por curva *precision–recall* e calibração de probabilidade.

Do ponto de vista diagnóstico, a predominância de *FN* sugere três fontes prováveis de erro: (i) focos pequenos e fragmentados, que desaparecem com o *downsampling* para 256×256 e com filtros de maior *stride*; (ii) bordas de fogo finas ou parcialmente encobertas por fumaça, nas quais a incerteza do modelo tende a ficar abaixo do limiar; e (iii) confusão com padrões espectrais de solo exposto/rocha brilhante em ângulos solares extremos, que reduzem o contraste com a classe fogo.

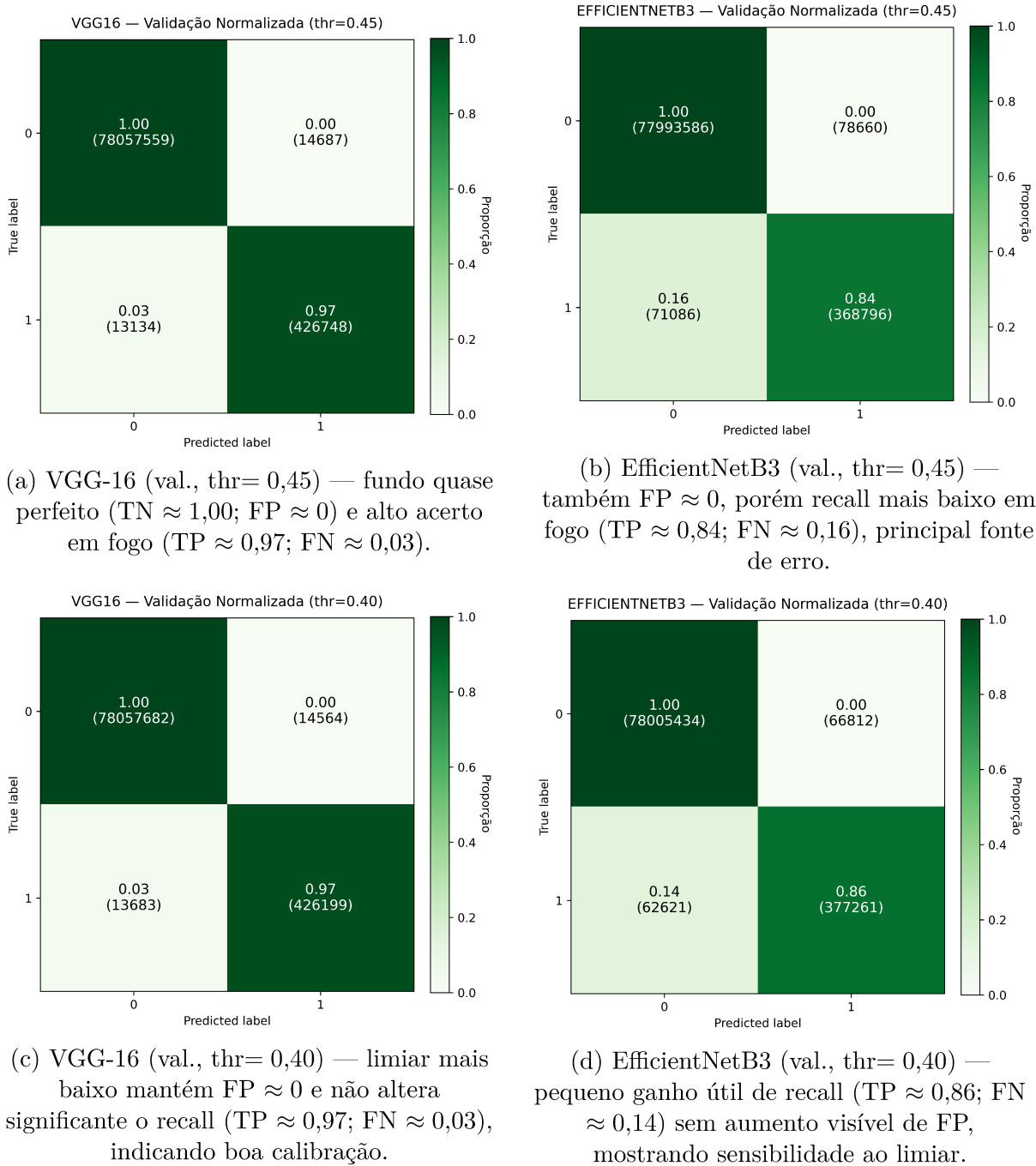


Figura 9 – Matrizes de confusão normalizadas na validação com limiares ótimos por função de perda (BCE: 0,45; *Dice*: 0,40). Ambos os modelos mantêm FP muito baixos devido ao fundo dominante; a diferença central está no *recall* da classe fogo.

Em síntese, as matrizes corroboram a narrativa das curvas de treinamento: ambos os *encoders* mantêm *precision* alta em virtude do fundo dominante, e a diferenciação entre modelos e perdas aparece principalmente no *recall* da classe de interesse. A *Dice* reduz parcialmente a lacuna do EfficientNetB3 relativamente ao VGG-16, mas, neste ciclo, o VGG-16 permanece superior e mais estável. Esse quadro estabelece uma linha de base sólida para avaliar, no ciclo 2, o impacto de perdas compostas, multiescala e calibração

final do limiar sobre a redução de FN sem degradação relevante de $precision$.

Os resultados no conjunto de teste confirmam o padrão observado na validação, a partir das Figuras 10(a)–(c), o VGG-16 permanece consistentemente superior ao EfficientNetB3 em todos os cenários, com diferenças mais pronunciadas quando se utiliza *Binary Crossentropy* (BCE). Com BCE, o VGG-16 preserva alta sensibilidade à classe fogo ($TP \approx 0,97$) e erro de omissão reduzido ($FN \approx 0,03$), sem penalizar a classe de fundo, que se mantém com FP próximos de zero. Ao migrar para *Dice*, o VGG-16 mantém o mesmo regime de pouca confusão no fundo e apenas ligeira queda de $recall$ ($TP \approx 0,96$), sugerindo boa calibração do limiar e robustez do *encoder* diante de mudanças na função de perda.

Para o EfficientNetB3, a troca de BCE por *Dice* é decisiva: o $recall$ de fogo sobe de aproximadamente 0,81 para 0,86, sem aumento visível de FP. Em termos práticos, *Dice* reduz o viés para a classe majoritária (fundo), melhorando a cobertura de regiões queimadas que tendem a ser pequenas, esparsas e fragmentadas. Essa melhora, ainda que não elimine a diferença para o VGG-16, reduz a lacuna entre os *encoders* e reforça que perdas *class-imbalance aware* favorecem tarefas em que a classe de interesse ocupa fração diminuta da cena.

As quatro matrizes de confusão evidenciam um traço comum: FP residuais para o fundo ($TN \approx 1,00$) em todos os modelos e perdas, refletindo o forte desbalanceamento espacial do problema. Assim, a métrica que realmente distingue as combinações é o $recall$ da classe fogo. No VGG-16, BCE e *Dice* produzem acertos elevados e estáveis em fogo ($TP \approx 0,97$ e $\approx 0,96$), ao passo que, no EfficientNetB3, *Dice* atenua omissões sistemáticas (queda de FN de $\approx 0,19$ para $\approx 0,14$). Esse comportamento é coerente com as curvas de treinamento: no EfficientNetB3 observam-se oscilações nas primeiras épocas e uma convergência mais lenta, o que torna a perda *Dice* particularmente útil para estabilizar o aprendizado da classe minoritária.

Sob a ótica de decisão, os limiares selecionados na validação (BCE: 0,45; *Dice*: 0,40) generalizam para o teste sem distorções visíveis, indicando boa *calibração* dos scores. Em problemas altamente desbalanceados, pequenas variações de limiar podem deslocar o ponto de operação no espaço *precision–recall*; aqui, o ajuste mais baixo para *Dice* preserva FP próximos de zero e melhora a sensibilidade do EfficientNetB3, sem sacrificar a seletividade do VGG-16. Isso sugere que, além da função de perda, a seleção criteriosa do limiar a partir de curvas *Precision–Recall* (curva Precisão–Sensibilidade) (PR) deve ser mantida como procedimento padrão.

