

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

ANDRÉ MINORO FUSIOKA

**DETECÇÃO DE FOGO ATIVO POR APRENDIZADO
PROFUNDO EM IMAGENS PROVENIENTES DO SATÉLITE
LANDSAT-8**

DISSERTAÇÃO

CURITIBA

2021

ANDRÉ MINORO FUSIOKA

**DETECÇÃO DE FOGO ATIVO POR APRENDIZADO
PROFUNDO EM IMAGENS PROVENIENTES DO SATÉLITE
LANDSAT-8**

**Active fire detection with Deep-Learning
in Landsat-8 imagery**

Dissertação apresentada como requisito para
obtenção do título de Mestre em Computa-
ção Aplicada da Universidade Tecnológica
Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Rodrigo Minetto
Coorientador: Prof. Dr. Bogdan Tomoyuki
Nassu

CURITIBA

2021



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es).

Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Curitiba



ANDRE MINORO FUSIOKA

DETECÇÃO DE FOGO ATIVO POR APRENDIZADO PROFUNDO EM IMAGENS PROVENIENTES DO SATÉLITE LANDSAT-8

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Computação Aplicada da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Engenharia De Sistemas Computacionais.

Data de aprovação: 07 de Outubro de 2021

Prof Rodrigo Minetto, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Bogdan Tomoyuki Nassu, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Fabio Augusto Faria, Doutorado - Universidade Federal de São Paulo (Unifesp)

Prof.a Leyza Elmeri Baldo Dorini, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Ricardo Dutra Da Silva, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 07/10/2021.

Dedico este trabalho a minha família e aos meus
amigos, pelos momentos de ausência.

AGRADECIMENTOS

Agradeço a minha família, por sempre estar presente, pelo carinho, incentivo e total apoio em todos os momentos da minha vida, principalmente durante esta jornada.

Aos meus orientadores, que mostraram os caminhos a serem seguidos, ensinaram e auxiliaram sempre que necessário.

A todos os professores e amigos, que ajudaram de forma direta e indireta na realização deste trabalho.

RESUMO

A detecção de fogo ativo em imagens de satélite é de grande importância para o desenvolvimento de políticas de conservação ambiental, fomentar a tomada de decisão e a aplicação da lei. As técnicas tradicionais de detecção de fogo ativo geralmente são baseadas na comparação de pixels, ou regiões da imagem, com limiares específicos para o sensor utilizado. O presente trabalho aborda a detecção de fogo ativo fazendo uso de técnicas de aprendizado profundo, do inglês *deep learning*. Nos últimos anos, técnicas de *deep learning* têm desfrutado de um enorme sucesso em diversos campos, mas o seu uso para detecção de fogo ativo é relativamente novo, assim, é necessário analisar a viabilidade do uso de aprendizado profundo para detecção de fogo ativo. O presente estudo analisa como diferentes arquiteturas de redes neurais convolucionais podem ser utilizadas para segmentar fogo ativo em imagens de satélite, utilizando como treinamento as máscaras produzidas por métodos conhecidos na literatura e comumente empregados para a detecção de fogo ativo, bem como a combinação de tais máscaras. Em específico, o estudo compreende imagens do satélite Landsat-8 no período de Agosto de 2020. Além disso, os modelos estudados foram avaliados sobre um conjunto de imagens do mesmo satélite, obtidas em Setembro de 2020, mas anotadas manualmente, com o objetivo de analisar a capacidade das redes em aproximar anotações de um especialista humano. De forma geral, as arquiteturas testadas foram capazes de aproximar os métodos clássicos da literatura, sendo que o melhor desempenho observado obteve uma métrica F -score de 94.2% e IoU de 89.0%. Quando comparadas com as anotações manuais, as redes obtiveram resultados superiores aos métodos clássicos da literatura na grande maioria dos casos, obtendo um F -score de 89.7% e IoU de 81.4%. O código desenvolvido, bem como os pesos treinados, foram disponibilizados em um projeto de código aberto no Github, disponível em: <https://github.com/pereira-gha/activefire>, possibilitando outros estudos na área.

Palavras-chave: Detecção de fogo ativo. Segmentação de fogo ativo. Redes Neurais Convolucionais. Landsat-8.

ABSTRACT

Active fire detection in satellite imagery is of great importance for the development of environmental conservation policies, supporting decision-making and law enforcement. Active fire detection techniques are generally based on comparing pixels or image regions with specific thresholds for the used sensor. This work addresses active fire detection using deep learning techniques. In recent years, deep learning techniques have enjoyed enormous success in various fields, but their use for active fire detection is relatively new, so it is necessary to analyze the feasibility of using deep learning for active fire detection. This study evaluates how different convolutional neural network architectures can be used to segment active fire in satellite images, using the masks produced by methods commonly used to detect active fire as training samples, as well as the combination of such masks, evaluating images from the Landsat-8 satellite from the period of August 2020. In addition, the studied models are evaluated on images from the same satellite obtained in September 2020, but manually annotated, verifying the capacity of the networks to approximate annotations made by a human specialist. The tested architectures were able to approximate the classical methods of the literature, and the best overall observed performance obtained an F -score metric of 94.2% and an IoU of 89.0%. When compared with the manual annotations, the networks obtained results superior to the classic methods of the literature in the vast majority of cases, obtaining an F -score of 89.7% and IoU of 81.4%. The developed code as well as the trained weights were made available in an open source project on GitHub, available at: <https://github.com/pereira-gha/activefire>, enabling other studies in the field.

Keywords: Active fire detection. Active fire segmentation. Convolutional Neural Network. Landsat-8.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Imagens disponíveis no <i>dataset</i> , utilizando as bandas 7, 6 e 2, e as máscaras produzidas pelos diferentes algoritmos.	22
Figura 2 –	Exemplo de anotação manual: a esquerda, uma imagem da Califórnia (Estados Unidos da América), capturada pelo Landsat-8 em 8 de Setembro de 2020. A imagem possui muitas regiões de fogo ativo, totalizando 55,386 pixels de fogo ativo identificados. Destacada em vermelho uma das duas maiores áreas de fogo ativo, com uma extensão de 22×22 km, com aproximadamente 10,000 pixels de fogo ativo; a marcação manual correspondente pode ser vista no canto inferior direito da figura.	23
Figura 3 –	Arquitetura da rede U-Net utilizada para segmentação de imagens.	25
Figura 4 –	Exemplo de combinações distintas de canais do Landsat-8: à esquerda a composição das bandas 4 (vermelho), 3 (verde) e 2 (azul); à direita a combinação de bandas 7 (Infravermelho de Ondas Curtas, do inglês <i>Short-wave Infrared</i> (SWIR)), 6 (SWIR) e 2 (azul), como utilizado para treinar as CNNs com três canais. Note que apenas a fumaça é visível na imagem a esquerda, enquanto o fogo aparece visivelmente na imagem da direita.	26
Figura 5 –	Diferentes situações para treino e teste das redes neurais convolucionais. Cada um das três arquiteturas (U-Net (10c), U-Net (3c) and U-Net-Light (3c)) foi treinada e testada para aproximar a saída de cinco conjuntos de condições distintos: os algoritmos propostos por Schroeder <i>et al.</i> (2016), Murphy <i>et al.</i> (2016) e Kumar e Roy (2018), como podem ser visto nas figuras (a), (b) e (c), bem como a intersecção e a votação (2 votos), como visto em (d).	27
Figura 6 –	Função de perda em relação às épocas de treino para a U-Net (3c), utilizando a intersecção dos três algoritmos como máscara de treino. Os demais treinos possuem um comportamento similar. Após aproximadamente 10 épocas há pequenas variações na função de perda, que podem ser difíceis de visualizar devido a escala do eixo <i>y</i> , que foi gerado automaticamente com a biblioteca utilizada.	28
Figura 7 –	Resultados da segmentação de fogo ativo obtidos pelo algoritmo de Schroeder <i>et al.</i> (2016), e pelas redes treinadas utilizando a saída desse método como máscara desejada. Em vermelho, detecções geradas pelas redes, mas não geradas pelas máscaras originais (falsos positivos).	33
Figura 8 –	Resultados da segmentação de fogo ativo obtidos pelo algoritmo de Murphy <i>et al.</i> (2016), e pelas redes treinadas utilizando a saída desse método como máscara desejada. Em vermelho, detecções geradas pelas redes, mas não geradas pelas máscaras originais (falsos positivos).	33
Figura 9 –	Resultados da segmentação de fogo ativo obtidos pelo algoritmo de Kumar e Roy (2018), e pelas redes treinadas utilizando a saída desse método como máscara desejada. Em vermelho, detecções geradas pelas redes, mas não geradas pelas máscaras originais (falsos positivos).	34
Figura 10 –	Os algoritmos propostos por Schroeder <i>et al.</i> (2016), Murphy <i>et al.</i> (2016) e Kumar e Roy (2018) geram detecções falsas para ambientes urbanos, mas a intersecção e a U-Net de três canais treinada para aproximá-la foram capazes de evitar tais detecções.	34

Figura 11 – A U-Net de três canais foi capaz de aproximar as máscaras de votação geradas pela combinação das saídas produzidas pelos algoritmos de Schroeder <i>et al.</i> (2016), Murphy <i>et al.</i> (2016) e Kumar e Roy (2018). O esquema de votação reduz erros causados por artefatos produzidos por apenas um dos algoritmos. Em verde, na máscara de votação, as detecções geradas por apenas dois dos três métodos. Em vermelho, na máscara gerada pela rede U-Net (3c), as detecções falsas detectadas em relação a máscara de votação.	35
Figura 12 – Resultados da segmentação de fogo ativo produzidos pelos conjuntos de condições de Schroeder <i>et al.</i> (2016), Murphy <i>et al.</i> (2016) e Kumar e Roy (2018), a combinação desses métodos (intersecção e votação) e as redes profundas treinadas com cada abordagem para um recorte de uma imagem do Landsat-8. Em magenta, na linha das máscaras dos algoritmos, falsos positivos gerados pelos algoritmos quando comparados com a marcação manual. Em verde, as detecções feitas pelas redes em concordância com as anotações do especialista e que não foram capturadas pela máscara da respectiva abordagem. Em vermelho, as detecções feitas pelas redes e não anotadas pelo especialista nem identificadas pelas máscara da respectiva abordagem.	37
Figura 13 – Ampliação em uma região da imagem, bem como os resultados produzidos pelos métodos da literatura e pelas CNNs para segmentação de fogo ativo. Em magenta, na linha das máscaras dos algoritmos handcrafted tradicionais na literatura, falsos positivos gerados pelos algoritmos quando comparados com a marcação manual. Em verde, as detecções feitas pelas redes em concordância com as anotações do especialista e que não foram capturadas pelos algoritmos originais. Em vermelho, as detecções feitas pelas redes e não anotadas pelo especialista nem identificadas pelos algoritmos originais. . . .	39
Figura 14 – Erros de classificação: (a) região do deserto do Sahara, WRS: 189/046, data: 2020/09/18; (b) e (c) região da Groenlândia, WRS: 003/006, data: 2020/09/11; e (d) cidade de Milão, Itália, WRS: 193/029, data: 2020/09/14, onde todas as condições originais e também as arquiteturas CNN reportaram erros de classificação em um ambiente urbano.	40