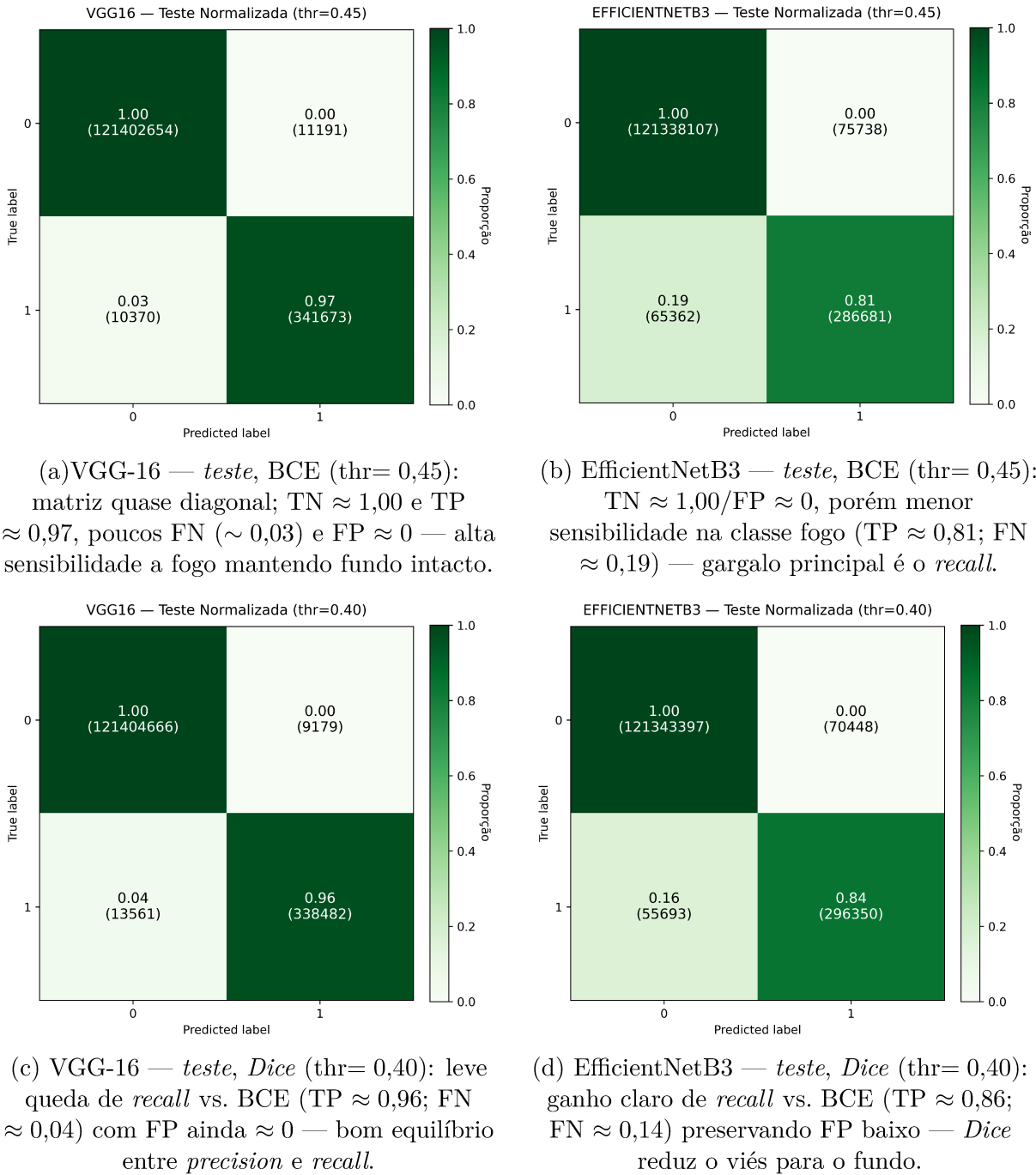


Figura 10 – Matrizes de confusão normalizadas no conjunto de teste para U-Net com VGG-16 e EfficientNetB3, avaliadas com limiares escolhidos na validação (BCE: 0,45; Dice: 0,40). Em todos os casos, o fundo apresenta FP ≈ 0 (TN $\approx 1,00$), evidenciando o desequilíbrio de classes.

O conjunto de teste confirma que: (i) o VGG-16 é um ponto de referência estável e de alta sensibilidade, relativamente indiferente à escolha entre BCE e Dice; (ii) o EfficientNetB3 depende mais de perdas *class-imbalance aware* para atingir patamares competitivos de *recall*; e (iii) a combinação entre função de perda adequada e seleção de limiar orientada por validação é determinante para controlar o compromisso entre *precision* e *recall*.

em cenários extremos de desbalanceamento.

A análise dos erros indica que os principais falsos negativos concentram-se em frentes de fogo muito delgadas, bordas semânticas ruidosas (transições fogo–solo exposto) e artefatos espectrais que mimetizam cicatrizes (sombra de nuvens, trilhas viárias e campos recém-expostos). Esses padrões são consistentes com a superioridade do *VGG-16* em limites e texturas de baixa frequência e com a maior sensibilidade do EfficientNetB3 ao regime de *optimization warm-up*, que, quando combinado à *Dice*, aprende a “valorizar” regiões positivas raras e descontínuas. Tais achados motivam, para o ciclo 2, explorar perdas compostas (por exemplo, *Dice+Focal*) visando reduzir omissões residuais, incorporar *data augmentation* espacial/fotométrica mais agressiva para mitigar a variância inicial do EfficientNetB3, e avaliar pós-processamentos morfológicos leves ou condicionados a contexto para recompor fragmentação em cicatrizes finas, sem inflacionar FP.

As Tabelas 4 e 5 sintetizam o Ciclo 1. No conjunto de *validação*, o *VGG-16* apresenta os melhores valores de *F1-Score* (0.968) e IoU (0.939) com BCE, mantendo desempenho praticamente idêntico com *Dice*, o que indica robustez do *encoder* às escolhas de função de perda. Para o EfficientNetB3, a substituição de BCE por *Dice* eleva *F1-Score* de 0.831 para 0.854 e IoU de 0.711 para 0.745, reduzindo o viés para o fundo e melhorando a cobertura de regiões queimadas.

Tabela 4 – Métricas na validação (limiares oriundos da validação).

Backbone	Loss	Thr	<i>precision</i>	Recall	<i>F1-Score</i>	IoU
EfficientNetB3	BCE	0.450	0.824	0.839	0.831	0.711
EfficientNetB3	Dice	0.400	0.850	0.858	0.854	0.745
VGG-16	BCE	0.450	0.967	0.970	0.968	0.939
VGG-16	Dice	0.400	0.967	0.969	0.968	0.938

Nota: negrito destaca o melhor *F1-Score* e IoU na validação.

Tabela 5 – Métricas no teste (usando os limiares da validação).

Backbone	Loss	Thr	<i>precision</i>	Recall	<i>F1-Score</i>	IoU
EfficientNetB3	BCE	0.450	0.791	0.814	0.803	0.670
EfficientNetB3	Dice	0.400	0.808	0.842	0.825	0.701
VGG-16	BCE	0.450	0.968	0.971	0.969	0.941
VGG-16	Dice	0.400	0.974	0.961	0.968	0.937

Nota: negrito destaca o melhor F1 e IoU no teste.

Limiar adotado: BCE = 0,45; *Dice* = 0,40.

No teste, o comportamento se confirma: o VGG-16 com BCE atinge os maiores *F1-Score* (0.969) e IoU (0.941), enquanto o EfficientNetB3 volta a se beneficiar de *Dice* (*F1-Score* de 0.825 e IoU de 0.701, superiores aos de BCE). Importante notar que os limiares definidos na validação (0.45 para BCE e 0.40 para *Dice*) transferem-se bem para o teste, sem inflação de falsos positivos: o ganho de EfficientNetB3 com *Dice* decorre majoritariamente de maior *recall* da classe fogo.

Em síntese, o VGG-16 estabelece o patamar de referência no Ciclo 1, com alta sensibilidade e sobreposição espacial; o EfficientNetB3, por sua vez, requer funções de perda *class-imbalance aware* (*Dice*) para operar em regime competitivo. Esses achados orientam as intervenções do Ciclo 2 (perdas compostas, *data augmentation* mais agressiva e calibração de limiar via curvas PR) para reduzir omissões residuais sem sacrificar a seletividade.

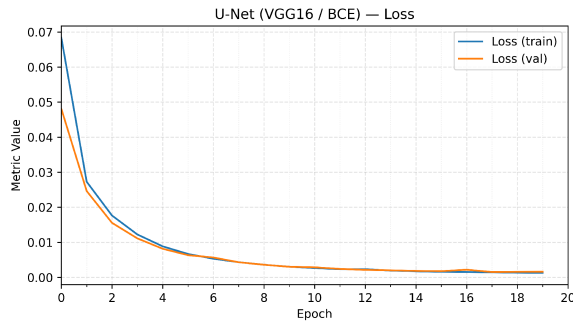
4.0.2 Ciclo 2: objetivos, desenho experimental e comparativo com o Ciclo 1

Considerando o contexto do Ciclo 1 e as alterações implementadas no Ciclo 2, algumas análises estão disponíveis abaixo.

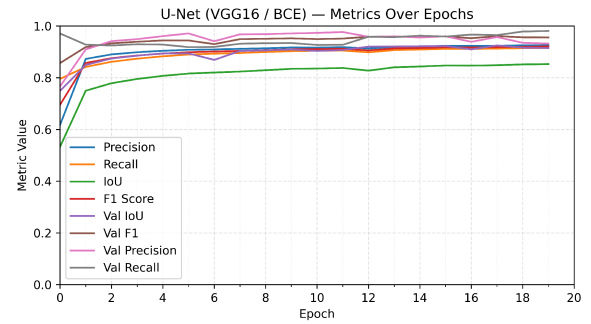
A Figura 11 sintetiza o comportamento de treino e validação da U-Net com *backbone* VGG-16 no Ciclo 2 para as quatro funções de perda avaliadas. De forma geral, observa-se que o uso de *undersampling* aliado às *augmentations* permitiu que todos os cenários convergissem rapidamente nas primeiras épocas, com reduções consistentes da função de perda e estabilização das métricas acima de 0,9 para a maior parte das combinações. Nas subfiguras (a1) e (a2), correspondentes à BCE, a perda decresce de maneira monótona e suave, enquanto *F1-Score*, IoU, *precision* e *recall* de treino e validação apresentam trajetórias muito próximas desde as primeiras épocas. Esse comportamento indica um regime de treino estável, com baixa discrepância entre conjuntos e ausência de sobreajuste marcado. A BCE, embora originalmente menos sensível ao desbalanceamento extremo, passa a se beneficiar da distribuição mais equilibrada de *pixels* obtida com o *undersampling* do Ciclo 2, o que explica a elevação dos valores de IoU e *F1-Score* em relação ao Ciclo 1, sem sacrificar a *recall* da classe fogo.