LISTA DE TABELAS

Tabela 1 – Bandas disponíveis para o satélite Landsat-8, com as respectivas resoluções espectrais (em μm) e espaciais (em metros).	17
Tabela 2 – Arquiteturas baseadas em uma rede U-Net para segmentação de fogo ativo. .	29
Tabela 3 – Desempenho das CNNs para reconhecimento de fogo ativo utilizando conjunto de condições da literatura	32
Tabela 4 – Desempenho das CNNs para reconhecimento de fogo ativo em relação as anotações manuais	36

LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

SIGLAS

AVHRR	<i>Advanced Very-High-Resolution Radiometer</i>
AWS	<i>Amazon Web Service</i>
EOSDIS	<i>Earth Observing System Data and Information System</i>
GOES	<i>Geostationary Operational Environmental Satellite</i>
IoU	Intersecção Sobre União, do inglês <i>Intersection Over Union</i>
MODIS	<i>Moderate-Resolution Imaging Spectroradiometer</i>
SWIR	Infravermelho de Ondas Curtas, do inglês <i>Short-wave Infrared</i>
TRMM	<i>Tropical Rainfall Measuring Mission</i>
VIIRS	<i>Visible Infrared Imaging Radiometer Suite</i>
VIRS	<i>Visible and Infrared Scanner</i>

ACRÔNIMOS

COCO	<i>Common Objects in Context</i>
KITTI	<i>Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago</i>
VOC	<i>Visual Object Challenge</i>

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO DA LITERATURA	15
2.1	MÉTODOS DE LIMIAÇÃO PARA DETECÇÃO DE FOGO ATIVO	15
2.2	ALGORITMOS TRADICIONAIS DE DETECÇÃO DE FOGO ATIVO DO LANDSAT-8	16
2.2.1	Método de SCHROEDER <i>et al.</i>	17
2.2.2	Método de MURPHY <i>et al.</i>	18
2.2.3	Método de Kumar e Roy (2018)	18
2.3	MÉTODOS DE APRENDIZADO PROFUNDO PARA DETECÇÃO DE FOGO ATIVO	19
3	MATERIAL E MÉTODOS	21
3.1	DATASET	21
3.2	REDES NEURAS CONVOLUCIONAIS PARA DETECÇÃO DE FOGO ATIVO	23
3.3	IMPLEMENTAÇÃO E AMBIENTE DE DESENVOLVIMENTO . . .	27
3.4	MÉTRICAS DE AVALIAÇÃO	29
4	RESULTADOS E DISCUSSÃO	31
4.1	DESEMPENHO NAS IMAGENS SEGMENTADAS AUTOMATICA- MENTE	31
4.2	DESEMPENHO NAS IMAGENS ANOTADAS MANUALMENTE . .	35
4.3	DISCUSSÃO	41
5	CONCLUSÕES	43
	REFERÊNCIAS	44

1 INTRODUÇÃO

O fogo faz parte do ciclo ecológico de alguns biomas, favorecendo a manutenção da vegetação local, além disso, também pode ser utilizado de maneira controlada para eliminar matéria orgânica acumulada, reduzindo a presença de possíveis combustíveis para incêndios de grande escala e com isso evitando queimadas descontroladas (CERTINI, 2005). Apesar disso, o uso indiscriminado do fogo pode acarretar em mudanças no solo, aumento na emissão de carbono e alterações no ciclo da água da região afetada (ARAÚJO *et al.*, 2012). Em alguns biomas o fogo é um dos principais responsáveis pela degradação do ambiente, sendo empregado para abrir espaço para a exploração agrícola e pecuária (LIZUNDIA-LOIOLA *et al.*, 2020).

A detecção de fogo ativo é um campo de pesquisa que extrai informações relevantes de imagens para dar suporte à tomada de decisão (CARDIL *et al.*, 2019), de modo a localizar e mensurar a extensão do evento, além de permitir estimativas sobre os danos causados. Nesse sentido, a utilização de imagens de satélite mostra-se como uma ferramenta de grande auxílio para observar e dimensionar o impacto de incêndios, de modo a cobrir uma grande área em uma única imagem, o que permite monitorar até mesmo regiões de difícil acesso físico.

Os satélites fornecem imagens com diferentes resoluções temporais, espectrais e espaciais de diversas regiões do globo, tornando-se uma grande fonte de informação para observações da superfície. Com estas imagens, é possível observar regiões que passaram por uma catástrofe, muitas vezes fornecendo as primeiras informações visuais sobre o problema (DUARTE *et al.*, 2018), e em casos de regiões remotas pode ser a única fonte de informação.

Pesquisas na área de detecção de incêndios, têm dentre algum dos objetivos, o desenvolvimento de métodos apropriados para determinar se um dado pixel — correspondente a uma área georreferenciada em uma imagem multiespectral — corresponde a fogo ativo. Como os satélites orbitam em diferentes altitudes e possuem sensores distintos, com diferentes comprimentos de ondas, normalmente esses métodos precisam ser desenvolvidos especificamente para cada satélite. Por exemplo, imagens provenientes do satélite Landsat-8, foco deste trabalho, são normalmente processadas por uma série de condições (limiares) conforme proposto por Schroeder *et al.* (2016), Murphy *et al.* (2016) ou Kumar e Roy (2018). Estas condições, descritas na Seção 2.2, são ajustadas de modo a encontrar os melhores parâmetros para o sensor a bordo do Landsat-8 e, de modo geral, baseiam-se em comparações de limiares fixos em certas bandas e estatísticas de sua região circundante.

Nos últimos anos métodos de *deep learning* (LECUN *et al.*, 2015) têm promovido grandes avanços em inteligência artificial, inspirando o design de novas arquiteturas e estratégias para um grande número de aplicações de sensoriamento remoto. Redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN), são destaque em *deep learning* e têm como característica a busca por padrões intrínsecos em uma grande quantidade de dados ao usar uma sequência de camadas de processamento, o que permite aprender representações de dados em vários níveis de abstração, dispensando a definição explícita de características para classificação de regiões de interesse.

Desse modo, o presente trabalho realiza um estudo sobre como redes neurais convolucionais, mais precisamente variações de uma rede U-Net (RONNEBERGER *et al.*, 2015), podem ser utilizadas para aproximar máscaras de segmentação de fogo ativo geradas por algoritmos tradicionais para este problema (Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018)), bem como a combinação destes algoritmos, para produzir resultados mais robustos. Também é verificada a possibilidade de redução do número de bandas espectrais para geração das máscaras, de modo que durante o download das imagens possam ser selecionadas apenas as bandas necessárias para produção da saída (economia de recursos). Além disso, é realizado um estudo comparativo entre os algoritmos tradicionais e redes convolucionais, para o problema da detecção de fogo ativo, em imagens anotadas manualmente por um especialista da área.

Assim, o principal foco do presente estudo não é o projeto de uma nova arquitetura CNN otimizada para detecção de fogo ativo, mas sim: 1) mostrar que uma CNN treinada com os devidos dados é capaz de produzir resultados similares aos métodos convencionais encontrados na literatura (descritos na Seção 2.2); 2) melhorar as respostas de detecção em regiões onde eles falham; e ainda, 3) mostrar que é possível modelar com apenas uma rede CNN combinações de vários algoritmos, por exemplo, modelar a intersecção de respostas de algoritmos tradicionais para gerar saídas que indiscutivelmente são grandes candidatas a regiões de fogo ativo em uma imagem, e modelar a resposta de votação majoritária entre eles, para produzir detecções que não são nem restritivas nem permissivas demais.

Todo o código desenvolvido no presente trabalho é disponibilizado como parte de um projeto de código aberto no GitHub¹, em linguagem Python, para que outros pesquisadores possam utilizá-lo como referência, de modo a criar oportunidades para o estudo de imagens do Landsat-8 coletadas em outros anos ou em diferentes épocas, ou ainda, para o uso de diferentes

¹ <https://github.com/pereira-gha/activefire>

satélites. Este é um primeiro esforço nessa direção, visto que os algoritmos tradicionais que foram analisados aqui, até onde se sabe, não têm versões de código aberto disponíveis.

Parte das discussões e resultados aqui apresentados foram aceitos para publicação em um conceituado periódico da área:

- **Active Fire Detection in Landsat-8 Imagery: a Large-Scale Dataset and a Deep-Learning Study.** *Gabriel Henrique de Almeida Pereira, André Minoro Fusioka, Bogdan Tomoyuki Nassu e Rodrigo Minetto.* ISPRS Journal of Photogrammetry and Remote Sensing, Elsevier, 2021 (fator de impacto: 8.9).

Esta dissertação é dividida da seguinte forma: o Capítulo 2 apresenta trabalhos relacionados à detecção de fogo ativo e ao emprego de *deep learning* para tal fim. O Capítulo 3 apresenta os materiais e métodos utilizados no trabalho. O Capítulo 4 descreve os experimentos, resultados obtidos e a discussão dos resultados. O Capítulo 5 contém as conclusões e trabalhos futuros.

2 REVISÃO DA LITERATURA

Neste capítulo são discutidos os principais trabalhos da literatura que abordam a detecção de fogo ativo por meio de imagens de satélite. A Seção 2.1 aborda os métodos tradicionais de detecção de fogo que recorrem ao uso de limiares sobre as bandas de diferentes satélites. A Seção 2.2 apresenta os métodos comumente empregados para detecção de fogo ativo para o satélite Landsat-8 — estes métodos foram empregados para a geração das máscaras utilizadas para treinamento dos modelos de aprendizado profundo do presente trabalho. A Seção 2.3 aborda os trabalhos que fazem uso de aprendizado profundo para a detecção de fogo ativo.

2.1 MÉTODOS DE LIMIAZIZAÇÃO PARA DETECÇÃO DE FOGO ATIVO

Estudos sobre métodos automáticos de detecção de regiões ativas de fogo datam da década de 1970, quando foram analisados os intervalos espectrais mais adequados para detecção de incêndios florestais (KONDRATYEV *et al.*, 1972). Matson e Holben (1987) argumentam que o sensoriamento remoto é a única alternativa viável de monitorar fogo ativo em regiões isoladas. No estudo realizado pelos autores, eles empregaram o sensor *Advanced Very-High-Resolution Radiometer* (AVHRR) para detectar fogo ativo na região amazônica. Outros algoritmos de detecção de fogo ativo também foram desenvolvidas para o AVHRR (FLANNIGAN; HAAR, 1986; LEE; TAG, 1990), sendo alguns deles utilizados como base para o desenvolvimento de algoritmos para sensores de outros satélites. Ji e Stocker (2002) e Giglio *et al.* (2003) desenvolveram algoritmos que incluem análise contextual para detecção de fogo ativo para o sensor *Visible and Infrared Scanner* (VIRS) embarcado no satélite *Tropical Rainfall Measuring Mission* (TRMM).

Por sua vez, Kaufman *et al.* (1998) propuseram uma abordagem para detecção de fogo ativo utilizando o sensor *Moderate-Resolution Imaging Spectroradiometer* (MODIS), que posteriormente foi melhorado por Giglio *et al.* (2003). Esse método continua sendo empregado nos dias atuais, dada a sua relevância e também porque o sensor MODIS continua ativo e possui uma cobertura global quase diária, sendo capaz de detectar focos de incêndio com um tamanho entre 100 e 300 m² (MAIER *et al.*, 2013). Além disso, o algoritmo de detecção do MODIS foi melhorado de diferentes modos (MORISSETTE *et al.*, 2005; GIGLIO *et al.*, 2016) e serviu de referência para o desenvolvimento de equações para outros satélites. Por exemplo, em 2014,

Schroeder *et al.* (2014) propuseram um algoritmo para detecção de fogo ativo para o sensor *Visible Infrared Imaging Radiometer Suite* (VIIRS) baseado no algoritmo utilizado no MODIS.

Esses métodos tradicionais de detecção de fogo ativo recorrem ao uso de limiares sobre os valores das bandas, identificando os pixels que claramente representam fogo e possíveis corpos quentes e, em alguns deles, realizando uma análise contextual na vizinhança de cada pixel em potencial para determinar se o mesmo representa fogo ou não (SCHROEDER *et al.*, 2016). Visando reduzir os alarmes falsos de fontes de calor constantes, como por exemplo fábricas, refinarias, e telhados, alguns métodos de aplicação de limiar fazem uma análise temporal observando a persistência do foco de calor no decorrer do tempo (KUMAR; ROY, 2018; SCHROEDER *et al.*, 2016). Apesar dos bons resultados produzidos, analisar a vizinhança de cada pixel que potencialmente representa fogo em imagens no decorrer do tempo é uma tarefa computacionalmente cara, sendo indicada para pequenos *datasets* (MURPHY *et al.*, 2016).

2.2 ALGORITMOS TRADICIONAIS DE DETECÇÃO DE FOGO ATIVO DO LANDSAT-8

Esta seção apresenta uma breve descrição sobre os métodos de detecção de fogo ativo comumente utilizados para o satélite Landsat-8. Serão descritos os algoritmos de Schroeder *et al.* (2016) (Seção 2.2.1), Murphy *et al.* (2016) (Seção 2.2.2) e Kumar e Roy (2018) (Seção 2.2.3). Esses algoritmos foram empregados por Pereira *et al.* (2021) para a construção do *dataset* utilizado pelo presente trabalho para treinar e avaliar as redes convolucionais para detecção de fogo ativo. Para um entendimento mais profundo dos métodos recomenda-se ao leitor consultar os trabalhos originais.

Os métodos descritos recorrem à aplicação de limiares sobre a reflectância das bandas do satélite Landsat-8. A Tabela 1 exibe uma listagem das bandas disponíveis. A conversão do valor capturado pelo sensor do Landsat para reflectância pode ser feita com auxílio dos metadados específicos, disponíveis para cada imagem do satélite (USGS, s.d.).

No restante da dissertação será utilizada a notação ρ_i para representar o valor de reflectância para o canal c_i , e a notação R_{ij} para representar a razão entre a reflectância do canal i e j , ou seja, $R_{ij} = \rho_i / \rho_j$.

Tabela 1 – Bandas disponíveis para o satélite Landsat-8, com as respectivas resoluções espectrais (em μm) e espaciais (em metros).

# Banda Espectral	Resolução Espectral (μm)	Resolução Espacial (m)
1 - Costal	0.433 — 0.453	30
2 - Azul	0.450 — 0.515	30
3 - Verde	0.525 — 0.600	30
4 - Vermelho	0.630 — 0.680	30
5 - Infravermelho próximo	0.845 — 0.885	30
6 - Infravermelho de ondas curtas	1.560 — 1.660	30
7 - Infravermelho de ondas curtas	2.100 — 2.300	30
8 - Pancromático	0.500 — 0.680	15
9 - Cirrus	1.360 — 1.390	30
10 - Infravermelho de ondas longas	10.6 — 11.2	100
11 - Infravermelho de ondas longas	11.5 — 12.5	100

Fonte: Adaptado de Loyd (2013).

2.2.1 Método de SCHROEDER *et al.*

O método proposto por Schroeder *et al.* (2016) aplica um conjunto de condições sobre as sete primeiras bandas do Landsat-8. O primeiro conjunto de limiares visa detectar casos em que o dado pixel inequivocamente representa fogo ativo:

$$((R_{75} > 2.5) \text{ E } (\rho_7 - \rho_5 > 0.3) \text{ E } (\rho_7 > 0.5)) \quad (1)$$

$$\text{OU } ((\rho_6 > 0.8) \text{ E } (\rho_1 < 0.2) \text{ E } (\rho_5 > 0.4 \text{ OU } \rho_7 < 0.1)).$$

Após o primeiro conjunto de limiares, o método identifica pixels que potencialmente representam fogo através da seguinte condição:

$$(R_{75} > 1.8) \text{ E } (\rho_7 - \rho_5 > 0.17). \quad (2)$$

Os pixels identificados pela condição (2) são avaliados novamente para determinar se representam fogo. É computada a média (μ) e desvio padrão (σ) de uma janela de 61×61 pixels centrada em cada pixel candidato identificado e então é aplicado um grupo de limiares com auxílio das informações estatísticas calculadas para capturar focos de fogo não observados anteriormente:

$$(R_{76} > 1.6) \text{ E } (R_{75} > \mu_{R_{75}} + \max(3\sigma_{R_{75}}, 0.8)) \text{ E } (\rho_7 > \mu_{\rho_7} + \max(3\sigma_{\rho_7}, 0.08)). \quad (3)$$

A média e desvio padrão são computados desconsiderando os pixels de fogo e água, tal que os pixels de água são identificados pelo seguinte conjunto de limiares:

$$(\rho_4 > \rho_5) \text{ E } (\rho_5 > \rho_6) \text{ E } (\rho_6 > \rho_7) \text{ E } (\rho_1 - \rho_7 < 0.2) \quad (4)$$

$$\text{E } ((\rho_3 > \rho_2) \text{ OU } ((\rho_1 > \rho_2) \text{ E } (\rho_2 > \rho_3) \text{ E } (\rho_3 > \rho_4))).$$

Vale ressaltar que o método originalmente proposto analisa imagens de um mesmo local em períodos distintos de tempo, com o intuito de reduzir falsos positivos causados por fontes constantes de calor, como chaminés e telhados. Essa análise temporal não foi aplicada na construção do *dataset* utilizado neste trabalho (PEREIRA *et al.*, 2021) e portanto não será discutida nesta seção.

2.2.2 Método de MURPHY *et al.*

O método proposto por Murphy *et al.* (2016) se baseia em conjuntos de limiares sobre as bandas 5, 6 e 7. Inicialmente o método identifica pixels que inequivocamente representam fogo através da seguinte verificação:

$$(R_{76} \geq 1.4) \text{ E } (R_{75} \geq 1.4) \text{ E } (\rho_7 \geq 0.15). \quad (5)$$

Então uma vizinhança de 3×3 pixels é utilizada sobre cada pixel detectado como fogo na equação (5), pixels nessa vizinhança também são assinalados como fogo em potencial se satisfizerem a seguinte condição:

$$((R_{65} \geq 2) \text{ E } (\rho_6 \geq 0.5)) \text{ OU } S. \quad (6)$$

O termo S da condição anterior assume o valor verdadeiro quando a banda 6 ou 7 está saturada. A identificação da saturação de um pixel pode ser feita através de dados auxiliares disponibilizados para cada imagem do Landsat-8 (USGS, s.d.).

2.2.3 Método de Kumar e Roy (2018)

O método proposto por Kumar e Roy (2018) inicialmente avalia a ocorrência inequívoca de fogo ativo para cada pixel presente na imagem através da condição:

$$\rho_4 \leq 0.53 \times \rho_7 - 0.214. \quad (7)$$

Pixels vizinhos ao detectado pela condição anterior são definidos como fogo caso satisfaçam a seguinte condição:

$$\rho_4 \leq 0.35 \times \rho_6 - 0.044. \quad (8)$$

Então, o método avalia pixels que potencialmente representam fogo ativo. Os pixels que satisfazem a seguinte condição são identificados como tal:

$$(\rho_4 \leq 0.53 \times \rho_7 - 0.125) \text{ OU } (\rho_6 \leq 1.08 \times \rho_7 - 0.048). \quad (9)$$

Os pixels identificados pela condição anterior passam por uma análise contextual, similar à proposta por Schroeder *et al.* (2016), entretanto é analisada uma vizinhança crescente ($5 \times 5, 7 \times 7, \dots$, até 61×61) parando quando a janela observada tem ao menos 25% de pixels que satisfaçam as condições propostas:

$$(R_{75} > \mu_{R_{75}} + \max(3\sigma_{R_{75}}, 0.8)) \text{ E } (\rho_7 > \mu_{\rho_7} + \max(3\sigma_{\rho_7}, 0.08)). \quad (10)$$

Assim como o método de Schroeder *et al.* (2016), a média (μ) e desvio padrão (σ) não levam em consideração os valores dos pixels identificados como fogo, fogo em potencial e água. Para o método de Kumar e Roy (2018) um pixel é definido como água se atender a seguinte condição:

$$\rho_2 \geq \rho_3 \geq \rho_4 \geq \rho_5. \quad (11)$$

O método original realiza uma análise temporal para reduzir o número de falsos positivos.

2.3 MÉTODOS DE APRENDIZADO PROFUNDO PARA DETECÇÃO DE FOGO ATIVO

Nos últimos anos, métodos de aprendizado profundo têm sido empregados em diferentes áreas, incluindo aplicações no campo de sensoriamento remoto. Algoritmos por *deep learning* podem ser aplicados, por exemplo, para prever regiões que correm risco de incêndio (YALOVEHA *et al.*, 2019), mensurar os danos causados pelo fogo (TREKIN *et al.*, 2018; KHRYASHCHEV; LARIONOV, 2020) e até mesmo para melhorar a resolução de imagens para aplicação de limiares para detecção de fogo ativo (GARGIULO *et al.*, 2019).

Fazendo uso de *deep learning*, Priya e Vani (2019) aplicaram uma CNN para identificar a ocorrência de fogo ativo em imagens de satélite obtidas de diversas fontes, como *NASA Worldview*, MODIS, *Earth Observing System Data and Information System* (EOSDIS), e imagens de satélite disponibilizadas pelo Google. Após a detecção, foi aplicada uma técnica de limiarização local para segmentar e detectar a área afetada da imagem. O método dos autores apresentou bons resultados para detecção (98.0% de *F*-score), entretanto não há um enfoque em produzir máscaras de segmentação que identifiquem a área afetada de modo preciso. Assim, apesar do bom desempenho na detecção, a saída produzida não permite a produção de máscaras que possibilitem mensurar o evento a nível de pixel e conseqüentemente a extensão do fogo de forma precisa.

Langford *et al.* (2019) utilizaram redes neurais profundas para determinar quais pixels da imagem representam fogo. Os autores avaliaram imagens do Alasca obtidas pelo satélite MODIS, com 500 m de resolução espacial. Os autores utilizaram de uma rede totalmente conectada, com cinco camadas, para classificar cada pixel de uma imagem como fogo ou fundo. O método proposto obteve aproximadamente 96.0% de recall na classificação de pixels. Vale ressaltar que a abordagem proposta não tira vantagem de propriedades espaciais da imagem, classificando cada pixel individualmente. Além disso, o estudo se limitou a apenas uma localidade, assim a performance do método não foi atestada em outras regiões do globo.

Toan *et al.* (2019) propuseram um método para detectar fogo ativo em imagens de satélite obtidas da missão *Geostationary Operational Environmental Satellite* (GOES), mais precisamente do satélite GOES-16, com uma resolução espacial de 1 km. As imagens foram divididas em recortes de 12×12 blocos, que então passaram por uma CNN que determinou se os *patches* representavam fogo. O trabalho avaliou quatro regiões dos Estados Unidos da América, obtendo um *F*-score de 93.8%. Apesar do resultado, o método atribui a mesma classe para todos pixels de um mesmo recorte, o que pode gerar falsos positivos quando a saída é analisada a nível de pixel.