As subfiguras (b1) e (b2), associadas à perda *Dice*, revelam um comportamento mais irregular nas épocas iniciais, com oscilações nas métricas de validação e um distanciamento maior em relação às curvas de treino. Essa instabilidade inicial é consistente com a natureza da *Dice Loss* em cenários fortemente desbalanceados, uma vez que pequenas variações na segmentação dos poucos *pixels* positivos podem provocar flutuações significativas na métrica. Ainda assim, após algumas épocas, as curvas tendem à estabilização e *F1-Score* e IoU atingem patamares satisfatórios, embora inferiores aos observados com BCE, *Focal* e *DiceFocal*. Em termos práticos, isso indica que a perda puramente baseada em sobreposição de regiões melhora a detecção de áreas queimadas em relação ao Ciclo 1, mas ainda não explora todo o potencial da VGG-16 quando comparada às abordagens que combinam ou reponderam exemplos difíceis.

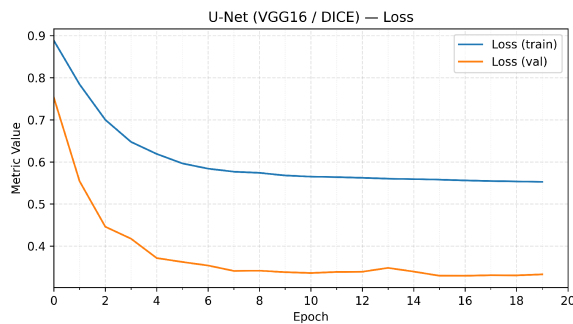
Já ao observar as subfiguras (c1) e (c2), que representam a *Focal Loss*, nota-se um rápido decaimento da perda, com métricas de validação elevadas desde as primeiras épocas. Há, contudo, um pico pontual de instabilidade em uma das épocas, visível como um aumento abrupto no valor da perda e queda temporária nas métricas, provavelmente associado a um *minibatch* com distribuição atípica de *pixels* positivos ou à alta sensibilidade da *Focal* a exemplos considerados “difíceis”. Essa oscilação é rapidamente corrigida nas épocas subsequentes, e o treinamento converge para um regime em que *precision*, *recall* e *F1* permanecem próximas de 1,0. O resultado final é uma rede capaz de manter alta *recall* dos *pixels* de fogo, reduzindo falsos negativos em comparação à BCE e principalmente à *Dice*, o que é coerente com os altos valores de *F1-Score* e IoU observados nas tabelas de desempenho do Ciclo 2.



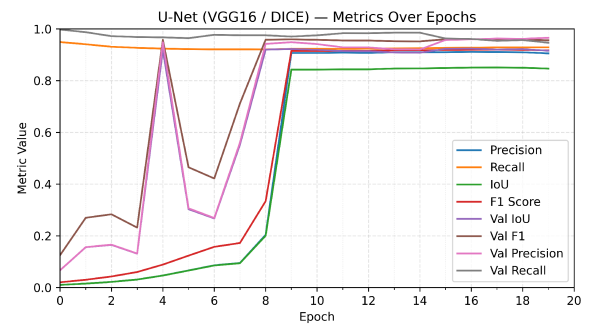
(a1) BCE: perda decresce e estabiliza (treino e val).



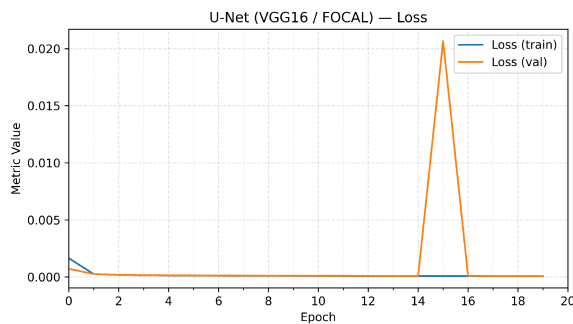
(a2) BCE: $F1\text{-Score}$ /IoU sobem e entram em platô; precision/recall estáveis.



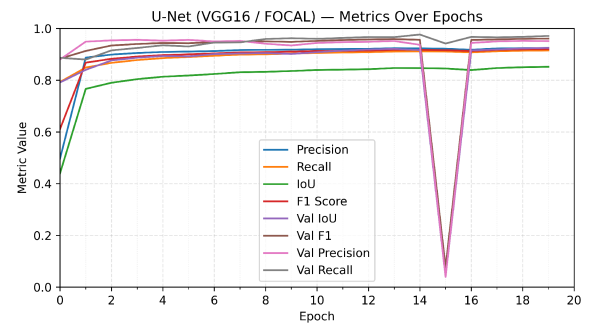
(b1) *Dice*: perda cai no início e estabiliza; maior *gap* treino-val.



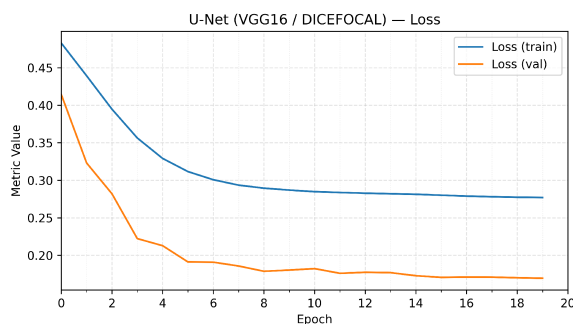
(b2) *Dice*: métricas oscilam no início e estabilizam após poucas épocas.



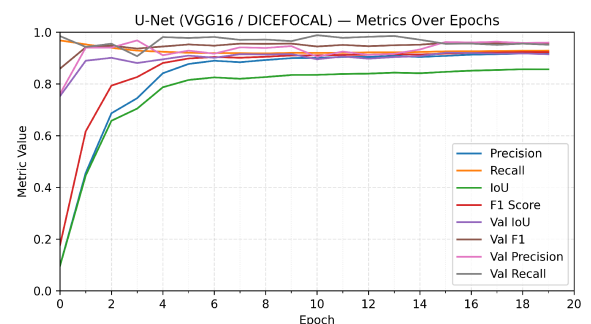
(c1) *Focal*: perda baixa e estável; pico isolado na validação.



(c2) *Focal*: métricas em platô; queda pontual (ép. 15) e recuperação.



(d1) *DiceFocal*: perda decresce e estabiliza (val < treino).

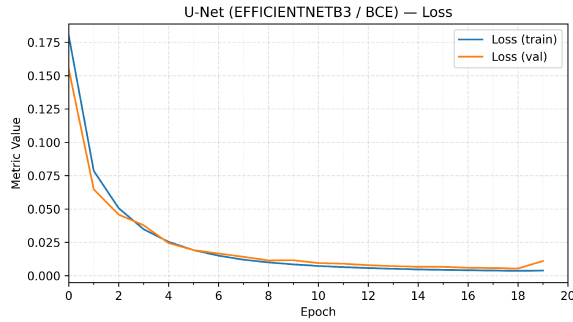


(d2) *DiceFocal*: métricas sobem rápido e estabilizam, com pouca oscilação.

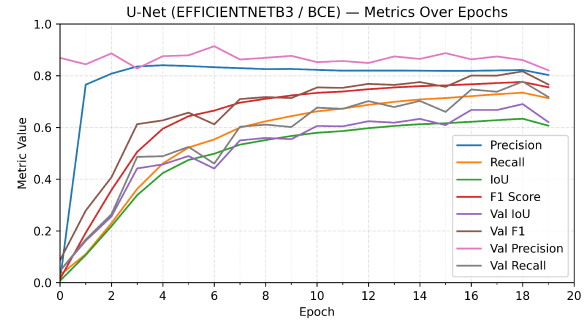
Figura 11 – Curvas de perda e métricas de treino/validação da U-Net com *backbone* VGG-16 no Ciclo 2, considerando as diferentes funções de perda.

Por fim, as subfiguras (d1) e (d2) mostram o comportamento da função composta *DiceFocal*. Ao combinar um termo de sobreposição (*Dice*) com o mecanismo de reponderação da *Focal*, observa-se um compromisso mais equilibrado entre estabilidade e sensibilidade à classe minoritária. As curvas de perda apresentam decaimento suave, sem os picos abruptos observados em *Focal* isolada, enquanto as métricas de treino e validação se mantêm altas e com pequena lacuna entre si ao longo de todo o treinamento. Esse padrão indica que a função composta conseguiu explorar a capacidade representacional da VGG-16 ao mesmo tempo em que mitiga as principais limitações de cada componente individual: reduz a instabilidade típica da *Dice* em cenários extremamente desbalanceados e, ao mesmo tempo, controla a agressividade da *Focal* em relação a exemplos raros. Em conjunto com os resultados de validação e teste, conclui-se que, no contexto do Ciclo 2, a VGG-16 combinada com perdas *Focal* e, sobretudo, *DiceFocal* se destaca como a configuração mais robusta, superando de forma consistente os desempenhos obtidos no Ciclo 1.

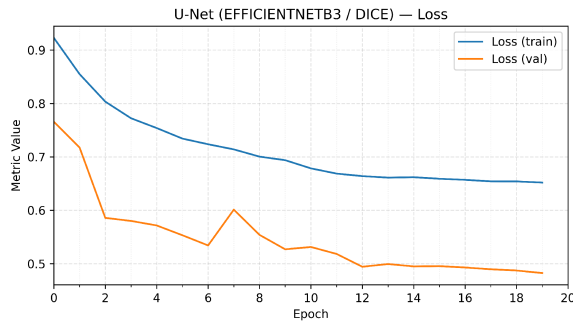
A Figura 12 apresenta as curvas de perda e métricas para a U-Net com *backbone* EfficientNetB3 no Ciclo 2, também sob as quatro funções de perda avaliadas.