Diferente das abordagens de *deep learning* listadas, o presente trabalho explorou a arquitetura de uma rede U-Net para segmentar regiões de fogo ativo em uma imagem. O uso de tal arquitetura dispensa a definição manual de diferentes limiares, ao mesmo tempo que produz máscaras que permitem mensurar, a nível de pixel, a extensão do evento observado. Para isso, foram utilizadas imagens do Landsat-8 que permitem observar o solo com uma melhor resolução espacial (30 m) quando comparado com os trabalhos desenvolvidos por Priya e Vani (2019), Langford *et al.* (2019) e Toan *et al.* (2019). Além disso, o trabalho utilizou um *dataset* com cobertura global, o que permitiu à rede observar diferentes características ao redor do mundo.

3 MATERIAL E MÉTODOS

O presente capítulo apresenta as bases de dados utilizadas (Seção 3.1), bem como as topologias de redes neurais convolucionais exploradas (Seção 3.2), o ambiente de desenvolvimento e configurações das redes (Seção 3.3) e também as métricas utilizadas nos experimentos (Seção 3.4).

3.1 DATASET

Para os experimentos foram utilizados os *datasets* construídos por Pereira *et al.* (2021). Os autores disponibilizaram dois conjuntos de dados que somam aproximadamente 220 GB de informação, com cobertura de grande parte do globo. Os *datasets* são constituídos de recortes de imagens do satélite Landsat-8, sendo que cada imagem original de aproximadamente $7,600 \times 7,600$ pixels (cerca de 185 km x 180 km) foi recortada em sub-imagens com tamanho de 256×256 pixels, sem sobreposição. Os recortes são disponibilizados em um formato TIFF de 16-bits, contendo 10 bandas, com 30 metros de resolução espacial.

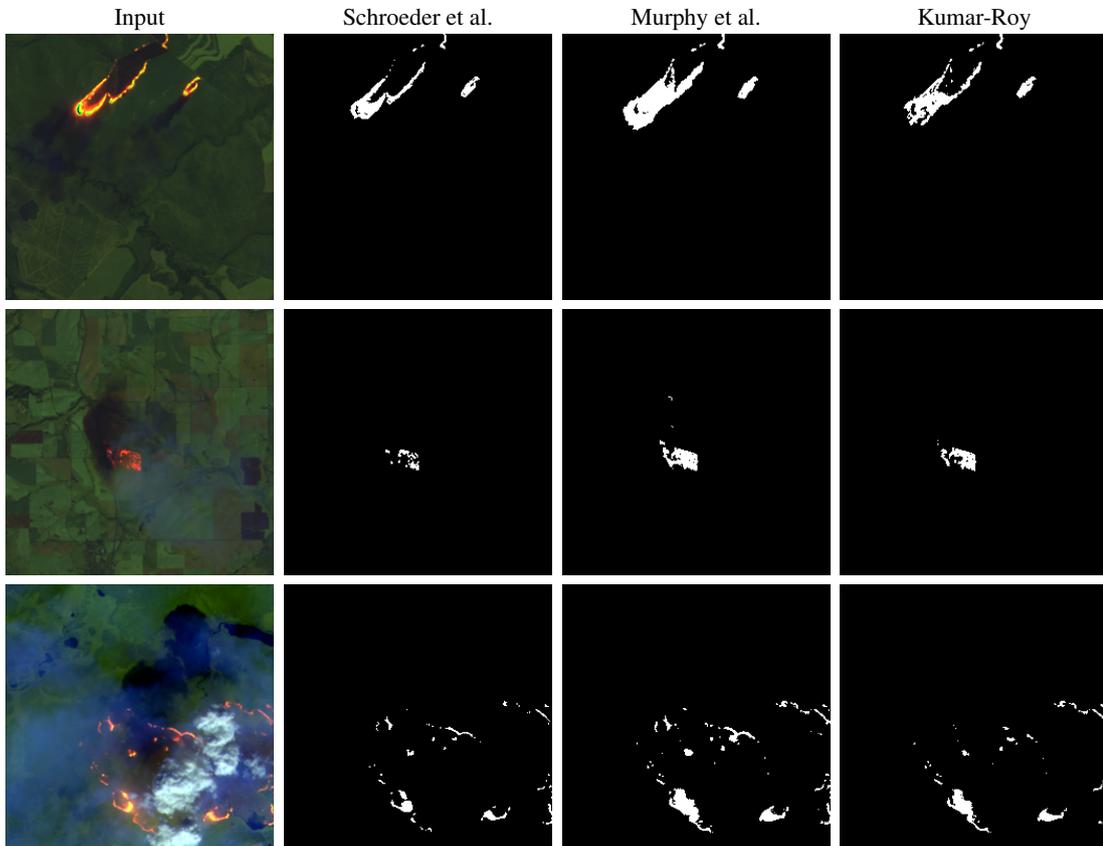
É importante observar que as imagens disponíveis no *dataset* construído por Pereira *et al.* (2021) contam com 10 das 11 bandas capturadas pelo satélite Landsat-8. O canal pancromático (banda 8), de 15 metros de resolução espacial, não está disponível nos *datasets* devido a resolução espacial diferente das demais e ao elevado custo para armazenar.

O primeiro *dataset* fornecido por Pereira *et al.* (2021), dispõe de imagens de praticamente todo o globo, com exceção da Antártida, totalizando 146.214 *patches*, referentes ao mês de Agosto de 2020. Juntamente com os recortes, também são disponibilizadas as máscaras de fogo ativo geradas pelos algoritmos tradicionais da literatura (SCHROEDER *et al.*, 2016; MURPHY *et al.*, 2016; KUMAR; ROY, 2018), bem como combinações desses métodos através da intersecção e votação majoritária das máscaras. O conjunto de imagens deste *dataset* corresponde a aproximadamente 180 GB do total de dados disponíveis.

A Figura 1 exibe exemplos de imagens disponibilizadas no *dataset* utilizado, bem como as máscaras geradas pelos algoritmos da literatura. É possível observar que, apesar dos algoritmos geralmente concordarem sobre a presença de fogo ativo, há diferenças significativas para os resultados produzidos. Visto que os métodos produzem segmentações distintas, o *dataset* também disponibiliza a combinação das máscaras por meio da intersecção das saídas dos métodos, bem

como a combinação das máscaras construídas com a abordagem de votação, em que ao menos dois métodos devem concordar sobre a ocorrência de fogo.

Figura 1 – Imagens disponíveis no *dataset*, utilizando as bandas 7, 6 e 2, e as máscaras produzidas pelos diferentes algoritmos.

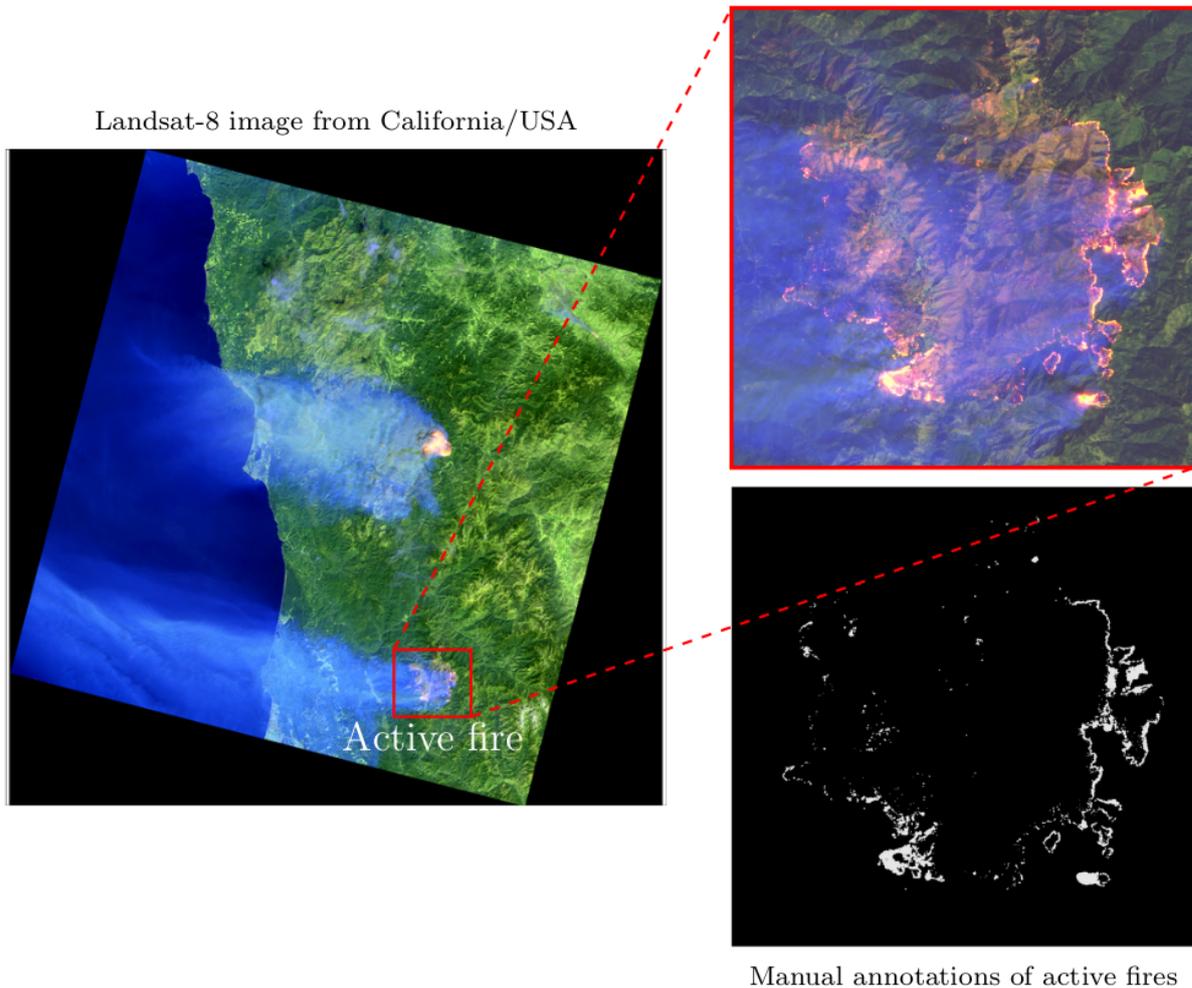


Fonte: Pereira *et al.* (2021).

Já o segundo *dataset* disponibilizado por Pereira *et al.* (2021) foi construído a partir de 13 imagens do Landsat-8 (10 delas com fogo ativo), capturadas no mês de Setembro de 2020, manualmente anotadas por um especialista da área para a tarefa de identificação de fogo ativo. Essas imagens foram recortadas em 9,044 *patches* de 256×256 pixels, sem sobreposição, e disponibilizadas em um formato TIFF de 16-bits, com 10 das 11 bandas do Landsat-8. Essas imagens correspondem a aproximadamente 12 GB do conjunto de dados. Um exemplo de imagem anotada manualmente pode ser visto na Figura 2.

As redes convolucionais exploradas no presente trabalho foram treinadas e avaliadas sobre o conjunto de dados processado pelos algoritmos da literatura descritos na Seção 2.2, bem como pela combinação desses métodos pela interseção e votação das máscaras. As imagens anotadas manualmente foram utilizadas exclusivamente para avaliar o desempenho dos modelos. O propósito de avaliar os modelos treinados sobre essas imagens é ter uma segunda referência

Figura 2 – Exemplo de anotação manual: a esquerda, uma imagem da Califórnia (Estados Unidos da América), capturada pelo Landsat-8 em 8 de Setembro de 2020. A imagem possui muitas regiões de fogo ativo, totalizando 55,386 pixels de fogo ativo identificados. Destacada em vermelho uma das duas maiores áreas de fogo ativo, com uma extensão de 22×22 km, com aproximadamente 10,000 pixels de fogo ativo; a marcação manual correspondente pode ser vista no canto inferior direito da figura.