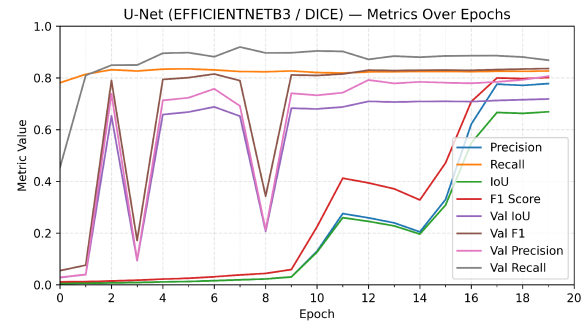
(a1) BCE: perda cai rápido e estabiliza baixa.



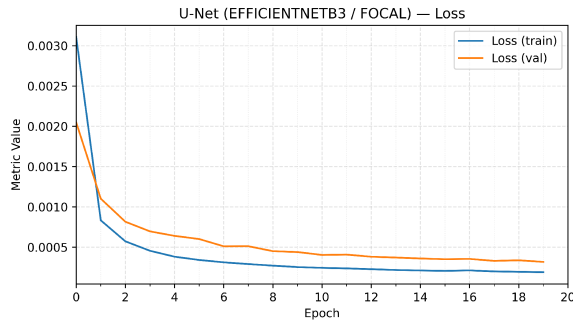
(a2) BCE: métricas sobem cedo e entram em platô.



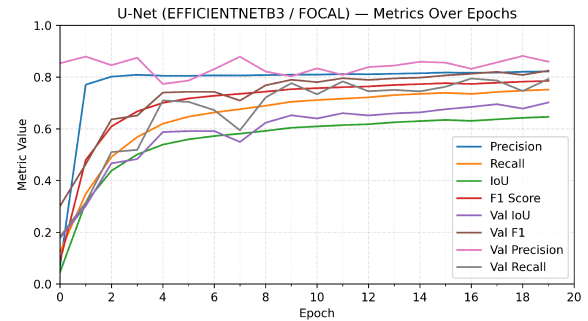
(b1) Dice: perda diminui, mas estabiliza mais alta.



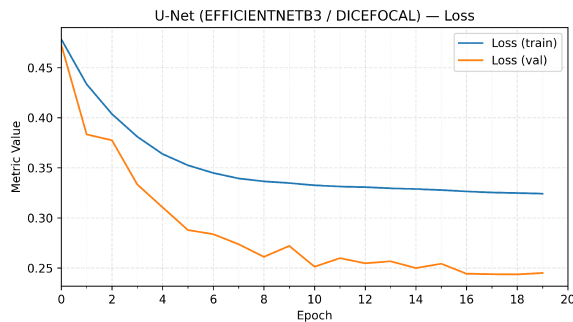
(b2) Dice: instável no início; melhora só no fim.



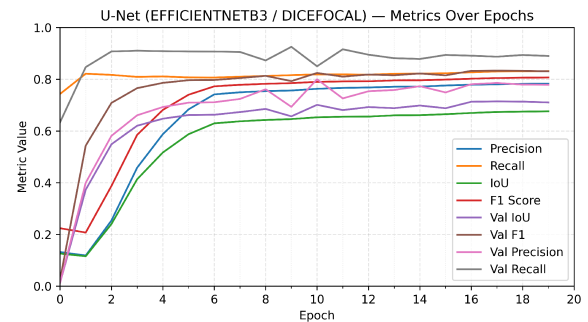
(c1) Focal: perda cai rápido e estabiliza baixa.



(c2) Focal: métricas sobem cedo e entram em platô.



(d1) Dice+Focal: perda reduz mais lentamente e estabiliza mais alta.



(d2) Dice+Focal: melhora gradual e platô mais alto nas métricas.

Figura 12 – Curvas de perda e métricas de treino/validação da U-Net com *backbone EfficientNetB3* no Ciclo 2, considerando as diferentes funções de perda.

De maneira geral, todas as combinações convergem para valores altos de $F1$ -Score e IoU, porém com patamares ligeiramente inferiores aos observados para a VGG-16 o que reforça a tendência já identificada no Ciclo 1 de que a VGG-16 se adapta melhor às características espectrais e espaciais das cenas de queimadas consideradas neste trabalho. Nas subfiguras (a1) e (a2), correspondentes à BCE, a perda apresenta decaimento rápido e estável, enquanto as métricas de treino e validação crescem de forma quase monótona, alcançando valores em torno de 0,8–0,85 de $F1$ -Score e IoU. A distância entre curvas de treino e validação é pequena, sugerindo boa capacidade de generalização após a aplicação do *undersampling* e das *augmentations*. No entanto, o patamar final permanece abaixo daquele obtido pela VGG-16 com a mesma função de perda, indicando que a maior profundidade e complexidade da EfficientNetB3 não se convertem necessariamente em ganhos de desempenho para esse tipo de segmentação binária em dados Sentinel-2 altamente desbalanceados.

O comportamento observado com a *Dice Loss*, nas subfiguras (b1) e (b2), reforça essa diferença de sensibilidade entre *backbones*. Assim como na VGG-16, as curvas de perda e métricas apresentam maior variabilidade nas primeiras épocas, mas no caso da EfficientNetB3 essa instabilidade é ainda mais pronunciada e persiste por um número maior de épocas, refletindo uma resposta mais sensível às flutuações na proporção de *pixels* positivos após o *undersampling*. Apesar disso, o modelo converge para valores de $F1$ -Score e IoU satisfatórios, ligeiramente superiores aos obtidos com BCE, o que sugere que a *Dice* contribui para melhorar a segmentação das regiões afetadas pelo fogo.

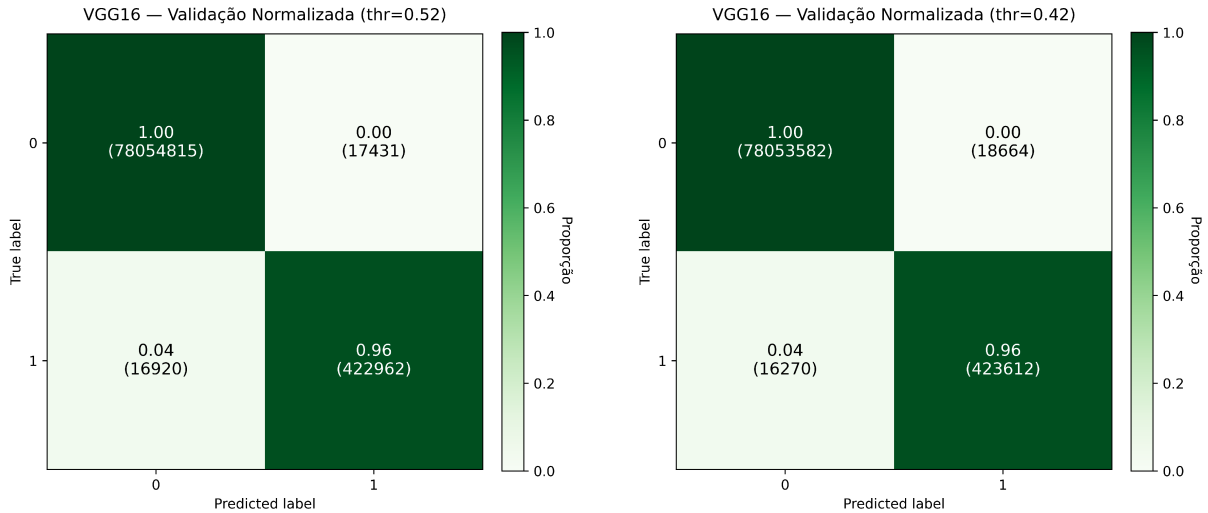
Nas subfiguras (c1) e (c2), associadas à Focal Loss, a EfficientNetB3 apresenta curvas de perda decrescentes e métricas que aumentam de forma relativamente suave, sem os picos acentuados de instabilidade observados na VGG-16. Isso sugere que, para esse *backbone*, a *Focal Loss* é numericamente mais estável ao longo do treino, possivelmente em função da forma como a EfficientNetB3 distribui a ativação ao longo das camadas profundas. Ainda assim, os patamares finais de $F1$ -Score e IoU permanecem ligeiramente abaixo daqueles alcançados pela ' com *Focal*, indicando que a maior estabilidade numérica não se traduz, necessariamente, em superioridade de desempenho. Em contrapartida, compara-se favoravelmente à BCE e à *Dice* no próprio EfficientNetB3, reduzindo o número de falsos negativos e aproximando o modelo da performance desejada para aplicação prática.

Já as subfiguras (d1) e (d2) mostram o comportamento da perda composta *DiceFocal* na EfficientNetB3. As curvas de perda exibem decaimento regular, com estabilização nas últimas épocas, enquanto as métricas de treino e validação convergem para valores de $F1$ -Score próximos aos observados com *Focal*, mas com IoU ligeiramente superior e menor discrepância entre conjuntos. Esse padrão indica que a combinação dos termos *Dice* e *Focal* ajuda a EfficientNetB3 a explorar melhor as regiões de fronteira das queimadas, mantendo alta *recall* e *precision* mesmo diante da forte assimetria entre classes. Quando

comparada às curvas equivalentes da VGG-16, nota-se que o *backbone* EfficientNetB3 se beneficia mais claramente da função composta do que das perdas individuais, reduzindo o gap em relação ao desempenho da VGG-16, ainda que não o elimine.