Fonte: Pereira *et al.* (2021).

para a validação deles, evitando qualquer viés em potencial que possa surgir das saídas produzidas pelos métodos de Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018).

3.2 REDES NEURAI CONVOLUCIONAIS PARA DETECÇÃO DE FOGO ATIVO

Redes neurais convolucionais (GOODFELLOW *et al.*, 2016) são arquiteturas de rede neural do tipo *feedforward* que, nos últimos anos, têm recebido cada vez mais atenção devido ao sucesso em uma grande variedade de tarefas, muitas delas relacionadas ao processamento

(GARCIA *et al.*, 2018; WANG *et al.*, 2020), análise (PETERSSON *et al.*, 2016; LITJENS *et al.*, 2017) e reconhecimento (PAOLETTI *et al.*, 2019; YAO *et al.*, 2019) em imagens. Essa popularização se deu, principalmente, aos avanços tanto na capacidade de processamento — particularmente pelo uso de Unidades de Processamento Gráfico (GPU, do inglês *Graphic Processing Units*) para cálculos paralelos — quanto nos algoritmos e técnicas, que levaram à capacidade de aprender modelos mais profundos e complexos. As revisões da literatura feitas por Zhu *et al.* (2017), Ma *et al.* (2019) e Yuan *et al.* (2020) discutem os principais avanços do *deep learning* para o campo de sensoriamento remoto, incluindo os principais desafios, o estado da arte em recuperação de parâmetros ambientais, fusão e redução de escala de dados, classificação de cenas, detecção de objetos, classificação de uso do solo e segmentação de regiões.

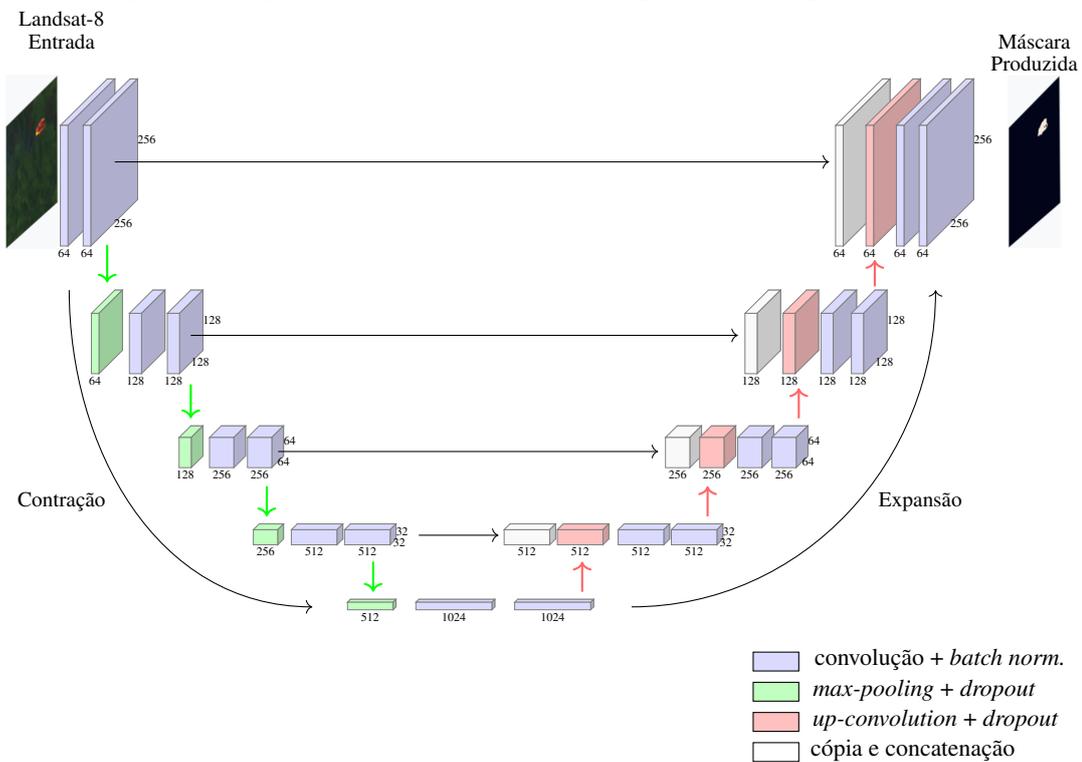
Uma rede convolucional profunda é composta por uma série de camadas, cada camada aplicando à saída da camada anterior uma série de filtros por meio da operação de convolução, seguida por algum tipo de função de ativação (GOODFELLOW *et al.*, 2016). Camadas mais profundas combinam as características extraídas pelas camadas anteriores, de modo que a rede é capaz de codificar progressivamente conceitos mais complexos (LECUN *et al.*, 2015). O processo de treinamento de uma CNN consiste em descobrir os pesos (coeficientes para os filtros), que permitem que os dados de entrada sejam transformados de modo a aproximar a saída desejada. No presente trabalho, foi considerado apenas o aprendizado supervisionado (CHAPELLE *et al.*, 2010), assim, a rede recebeu um número de amostras rotuladas — imagens de entrada e a saída associada — e iterativamente ajustou os pesos com o algoritmo de *backpropagation* (RUMELHART *et al.*, 1986) para fazer a saída produzida para um dado de entrada cada vez mais similar à amostra de saída fornecida. Embora CNNs sejam o estado da arte para um grande número de tarefas no campo de aprendizado de máquina e visão computacional, há diversos pontos que devem ser levados em consideração ao utilizá-las. Este trabalho não se aprofundará nos desafios encarados na construção de uma CNN, entretanto é importante ressaltar que arquiteturas complexas geralmente demandam maior poder computacional para serem treinadas, assim, sempre que possível é preferível utilizar arquiteturas menores e mais simples.

Os testes utilizaram uma variação de uma arquitetura U-Net (RONNEBERGER *et al.*, 2015), muito popular em segmentação de imagens. A U-Net é uma rede totalmente convolucional, com duas metades simétricas, a primeira com operações de *pooling* que reduzem a resolução dos dados e a segunda com operações de *upsampling* que restauram a resolução original dos dados. O número de filtros convolucionais aplicados é dobrado a cada operação de *pooling*, iniciando

com 64 filtros e chegando a um total de 1024, e a cada operação de *upsampling* os filtros são reduzidos pela metade, até retornar ao número inicial de 64 filtros. A primeira metade da rede extrai características básicas e contexto, enquanto a segunda metade permite uma localização a nível de pixel de cada característica, com o uso de *skip connections* entre as duas metades de tal forma a combinar características de uma camada com representações mais abstratas de camadas mais profundas.

A Figura 3 exibe a arquitetura básica da U-Net utilizada nos testes. Essencialmente é utilizado o conceito original da U-Net, mas modificada aqui com a adição de camadas de *Batch Normalization*, após cada camada convolucional, e também camadas de *Dropout*, que ajudam a evitar o problema de *overfitting* (GOODFELLOW *et al.*, 2016).

Figura 3 – Arquitetura da rede U-Net utilizada para segmentação de imagens.



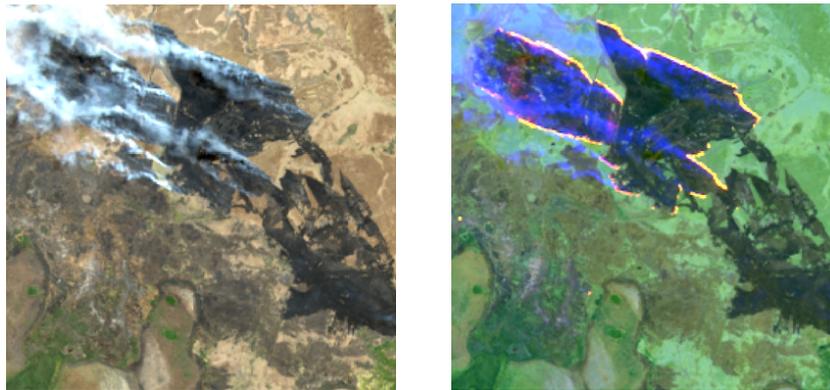
Fonte: Autoria Própria.

Para realizar os experimentos foram consideradas três variações da arquitetura U-Net. A primeira, chamada de U-Net (10c), utiliza a arquitetura básica da U-Net, usando como entrada as imagens com 10 canais do Landsat-8. A segunda arquitetura, chamada de U-Net (3c), mantém a mesma estrutura, mas utiliza como entrada apenas as bandas 7, 6 e 2 das imagens. A ideia por trás dessa arquitetura é analisar se é possível obter uma aproximação satisfatória dos resultados de modo a reduzir o número de bandas utilizadas, e com isso reduzir a largura de banda, espaço em disco e uso de memória, para obter, armazenar e processar as imagens. Assim, é possível

ter uma economia considerável de largura de banda ao processar as imagens na *Amazon Web Service* (AWS), onde estão hospedadas as imagens originais do Landsat-8, sendo possível baixar apenas as bandas desejadas.

Algoritmos simples de detecção de fogo ativo podem utilizar limiares sobre as bandas 7 e 6, referentes ao SWIR, em combinação com bandas menos sensíveis ao fogo para detectar fogo ativo (KUMAR; ROY, 2018). Desse modo, as redes convolucionais que utilizam apenas três bandas exploraram as bandas SWIR em conjunto com a banda 2, referente ao azul (Tabela 1) para compor uma imagem em RGB falso, com a banda 7 na componente vermelha, a banda 6 no canal verde e a banda 2 no respectivo canal azul. A Figura 4 exibe a diferença entre duas imagens da mesma cena, mas usando diferentes combinações de canais Landsat-8.

Figura 4 – Exemplo de combinações distintas de canais do Landsat-8: à esquerda a composição das bandas 4 (vermelho), 3 (verde) e 2 (azul); à direita a combinação de bandas 7 (SWIR), 6 (SWIR) e 2 (azul), como utilizado para treinar as CNNs com três canais. Note que apenas a fumaça é visível na imagem a esquerda, enquanto o fogo aparece visivelmente na imagem da direita.



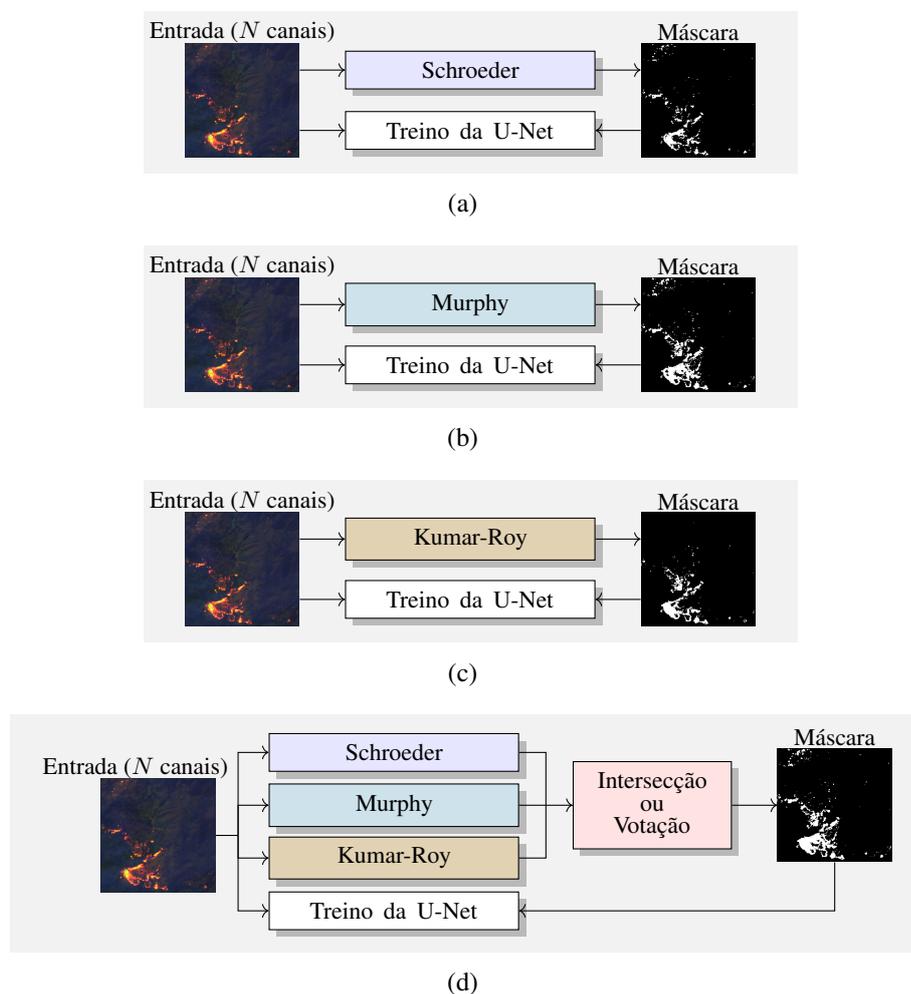
Fonte: Autoria Própria.

A terceira arquitetura, chamada U-Net-Light (3c), é uma versão reduzida da arquitetura U-Net de três canais, com o número de filtros por camada dividido por 4 (i.e. 16 filtros na primeira camada, 32 na segunda, etc). Essa é uma versão “*light*” da arquitetura original da rede U-Net.

Todas as arquiteturas produzem como saída uma imagem binária, com dimensões de 256×256 pixels, onde 0 e 1 representam, respectivamente, locais sem a presença de fogo e com a possível presença de fogo. Para obter uma saída binária, a saída produzida pela CNN é limiarizada de modo que para qualquer pixel com valor acima de 0.25 é atribuído o valor 1. Esse limiar foi definido empiricamente após alguns testes iniciais em uma pequena fração do *dataset*. Foi observado um ganho de $\approx 1 - 2\%$ na métrica *F*-score, computada conforme explicado na Seção 3.4, se comparada com o uso do valor padrão de 0.5.

Cada uma das três arquiteturas de CNN foi treinada e testada para aproximar cinco situações diferentes: as máscaras produzidas pelos conjuntos de limiares descritos na Seção 2.2, bem como a interseção e a votação dessas máscaras. Para as máscaras de votação, ao menos dois dos três conjuntos de condições devem concordar se um dado pixel representa a ocorrência de fogo. Desse modo, há um total de 15 cenários diferentes de teste, como ilustrado na Figura 5.

Figura 5 – Diferentes situações para treino e teste das redes neurais convolucionais. Cada um das três arquiteturas (U-Net (10c), U-Net (3c) and U-Net-Light (3c)) foi treinada e testada para aproximar a saída de cinco conjuntos de condições distintos: os algoritmos propostos por Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018), como podem ser visto nas figuras (a), (b) e (c), bem como a intersecção e a votação (2 votos), como visto em (d).



Fonte: Autoria Própria.

3.3 IMPLEMENTAÇÃO E AMBIENTE DE DESENVOLVIMENTO

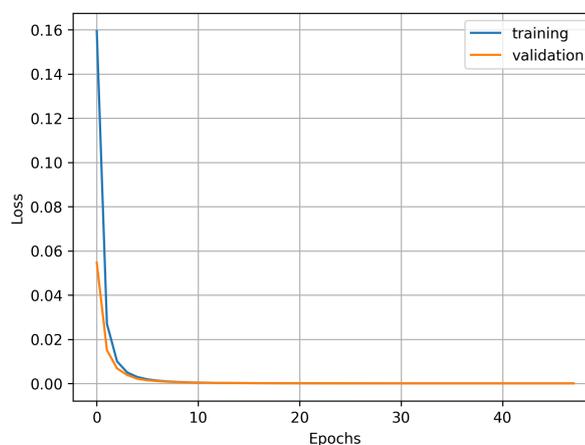
As arquiteturas de CNN propostas foram implementadas e testadas para verificar suas capacidades para aproximar os resultados das condições propostas por Schroeder *et al.* (2016),

Murphy *et al.* (2016) e Kumar e Roy (2018). Também foi avaliado o desempenho das CNNs e os métodos da literatura sobre o conjunto de imagens anotadas manualmente.

Os recortes das imagens do dataset foram divididos aleatoriamente em três conjuntos distintos, para treino, validação e teste, contendo, respectivamente, 40%, 10% e 50% do total de amostras.

Os experimentos foram realizados em uma máquina equipada com um processador Intel Core i8, 64 GB de RAM, em um sistema operacional Linux, com duas GPUs Titan XP (12 GB). O código foi desenvolvido com a linguagem Python, utilizando as bibliotecas TensorFlow¹ e Keras². As arquiteturas foram treinadas utilizando o otimizador Adam, com uma taxa de aprendizado de 0.001, um *batch* de tamanho 16 e entropia binária cruzada (do inglês, *binary cross entropy*) como função de perda, treinando por 50 épocas ou até que a função de perda sobre o conjunto de testes não obtivesse melhora por pelo menos cinco épocas. A Figura 6 exibe um exemplo de um treinamento (função de perda \times época) — no caso exibido, para a U-Net (3c) treinada sobre a intersecção das máscaras dos algoritmos da literatura (SCHROEDER *et al.*, 2016; MURPHY *et al.*, 2016; KUMAR; ROY, 2018). Os demais treinos tiveram um comportamento similar, com o número de épocas variando entre 13 e 50. A Tabela 2 exibe o número de parâmetros de cada arquitetura, o tamanho do modelo salvo em disco após o treinamento e o tempo médio de inferência por recorte.

Figura 6 – Função de perda em relação às épocas de treino para a U-Net (3c), utilizando a intersecção dos três algoritmos como máscara de treino. Os demais treinos possuem um comportamento similar. Após aproximadamente 10 épocas há pequenas variações na função de perda, que podem ser difíceis de visualizar devido a escala do eixo y , que foi gerado automaticamente com a biblioteca utilizada.



Fonte: Autoria Própria.

¹ <https://www.tensorflow.org/>

² <https://keras.io/>

Tabela 2 – Arquiteturas baseadas em uma rede U-Net para segmentação de fogo ativo.

	# Canais de entrada	# Parâmetros treináveis	Tamanho do modelo (MB)	Tempo médio de inferência por recorte (ms)
U-Net (10c)	10	34,529,153	132	36,8
U-Net (3c)	3	34,525,121	132	36
U-Net-Light (3c)	3	2,161,649	8,5	25,5

Fonte: Autoria Própria.

3.4 MÉTRICAS DE AVALIAÇÃO

As redes convolucionais foram avaliadas com relação às principais métricas utilizadas em desafios de segmentação de imagens, como PASCAL VOC (EVERINGHAM *et al.*, 2010), KITTI (FRITSCH *et al.*, 2013) e COCO³ (LIN *et al.*, 2014). A métrica *F*-score (CHINCHOR; SUNDHEIM, 1993), também conhecida como coeficiente Dice, é amplamente utilizado como métrica para ranqueamento, e consiste da média harmônica das métricas de precisão (*P*) e *recall* (*R*):

$$F = \frac{2}{1/P + 1/R} \quad (12)$$

tal que:

$$P = \frac{tp}{tp + fp} \quad R = \frac{tp}{tp + fn} \quad (13)$$

onde *tp* representa os verdadeiros positivos, ou seja, o número de pixels com presença de fogo classificados corretamente, *fp* denota os falsos positivos, esse valor corresponde ao número de pixels sem a presença de fogo classificados incorretamente como fogo, e *fn* os falsos negativos, sendo o número de pixels com a presença de fogo mas classificados incorretamente como fundo.

Também foi aplicada a métrica de Intersecção Sobre União, do inglês *Intersection Over Union* (IoU) (PINHEIRO *et al.*, 2015), também conhecida como índice de Jaccard, definida por:

$$\text{IoU} = \frac{tp}{tp + fn + fp}. \quad (14)$$

A métrica IoU cobre as mesmas informações que a métrica *F*-score — $\text{IoU} = F/(2 - F)$ — mas também é amplamente utilizada em diversos desafios de segmentação; para maior clareza ambas serão relatadas. Todas as métricas foram computadas com a implementação de avaliação do COCO: *tp*, *fp* e *fn* são acumulados para todas as imagens e as métricas são computadas uma única vez para todo o *dataset*, ou seja, os resultados obtidos não refletem a performance média por recorte, mas sim a performance global por pixel.

³ <http://cocodataset.org>

Como no desafio KITTI (LYU *et al.*, 2019), a classificação da performance dos algoritmos se deu pela métrica F -score. Uma métrica comumente utilizada que não foi listada nos experimentos é a acurácia por pixel, que leva em conta o número de verdadeiros negativos (pixels de não fogo classificados corretamente). Nos testes realizados foi possível observar um grande desbalanceamento entre o número de pixels sem fogo e com fogo – mais de 99% do total de amostras são pixels de não fogo — resultando em uma acurácia muito alta. Para um caso extremo de um detector que falha em detectar qualquer pixel de incêndio a acurácia ainda pode parecer funcionar razoavelmente bem, tornando-se uma métrica enganosa.

4 RESULTADOS E DISCUSSÃO

O presente capítulo apresenta os resultados obtidos pelas redes neurais artificiais treinadas e avaliadas sobre máscaras produzidas por algoritmos da literatura (Seção 4.1), bem como o desempenho obtido ao comparar as redes treinadas e os conjuntos de condições da literatura com máscaras produzidas por um especialista (Seção 4.2). Além disso, também são discutidas possibilidades para melhoria dos resultados obtidos (Seção 4.3).

4.1 DESEMPENHO NAS IMAGENS SEGMENTADAS AUTOMATICAMENTE

Para o primeiro grupo de experimentos, foram avaliadas as arquiteturas CNNs na tarefa de aproximar a segmentação original obtida pelas condições propostas por Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018), bem como a combinação desses métodos (intersecção e votação). As arquiteturas CNNs foram treinadas e validadas com 50% das imagens disponíveis, enquanto os 50% restantes foram reservados para teste. As tabelas desta seção mostram os resultados sobre o conjunto de teste.

No total foram avaliados 15 cenários de testes, mostrados na Tabela 3, onde são exibidas as métricas obtidas para as três arquiteturas de CNN (U-Net (10c), U-Net (3c) e U-Net-Light (3c)) na aproximação dos resultados de cinco conjuntos diferentes de máscaras.