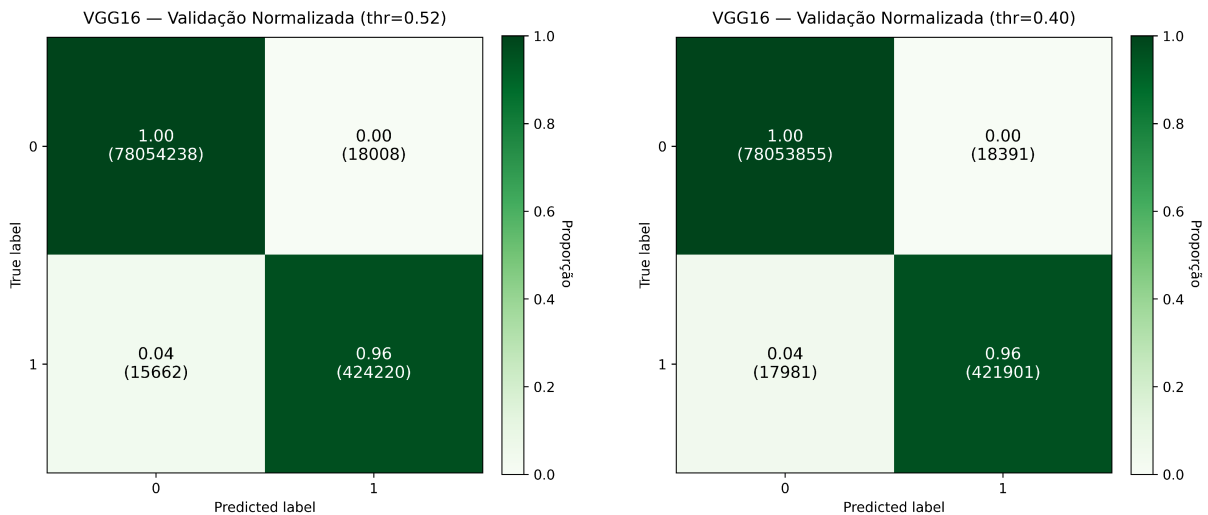
Enquanto a VGG-16 atinge os melhores resultados absolutos, a EfficientNetB3 torna-se competitiva quando associada a funções de perda mais sofisticadas, como a *DiceFocal*, o que é particularmente relevante em cenários operacionais em que se deseja explorar arquiteturas mais modernas preservando, ao mesmo tempo, a robustez frente ao desbalanceamento extremo dos dados.

Os mapas de confusão normalizados da UNet com *backbone* VGG-16 no conjunto de validação do Ciclo 2, representado nas figuras 13, 14, 15 e 16 evidenciam que todas as funções de perda mantêm desempenho extremamente alto para a classe “não fogo”. Em todos os casos, a proporção de verdadeiros negativos aproxima-se de 1,0, com praticamente ausência de falsos positivos, indicando que o modelo raramente rotula o fundo como fogo. Para a classe de interesse, observa-se que aproximadamente 0,96 dos *pixels* são corretamente identificados, enquanto cerca de 0,04 continuam sendo confundidos com fundo. Essa estabilidade ao longo de *BCE*, *Dice*, *Focal* e, sobretudo, *DiceFocal* reflete o efeito combinado do *undersampling* e da função de perda composta, que reduz a tendência observada no Ciclo 1 de favorecer excessivamente a classe majoritária e penaliza de forma mais equilibrada falsos positivos e falsos negativos para fogo. Além disso, os limiares ótimos distintos (por exemplo, $\text{thr} = 0,52$ para *BCE*, 0,42 para *Dice* e 0,40 para *DiceFocal*) mostram que o ajuste fino do limiar de decisão é relevante para explorar todo o potencial da métrica *F1-Score*, mas não altera qualitativamente o padrão de alta separação entre as classes.



BCE — validação: FP 0; principal erro é FN 4% (incêndios perdidos)

Dice — validação: mantém FP 0 e FN 4% (desempenho muito similar)



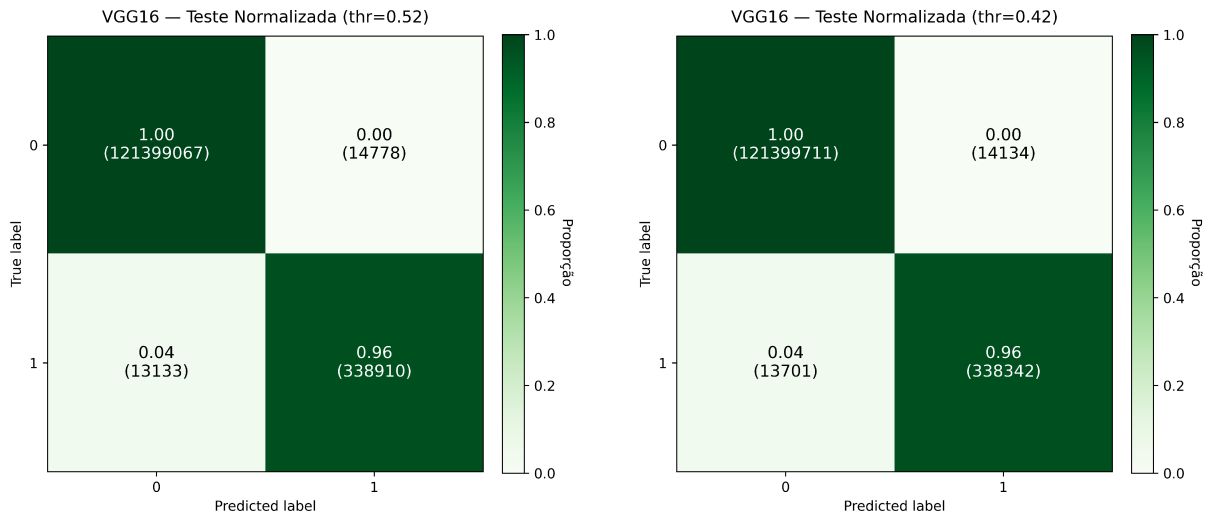
Focal — validação: FP 0; FN 4% continua sendo o erro dominante.

DiceFocal — validação: FP 0; FN 4%, com padrão praticamente igual às demais.

Figura 13 – Matrizes de confusão normalizadas da U-Net com *backbone*VGG-16 no conjunto de validação do Ciclo 2, considerando as diferentes funções de perda.

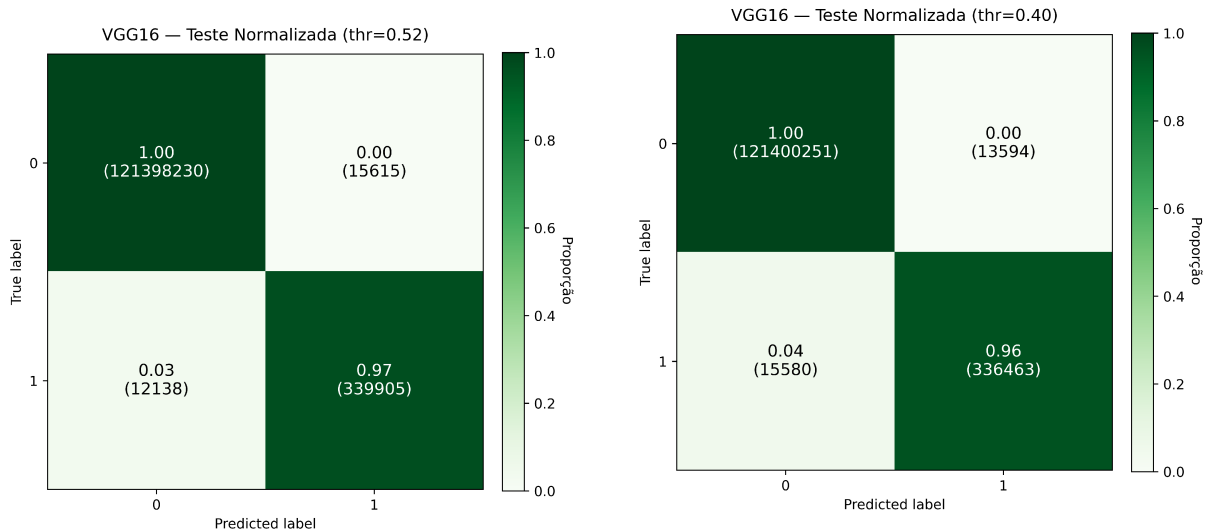
Quando se analisa as matrizes de confusão normalizadas do VGG-16 no conjunto de teste nas Figuras 14 e 16 abaixo, o comportamento observado em validação é essencialmente preservado. A diagonal principal permanece concentrada, com taxas de verdadeiros positivos para fogo entre 0,96 e 0,97 e fração de falsos negativos em torno de 0,03–0,04, o que indica boa capacidade de generalização do modelo para imagens não utilizadas no ajuste de hiperparâmetros. A manutenção dessa estrutura entre validação e teste é particularmente evidente para a combinação VGG-16 + *DiceFocal*, que apresentou os melhores valores globais de *F1-Score* e IoU no Ciclo 2. Em comparação com o Ciclo 1, as matrizes do Ciclo 2 exibem maior densidade na diagonal principal e redução visual de erros na classe minoritária, sinalizando que a estratégia atual de balanceamento e escolha de fun-

ção de perda foi mais eficaz para capturar cicatrizes de incêndio em cenários altamente desbalanceados.



BCE - Alta especificidade (quase zero FP) e *recall* da classe 1 de 96% (FN 4%).

Dice - Desempenho semelhante ao BCE: quase zero FP e *recall* da classe 1 de 96%.



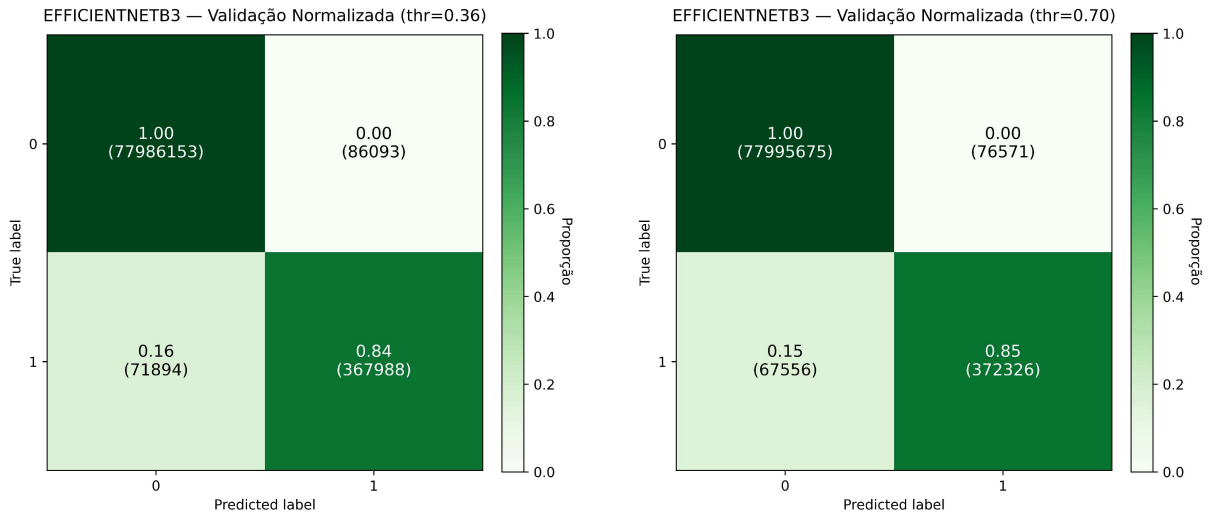
Focal - Melhor *recall* da classe 1 (97%), mantendo FP praticamente nulo.

DiceFocal — *zRecall* da classe 1 96% com FP praticamente nulo, porém com FN ligeiramente maior.

Figura 14 – Matrizes de confusão normalizadas da U-Net com *backbone* VGG-16 no conjunto de teste do Ciclo 2, considerando as diferentes funções de perda.

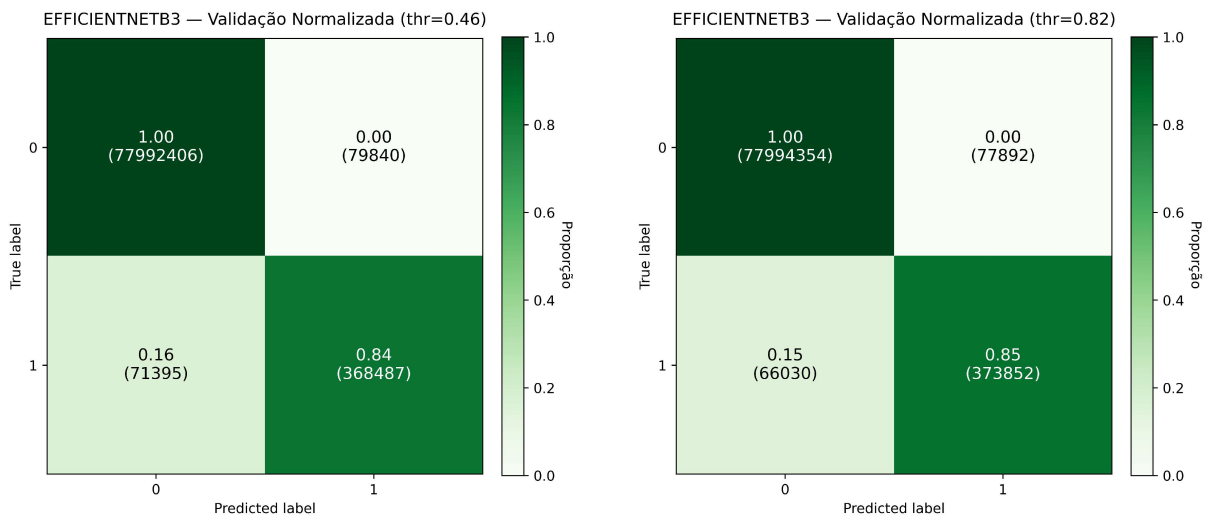
Em se tratando das matrizes de confusão da UNet com EfficientNetB3 para o conjunto de validação(Figura 15) revelam um comportamento distinto. Embora a classe “não fogo” também apresente proporção de verdadeiros negativos próxima de 1,0, a classe de fogo é menos favorecida que no VGG-16: a taxa de verdadeiros positivos varia em torno de 0,82–0,85, enquanto a fração de falsos negativos permanece entre 0,15 e 0,18. Isso significa que, mesmo após o *undersampling*, o EfficientNetB3 tende a deixar escapar uma proporção maior de *pixels* de fogo do que o VGG-16, o que se alinha com os valores de *F1-Score* e

IoU mais modestos observados nas tabelas de desempenho do Ciclo 2. Os limiares ótimos encontrados (por exemplo, $\text{thr} = 0,36$ para BCE, $0,70$ para Dice e $0,82$ para *DiceFocal*) são mais extremos do que aqueles do VGG-16, sugerindo que o EfficientNetB3 produz distribuições de probabilidade mais dispersas e exige ajustes mais agressivos no limiar para se aproximar de um compromisso adequado entre sensibilidade e *precision*:



BCE ($\text{thr}=0,36$) - alta capacidade de identificar fundo (poucos falsos alarmes relativos), mas ainda com perda perceptível de incêndios (falsos negativos moderados).

Dice ($\text{thr}=0,70$) - maior sensibilidade para incêndio entre as perdas, reduzindo omissões, com custo de alguns falsos positivos adicionais em relação ao melhor caso.



Focal ($\text{thr}=0,46$) - equilíbrio entre detectar incêndio e conter falsos positivos; comportamento estável e próximo ao BCE, com leve ganho na triagem.

DiceFocal ($\text{thr}=0,82$) - melhor compromisso geral; mantém alta detecção de incêndio com falsos positivos controlados, resultando na matriz mais “equilibrada”.

Figura 15 – Matrizes de confusão normalizadas da U-Net com *backbone* EfficientNetB3 no conjunto de validação do Ciclo 2, considerando as diferentes funções de perda.

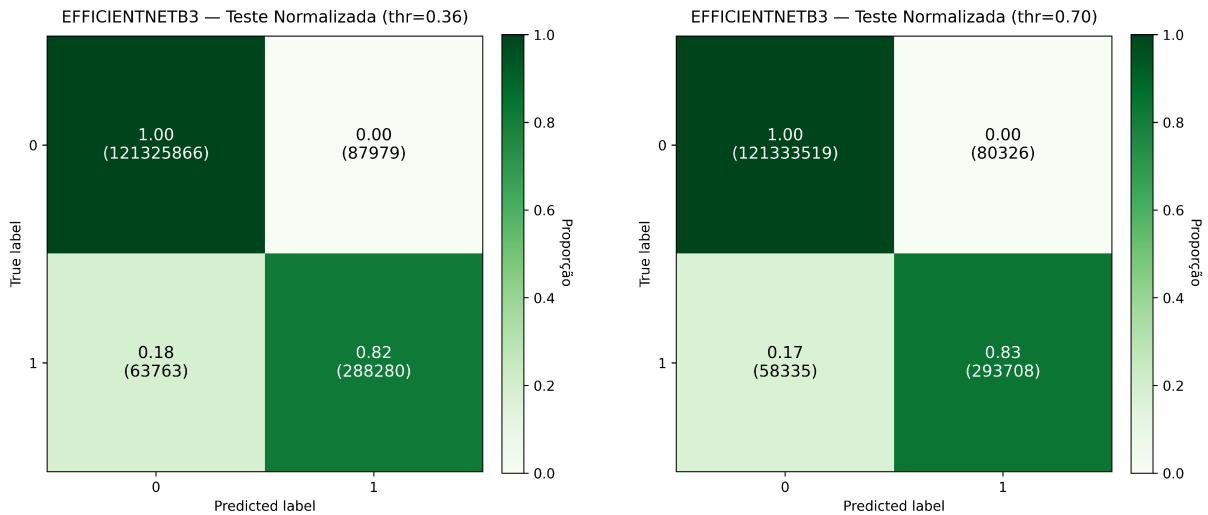
No conjunto de teste, as matrizes de confusão do EfficientNetB3 16 confirmam essa

tendência. A classe de fundo continua sendo bem representada, porém as taxas de verdadeiros positivos para fogo se mantêm abaixo das obtidas pelo VGG-16, com intervalos aproximados de 0,82–0,84 e frações correspondentes de falsos negativos em torno de 0,16–0,18. Outro ponto relevante é que, ao contrário do VGG-16, o EfficientNetB3 apresenta uma proporção visivelmente maior de falsos positivos (fundo rotulado como fogo), o que pode implicar maior taxa de alarmes falsos em cenários operacionais. Essa assimetria entre falsos negativos e falsos positivos é modulada pelos diferentes limiares ótimos — mais baixos para BCE e mais elevados para *Dice* e *DiceFocal* — refletindo o compromisso entre maximizar a detecção de áreas queimadas e controlar a supersegmentação sobre regiões de vegetação não queimada.

De forma integrada, a comparação das matrizes de confusão dos dois *backbones* no Ciclo 2 reforça a conclusão de que o VGG-16, especialmente em combinação com a função de perda *DiceFocal*, oferece o melhor compromisso entre redução de falsos negativos em fogo e manutenção de uma taxa quase nula de falsos positivos. O EfficientNetB3, por sua vez, apresenta um perfil mais arriscado, com maior sensibilidade a variações de limiar e tendência a produzir tanto mais falsos negativos quanto mais falsos positivos que o VGG-16. Em relação ao Ciclo 1, a visualização conjunta das matrizes evidencia que o uso de *undersampling* e de uma função de perda composta contribuiu para aumentar a separação entre classes, sobretudo no VGG-16, que passa a apresentar matrizes mais “limpas”, com diagonais mais marcadas e menor dispersão de erros. Esses resultados corroboram a escolha do *backbone* VGG-16 e da função *DiceFocal* como configuração de referência para o restante da dissertação.