Como pode ser observado na Tabela 3, todas as arquiteturas foram capazes de reproduzir o comportamento dos diferentes algoritmos razoavelmente bem, sendo o menor valor obtido para a métrica F -score igual a 84.2% e o maior igual a 94.2%. Um comportamento interessante a se notar é que a maior rede (U-Net (10c)) não é necessariamente superior à rede que utiliza um número reduzido de canais (U-Net (3c)), tampouco a uma rede menos complexa e com um número reduzido de canais (U-Net-Light (3c)). Isso significa que provavelmente essa aplicação, em particular, não demanda uma arquitetura complexa, e mais importante, as bandas 7, 6 e 2 contêm informações suficientes para a detecção de fogo ativo. Além disso, há uma tendência de que os modelos com maior precisão tenham menor *recall* e vice-versa, mas não com uma grande margem na maioria dos casos, o que mostra que o procedimento de aprendizagem está equilibrando detecções falsas e perdas.

As Figuras 7, 8 e 9 comparam as saídas produzidas pelas CNNs com, respectivamente, as máscaras produzidas pelos conjuntos de condições de Schroeder *et al.* (2016), Murphy *et al.*

Tabela 3 – Desempenho do reconhecimento de fogo ativo: foram utilizados os conjuntos de condições propostas por Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018) para treinar três redes neurais convolucionais: U-Net (10c) que utiliza como entrada uma imagem de 10 canais do Landsat-8; U-Net (3c), que utiliza como entrada uma imagem de três canais, contendo apenas as bandas 7, 6 e 2, que possuem uma boa resposta para fogo ativo; e U-Net-Light (3c) que é uma versão reduzida da U-Net de 3 canais. Os desempenhos estão relacionados à capacidade da CNN de reproduzir as máscaras de Schroeder *et al.* (2016), Murphy *et al.* (2016), Kumar e Roy (2018), e também as máscaras geradas pela intersecção e votação das máscaras originais, em um conjunto de teste nunca visto por essas arquiteturas.

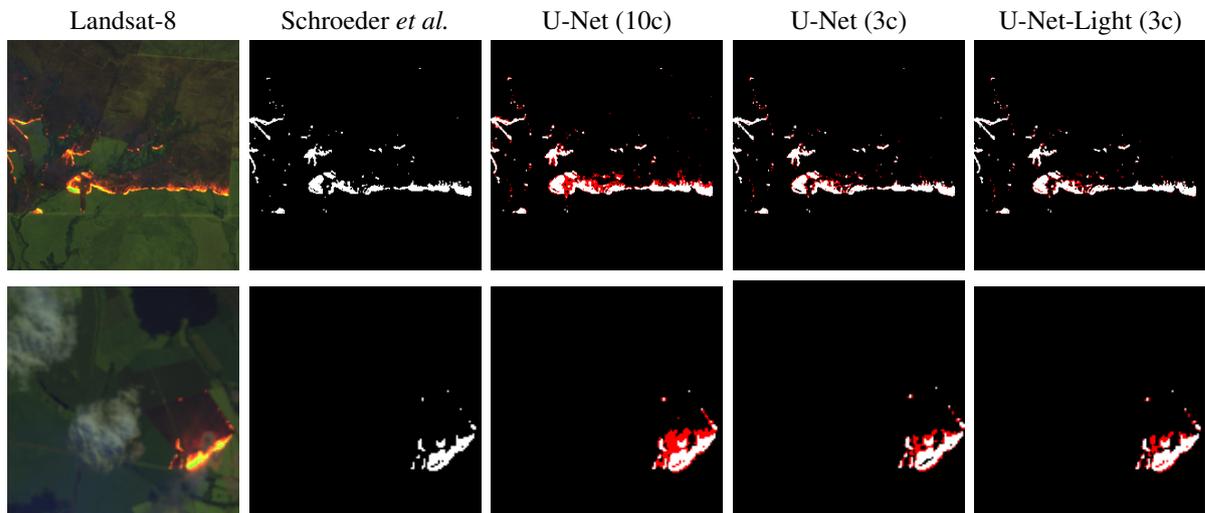
Máscara	Arquitetura	Métricas (%)				
		<i>P</i>	<i>R</i>	IoU	<i>F</i>	
Schroeder <i>et al.</i>	U-Net (10c)	86.8	89.7	78.9	88.2	
	U-Net (3c)	89.8	88.8	80.7	89.3	
	U-Net-Light (3c)	90.8	86.1	79.2	88.4	
Murphy <i>et al.</i>	U-Net (10c)	93.6	92.5	87.0	93.0	
	U-Net (3c)	89.1	97.6	87.2	93.2	
	U-Net-Light (3c)	92.6	95.1	88.4	93.8	
Kumar e Roy	U-Net (10c)	84.6	94.1	80.3	89.1	
	U-Net (3c)	84.2	90.6	77.4	87.3	
	U-Net-Light (3c)	76.8	93.2	72.7	84.2	
Intersecção	Schroeder <i>et al.</i>	U-Net (10c)	84.4	99.7	84.2	91.4
	Murphy <i>et al.</i>	U-Net (3c)	93.4	92.4	86.7	92.9
	Kumar e Roy	U-Net-Light (3c)	87.4	97.3	85.4	92.1
Votação	Schroeder <i>et al.</i>	U-Net (10c)	92.9	95.5	89.0	94.2
	Murphy <i>et al.</i>	U-Net (3c)	91.9	95.3	87.9	93.6
	Kumar e Roy	U-Net-Light (3c)	90.2	96.5	87.3	93.2

Fonte: Autoria Própria.

(2016) e Kumar e Roy (2018), para amostras do conjunto de teste. Pode ser observado que as CNNs obtiveram sucesso na aproximação das máscaras desejadas, com algumas diferenças nas máscaras produzidas. Mesmo a arquitetura com 3 canais, utilizando apenas as bandas 7, 6 e 2, obteve sucesso na maioria dos casos.

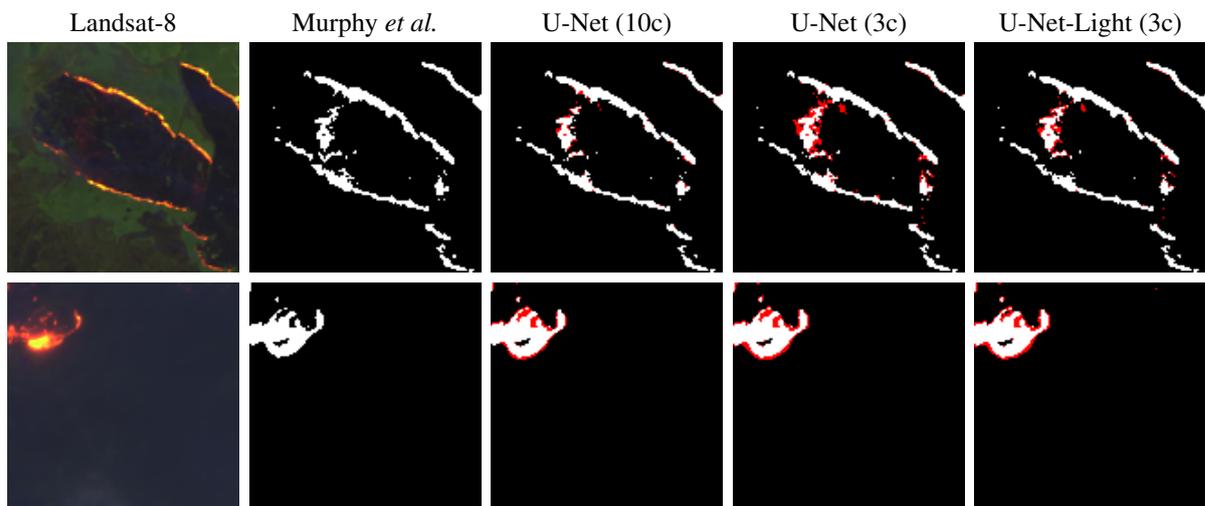
Ao observar os falsos positivos produzidos pelas redes (destaque em vermelho nas Figuras 7, 8 e 9) é notável que a maioria das falsas detecções estão próximas das regiões de fogo ativo. Esse comportamento pode ser justificado pela identificação de focos de calor menos intenso (brasa) próximos de regiões de fogo ativo que não foram capturados pelos algoritmos geradores das máscaras. Tais erros produzidos pelas redes podem indicar que os métodos de aprendizado profundo foram capazes de capturar algumas regiões que os algoritmos originais não capturaram. Essa diferença nos resultados deixa clara a necessidade da comparação dos resultados produzidos pelos algoritmos de Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018) bem como das saídas das CNNs com as anotações manuais de um especialista, como feito na Seção 4.2.

Figura 7 – Resultados da segmentação de fogo ativo obtidos pelo algoritmo de Schroeder *et al.* (2016), e pelas redes treinadas utilizando a saída desse método como máscara desejada. Em vermelho, detecções geradas pelas redes, mas não geradas pelas máscaras originais (falsos positivos).



Fonte: Autoria Própria.

Figura 8 – Resultados da segmentação de fogo ativo obtidos pelo algoritmo de Murphy *et al.* (2016), e pelas redes treinadas utilizando a saída desse método como máscara desejada. Em vermelho, detecções geradas pelas redes, mas não geradas pelas máscaras originais (falsos positivos).

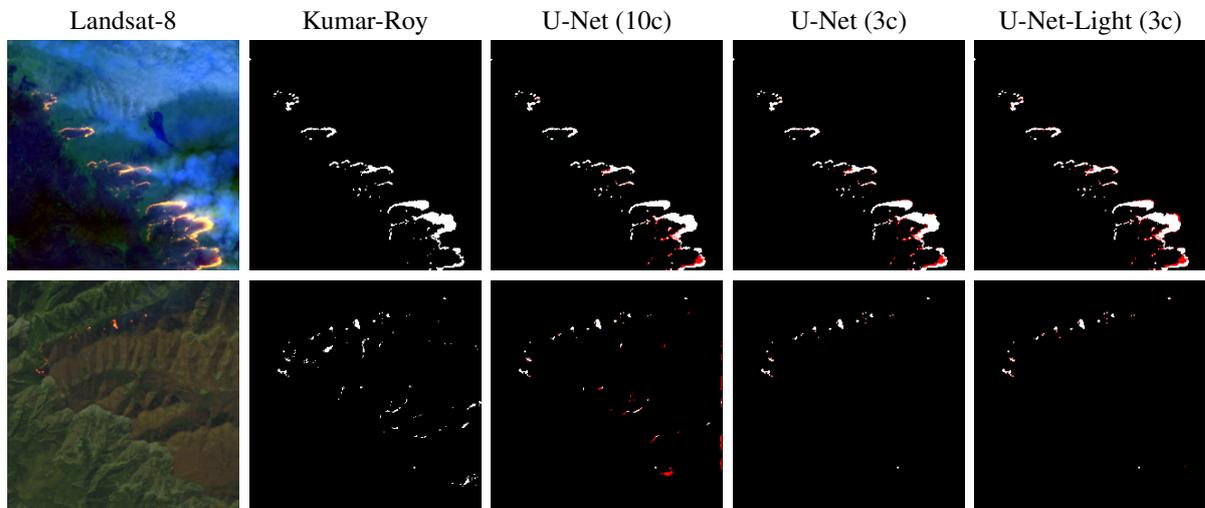


Fonte: Autoria Própria.

A segunda linha da Figura 9, que possui pequenas regiões de fogo ativo detectados pelo algoritmo original e pela CNN de 10 canais, mas que não foram encontrados pelas redes de três canais, é uma exceção de boa aproximação para as redes com 3 canais. As regiões não detectadas pelas redes de três canais são difíceis de visualizar, o que pode indicar que as redes aprenderam a evitar detecções falsas sob certas condições. Assim, é reforçada a importância da comparação dos métodos com a anotação de um especialista.