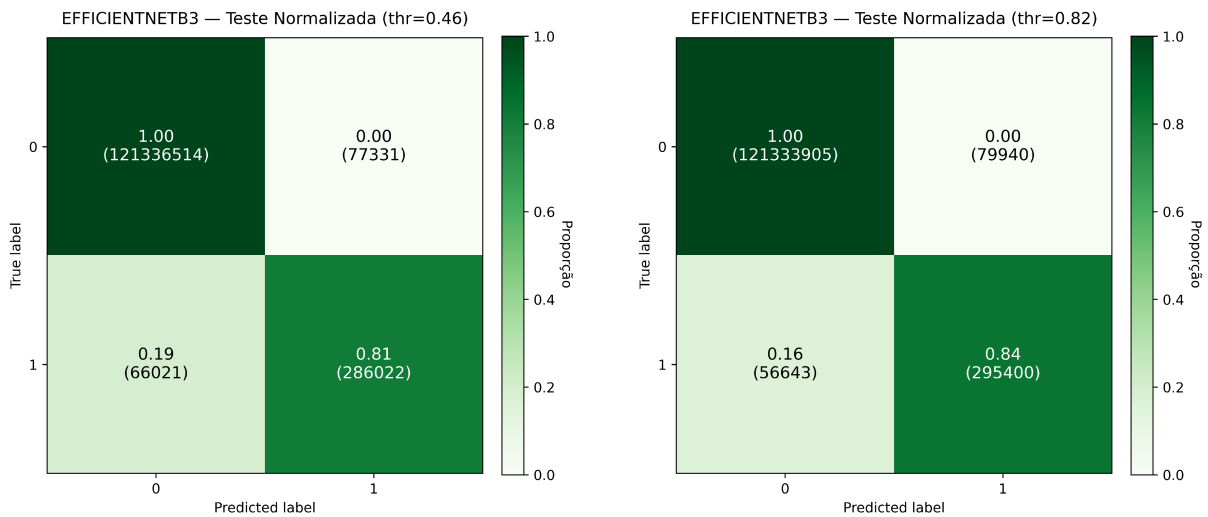
De maneira geral, observa-se que o VGG-16 supera consistentemente o EfficientNetB3 em todas as métricas avaliadas, tanto na validação quanto no teste. Esse comportamento sugere que a arquitetura mais simples e menos profunda do VGG-16 é particularmente adequada ao cenário de forte desbalanceamento e de baixa variabilidade espectral nas cicatrizes de incêndio de imagens Sentinel-2. Enquanto o EfficientNetB3 demonstra maior sensibilidade ao ajuste de limiar e tende a produzir mais erros tanto por omissão quanto por comissão, o VGG-16 apresenta um equilíbrio mais estável, com valores elevados de *F1* e IoU, mantendo uma taxa reduzida de falsos negativos.

A função *DiceFocal* mostrou-se a mais eficaz entre todas as avaliadas. No VGG-16, essa função de perda alcançou os maiores



BCE (thr=0,36): mantém o fundo praticamente intacto (poucos falsos positivos), mas ainda deixa escapar parte do incêndio (omissões moderadas).

Dice Loss (thr=0,70): ligeiro ganho na detecção de incêndio em relação ao BCE (menos omissões), preservando baixo nível de alarmes falsos.



Focal Loss (thr=0,46): perfil mais conservador; controla bem falsos positivos, porém com maior omissão de áreas de incêndio entre as perdas.

Dice+Focal (thr=0,82): melhor equilíbrio no teste; reduz as omissões e aumenta a detecção de incêndio, mantendo falsos positivos baixos.

Figura 16 – Matrizes de confusão normalizadas da U-Net com *backbone* EfficientNetB3 no conjunto de teste do Ciclo 2, considerando as diferentes funções de perda.

Conclusão

Os resultados apresentados ao longo deste capítulo evidenciam, de forma consistente, o impacto das escolhas metodológicas adotadas na tarefa de segmentação semântica de incêndios florestais em imagens multiespectrais do satélite Sentinel-2. A condução de dois ciclos experimentais permitiu não apenas avaliar o desempenho absoluto dos modelos, mas também compreender de maneira sistemática os fatores responsáveis pelas limitações observadas inicialmente e pelas melhorias obtidas posteriormente.

No Ciclo 1, de caráter exploratório, os experimentos revelaram dificuldades significativas na identificação da classe minoritária, especialmente em cenários de forte desbalanceamento entre classes. Embora a substituição da *Binary Cross-Entropy* pela *Dice Loss* tenha promovido ganhos pontuais, as métricas de IoU e *F1-Score* permaneceram limitadas, indicando que ajustes isolados na função de perda não eram suficientes para mitigar os efeitos do desbalanceamento e da baixa representatividade espacial do fogo nas imagens.

Em contraste, os resultados do Ciclo 2 demonstram de forma clara a eficácia do pipeline otimizado proposto neste trabalho. A introdução do *undersampling* no conjunto de treinamento, aliada ao uso de técnicas de *data augmentation* e à adoção de uma estratégia de treinamento em duas fases, resultou em melhorias expressivas tanto em desempenho quanto em estabilidade. A varredura sistemática de *thresholds* na validação, com seleção do limiar que maximiza o *F1-Score*, mostrou-se particularmente relevante para a calibração dos modelos, refletindo diretamente na qualidade das métricas reportadas no conjunto de teste.

Do ponto de vista arquitetural, os resultados indicam que a U-Net com *backbone* VGG-16 apresentou desempenho superior e mais estável em comparação à EfficientNetB3, especialmente no que se refere às métricas sensíveis ao desbalanceamento. Enquanto o VGG-16 manteve valores elevados e consistentes de *F1-Score* e IoU independentemente da função de perda utilizada, o EfficientNetB3 revelou maior sensibilidade à escolha da perda e do limiar de decisão, sugerindo maior instabilidade na separação entre as classes.

A análise comparativa das funções de perda reforça ainda a relevância de estratégias

sensíveis ao desbalanceamento. As perdas *Dice* e *DiceFocal* destacaram-se no Efficient-NetB3, enquanto no VGG-16 os resultados praticamente equivalentes entre as diferentes perdas evidenciam a robustez desse *backbone* frente às variações metodológicas avaliadas. Esse comportamento sugere que arquiteturas mais profundas e estáveis podem reduzir a dependência de ajustes finos na função de perda, embora tais ajustes permaneçam importantes em arquiteturas mais sensíveis.

Por fim, a análise qualitativa das máscaras preditas, apresentada ao final do capítulo, complementa os resultados quantitativos ao evidenciar melhorias não apenas métricas, mas também morfológicas. Observa-se maior coerência espacial das regiões segmentadas no Ciclo 2, com redução de falsos positivos dispersos e melhor delineamento dos focos de incêndio, corroborando as conclusões extraídas das matrizes de confusão e das métricas agregadas.

Em conjunto, os resultados obtidos confirmam que a combinação de balanceamento de classes, funções de perda adequadas, calibração criteriosa do limiar de decisão e estratégias de treinamento progressivo constitui um fator determinante para o sucesso da segmentação de incêndios florestais em imagens multiespectrais. Esses achados fornecem uma base sólida para as discussões e conclusões apresentadas no capítulo seguinte.

Como direções futuras, sugere-se investigar estratégias complementares como:

- ❑ Uso de *oversampling* com síntese de amostras (e.g., SMOTE adaptado para imagens) para enriquecer a classe minoritária;
- ❑ Aplicação de arquiteturas híbridas ou baseadas em atenção (e.g., *Attention*, U-Net, *Swin-UNet*);
- ❑ Testes em diferentes composições espectrais e resolução espacial;
- ❑ Avaliação em múltiplos sensores.

Esta dissertação contribui metodologicamente para o campo de visão computacional aplicada ao sensoriamento remoto ambiental, fornecendo subsídios para a construção de sistemas operacionais de detecção de queimadas com base em imagens orbitais multiespectrais.

5.1 Contribuições em Produção Bibliográfica

A partir dos resultados obtidos neste trabalho, foi elaborado e submetido o artigo intitulado Segmentação semântica de incêndios florestais em imagens do Sentinel-2 usando arquiteturas de aprendizado por transferência à revista científica Revista de Informática Teórica e Aplicada, encontrando-se atualmente em processo de avaliação.

O artigo apresenta uma análise comparativa entre diferentes funções de perda (*Binary Crossentropy* e *Dice Loss*) e a aplicação de *encoders* com transferência de aprendizado

(VGG-16 e EfficientNetb3), bem como os avanços alcançados com a adoção de técnicas como *undersampling*, *data augmentation* e função de perda composta. O trabalho destaca o impacto das decisões metodológicas sobre o desempenho dos modelos em tarefas de segmentação de regiões de incêndio, discutindo seus desdobramentos para aplicações operacionais em monitoramento ambiental e mitigação de desastres naturais.

Referências

ALPARONE, L.; GARZELLI, A.; ZOPPETTI, C. Fusion of vnir optical and c-band polarimetric sar satellite data for accurate detection of temporal changes in vegetated areas. **Remote Sensing**, v. 15, n. 3, 2023. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/15/3/638>>.

ANAYA, J.; TRIPATHY, P.; REINHARDT, M. Burned area detection using convolutional neural networks and sentinel-2 in manacacias natural park. In: **Anais da Space Week Nordeste 2023**. Fortaleza, CE, Brasil: Galoá, 2023. Disponível em: <<https://proceedings.science/swn-2023/trabalhos/burned-area-detection-using-convolutional-neural-networks-and-sentinel-2-in-mana?lang=en>>. Acesso em: 01 jan. 2026. Disponível em: <<https://proceedings.science/swn-2023/trabalhos/burned-area-detection-using-convolutional-neural-networks-and-sentinel-2-in-mana?lang=en>>.

BAN, Y. et al. Near real-time wildfire progression monitoring with sentinel-1 sar time series and deep learning. **Scientific Reports**, v. 10, n. 1, 2020. ISSN 2045-2322. Disponível em: <<https://doi.org/10.1038/s41598-019-56967-x>>.

BARMPOUTIS, P. et al. A review on early forest fire detection systems using optical remote sensing. **Sensors**, v. 20, n. 22, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/22/6442>>.

BOUGUETTAYA, A. et al. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. **Signal Processing**, v. 190, p. 108309, 2022. ISSN 0165-1684. Disponível em: <<https://doi.org/10.1016/j.sigpro.2021.108309>>.

BUMA, B. et al. Short-interval fires increasing in the alaskan boreal forest as fire self-regulation decays across forest types. **Scientific Reports**, v. 12, p. 4901, 03 2022. Disponível em: <<https://doi.org/10.1038/s41598-022-08912-8>>.

BUSLAEV, A. et al. Albumentations: Fast and flexible image augmentations. **Information**, MDPI AG, v. 11, n. 2, p. 125, fev. 2020. ISSN 2078-2489. Disponível em: <<https://doi.org/10.3390/info11020125>>.

CHAMBRELAN, A.; TEAM, S. M. **Sentinel-2 Level-1 Algorithm Theoretical Bases Document**. [S.l.], 2023. Disponível em: <<https://sentiwiki.copernicus.eu/>>.

attachments/1692737/S2-PDGS-MPC-ATBD-L1%20-%20Sentinel-2%20Level%201%20Algorithm%20Theoretical%20Bases%20Document%202023%20-%201.1.pdf?inst-v=2520e961-d5a2-4438-a386-65cfc671611c>.

CHEN, Y.; MORTON, D. C.; RANDERSON, J. T. Remote sensing for wildfire monitoring: Insights into burned area, emissions, and fire dynamics. **One Earth**, v. 7, n. 6, p. 1022–1028, 2024. ISSN 2590-3322. Disponível em: <<https://doi.org/10.1016/j.oneear.2024.05.014>>.

CHRISTIN, S.; HERVET, ; LECOMTE, N. Applications for deep learning in ecology. **Methods in Ecology and Evolution**, v. 10, n. 10, p. 1632–1644, 2019. Disponível em: <<https://doi.org/10.1111/2041-210X.13256>>.

CHUVIECO, E. et al. Historical background and current developments for mapping burned area from satellite Earth observation. **Remote Sensing of Environment**, v. 225, p. 45–64, 2019. ISSN 0034-4257. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2021.06.002>>.

de Almeida Pereira, G. H. et al. Active fire detection in landsat-8 imagery: A large-scale dataset and a deep-learning study. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 178, p. 171–186, 2021. ISSN 0924-2716. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S092427162100160X>>.

GOODMAN, R.; MILLER, J.; SMYTH, P. Objective functions for neural network classifier design. In: **Proceedings. 1991 IEEE International Symposium on Information Theory**. [s.n.], 1991. p. 87–87. Disponível em: <<https://doi.org/10.1109/ISIT.1991.695143>>.

HANES, C. C. et al. Fire-regime changes in canada over the last half century. **Canadian Journal of Forest Research**, v. 49, n. 3, p. 256–269, 2019. Disponível em: <<https://doi.org/10.1139/cjfr-2018-0293>>.

HE, H.; GARCIA, E. A. Learning from imbalanced data. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, n. 9, p. 1263–1284, 2009.

HOESER, T.; KUENZER, C. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. **Remote Sensing**, MDPI, v. 12, n. 10, 2020. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/12/10/1667>>.

KHELIFI, L.; MIGNOTTE, M. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. **IEEE Access**, v. 8, p. 126385–126400, 2020. Disponível em: <<https://doi.org/10.1109/ACCESS.2020.3009216>>.

_____. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. **IEEE Access**, v. 8, p. 126385–126400, 2020.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, maio 2017. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/3065386>>.

- KUMAR, S. S.; ROY, D. P. Global operational land imager landsat-8 reflectance-based active fire detection algorithm. **International Journal of Digital Earth**, v. 11, n. 2, p. 154–178, 2018. Disponível em: <<https://doi.org/10.1080/17538947.2017.1391341>>.
- LEE, D. et al. Single-temporal sentinel-2 for analyzing burned area detection methods: A study of 14 cases in republic of korea considering land cover. **Remote Sensing**, v. 16, n. 5, 2024. Disponível em: <<https://doi.org/10.3390/rs16050884>>.
- MACHADO, R. B. et al. **Aprendizado de máquina e uso de séries temporais de imagens Sentinel-2 no monitoramento da cobertura e uso da terra**. [S.l.], 2023.
- MapBiomass. **Queimadas atingem níveis alarmantes no Brasil em 2024**. 2024. Acesso em: 28 set. 2024. Disponível em: <<https://brasil.mapbiomas.org/2024/09/13/agosto-responde-por-quase-metade-da-area-queimada-no-brasil-em-2024/>>.
- MATSON, M.; HOLBEN, B. Satellite detection of tropical burning in brazil. **International Journal of Remote Sensing**, Taylor Francis, v. 8, n. 3, p. 509–516, 1987.
- MURPHY, S. W. et al. Hotmap: Global hot target detection at moderate spatial resolution. **Remote Sensing of Environment**, Elsevier, v. 177, p. 78–88, 2016. ISSN 0034-4257. Disponível em: <<https://doi.org/10.1016/j.rse.2016.02.027>>.
- OMER, M. et al. Visible near-infrared reflectance and laser-induced breakdown spectroscopy for estimating soil quality in arid and semiarid agroecosystems. **Soil Systems**, v. 4, 07 2020. Disponível em: <<https://doi.org/10.3390/soilsystems4030042>>.
- PEREIRA, E. et al. Mapping threatened canga ecosystems in the brazilian savanna using u-net deep learning segmentation and sentinel-2 images: a first step toward conservation planning. **Biota Neotropica**, v. 23, 03 2023. Disponível em: <<https://doi.org/10.1590/1676-0611-bn-2022-1384>>.
- PEREIRA, G. H. d. A. **Detecção de Incêndios em Imagens de Sensoriamento Remoto: Um Estudo Integrando os Satélites Landsat-8, Sentinel-2 e Inteligência Artificial**. Dissertação (Dissertação de Mestrado) — Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, 2021.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N. et al. (Ed.). **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4. Disponível em: <https://doi.org/10.1007/978-3-319-24574-4_28>.
- ROSTAMI, A. et al. Active fire detection from landsat-8 imagery using deep multiple kernel learning. **Remote Sensing**, v. 14, n. 4, p. 992, 2022. Disponível em: <<https://doi.org/10.3390/rs14040992>>.
- SANTOS, D. F. D. dos et al. Influence of data augmentation strategies on the segmentation of oral histological images using fully convolutional neural networks. **Journal of Digital Imaging**, Springer Science and Business Media LLC, apr 2023.
- SATHISHKUMAR, V. E. et al. Forest fire and smoke detection using deep learning-based learning without forgetting. **Fire Ecology**, SpringerOpen, v. 19, n. 9, 2023. Disponível em: <<https://doi.org/10.1186/s42408-022-00165-0>>.

SCHROEDER, W. et al. The new viirs 375m active fire detection data product: Algorithm description and initial assessment. **Remote Sensing of Environment**, v. 143, p. 85–96, 2014. ISSN 0034-4257. Disponível em: <<https://doi.org/10.1016/j.rse.2013.12.008>>.

_____. Active fire detection using landsat-8/oli data. **Remote Sensing of Environment**, v. 185, p. 210–220, 2016. ISSN 0034-4257. Landsat 8 Science Results. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0034425715301206>>.

SHIRVANI, Z.; ABDI, O.; GOODMAN, R. C. High-resolution semantic segmentation of woodland fires using residual attention unet and time series of sentinel-2. **Remote Sensing**, v. 15, n. 5, 2023. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/15/5/1342>>.

SOUNDARARAJAN, J. et al. Deep learning-driven multi-temporal detection: Leveraging deeplabv3+/efficientnet-b08 semantic segmentation for deforestation and forest fire detection. **Remote Sensing**, v. 17, n. 14, 2025. ISSN 2072-4292. Disponível em: <<https://doi.org/10.3390/rs17142333>>.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. 2020. Disponível em: <<https://arxiv.org/abs/1905.11946>>.

UNEP, U. N. E. P. **Spreading like wildfire—The rising threat of extraordinary landscape fires**. 2022. Disponível em: <<https://www.unep.org/resources/report/spreading-wildfire-rising-threat-extraordinary-landscape-fires>>.

WAGNER, F. H. et al. Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. **Remote Sensing in Ecology and Conservation**, v. 5, n. 4, p. 360–375, 2019. Disponível em: <<https://doi.org/10.1002/rse2.111>>.

XIONG, J. et al. A review of plant phenotypic image recognition technology based on deep learning. **Electronics**, MDPI, v. 10, n. 1, p. 81, 2021. Disponível em: <<https://doi.org/10.3390/electronics10010081>>.

ÖZEL, B.; ALAM, M.; KHAN, U. Review of modern forest fire detection techniques: Innovations in image processing and deep learning. **Information**, v. 15, p. 538, 09 2024. Disponível em: <<https://doi.org/10.3390/info15090538>>.