Nas Figuras 10 e 11, são comparadas as saídas dos três conjuntos de condições para detecção de fogo ativo (primeiro a intersecção e então a votação), bem como as saídas produzidas

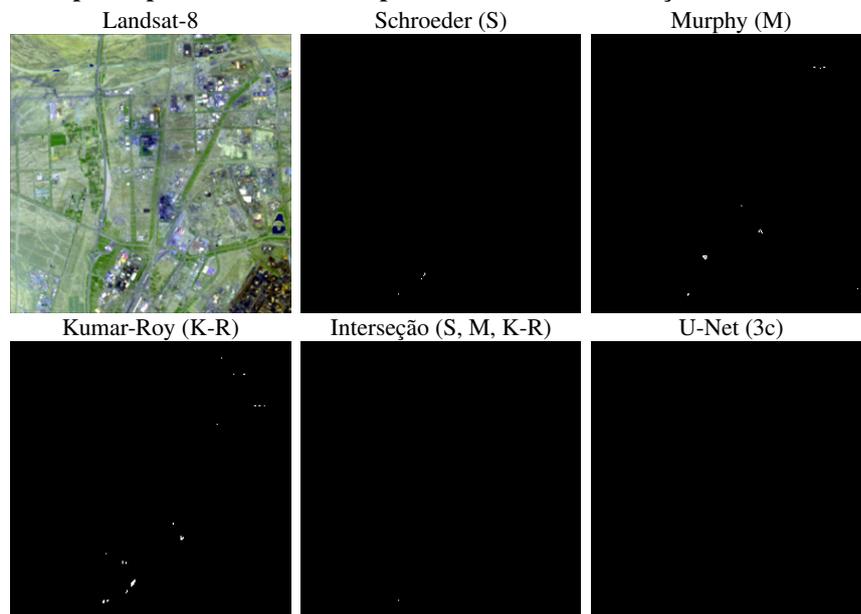
Figura 9 – Resultados da segmentação de fogo ativo obtidos pelo algoritmo de Kumar e Roy (2018), e pelas redes treinadas utilizando a saída desse método como máscara desejada. Em vermelho, detecções geradas pelas redes, mas não geradas pelas máscaras originais (falsos positivos).



Fonte: Autoria Própria.

pela U-Net de três canais treinada para aproximar essas combinações.

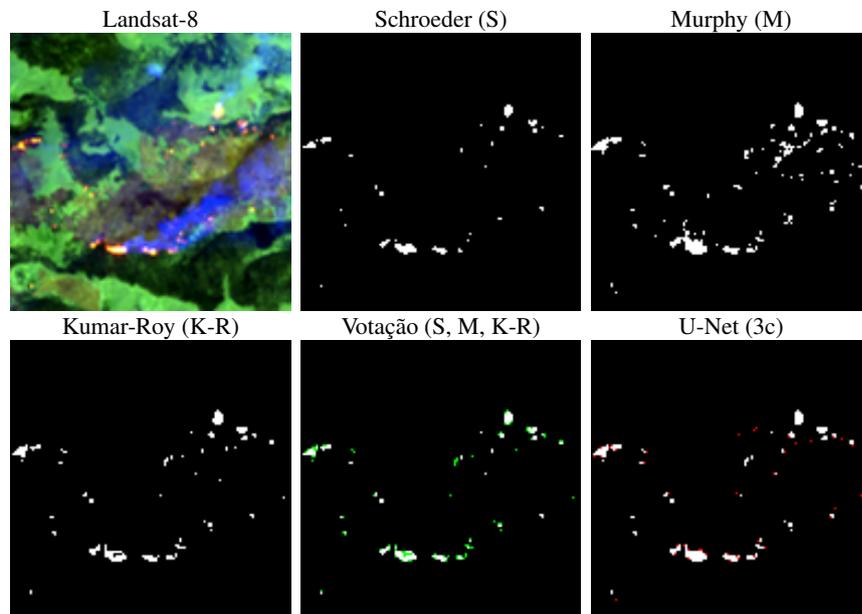
Figura 10 – Os algoritmos propostos por Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018) geram detecções falsas para ambientes urbanos, mas a intersecção e a U-Net de três canais treinada para aproximá-la foram capazes de evitar tais detecções.



Fonte: Autoria Própria.

A Figura 10 mostra que, enquanto os três conjuntos de condições geralmente detectam incorretamente fogo ativo em ambientes urbanos, a intersecção normalmente se mostra mais robusta, embora isso talvez resulte em um *recall* inferior. Já a Figura 11 mostra que o esquema de votação reduz comportamentos associados a apenas um dos conjuntos de condições, como a tendência das condições propostas por Murphy *et al.* (2016) de produzir aglomerados de fogo

Figura 11 – A U-Net de três canais foi capaz de aproximar as máscaras de votação geradas pela combinação das saídas produzidas pelos algoritmos de Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018). O esquema de votação reduz erros causados por artefatos produzidos por apenas um dos algoritmos. Em verde, na máscara de votação, as detecções geradas por apenas dois dos três métodos. Em vermelho, na máscara gerada pela rede U-Net (3c), as detecções falsas detectadas em relação a máscara de votação.



Fonte: Autoria Própria.

mais espessos e, neste exemplo, o “buraco” na região inferior do fogo das condições de Kumar e Roy (2018). A máscara de votação da Figura 11 exhibe em destaque (cor verde) as detecções geradas por apenas dois dos três métodos. Em ambos os casos, as CNNs treinadas tiveram sucesso em aproximar as máscaras combinadas. Outra questão levantada por essas imagens é: como os três conjuntos de condições se comparam entre si e com as redes que aprendem a combiná-los? Este assunto será analisado com mais profundidade na Seção 4.2.

4.2 DESEMPENHO NAS IMAGENS ANOTADAS MANUALMENTE

Para o segundo grupo de experimentos, as redes treinadas no primeiro grupo de experimentos (Seção 4.1) foram aplicadas sobre os recortes das imagens do Landsat-8 do *dataset* anotado manualmente, e os resultados comparados com as máscaras criadas manualmente. Assim, é possível observar o quão bem as redes treinadas sobre os *patches* segmentados automaticamente se comportam quando comparadas com anotações de um especialista humano. É preciso manter em mente que especialistas diferentes podem produzir segmentações distintas, já que pode ser subjetivo separar os limites de áreas queimadas de áreas que ainda estão queimando, além da presença de fumaça e dos diferentes níveis de intensidade do fogo. Para trabalhos sobre áreas

queimadas, o leitor pode consultar os trabalhos de Chen *et al.* (2020), Malambo e Heatwole (2020).

A Tabela 4 exibe o desempenho obtido para cada um dos 15 cenários de teste envolvendo as arquiteturas das CNNs, como descrito na Seção 4.1, bem como, o desempenho para o conjunto original de condições de Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018) sem nenhuma CNN. Como pode ser visto, os melhores valores para a métrica F -score ficaram em torno de 90% — isso é esperado, uma vez que existe viés tanto do especialista humano que rotula os dados quanto dos próprios algoritmos, que são ajustados para funcionar sob certas suposições. Como um exemplo, a Figura 12 exibe um recorte da imagem do Landsat-8, com uma grande área de fogo ativo, juntamente com a marcação manual associada, e também as máscaras produzidas por cada um dos cenários considerados.

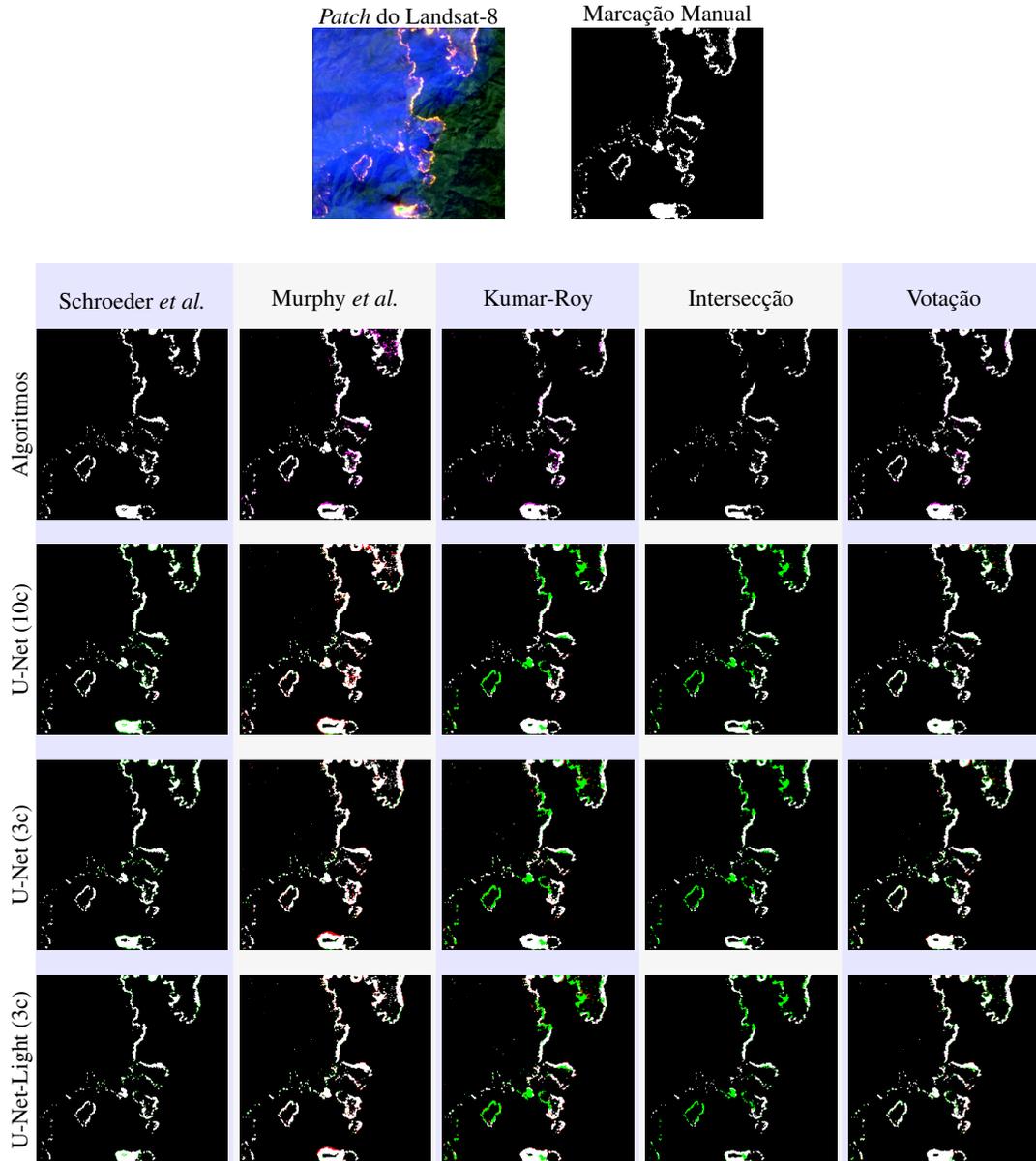
Tabela 4 – Desempenho do reconhecimento de fogo ativo em relação as anotações manuais para 13 cenas do Landsat-8. As linhas com “—” na coluna de arquitetura não utilizam nenhuma CNN, o desempenho é computado comparando as máscaras geradas pelos conjuntos de condições de Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018) com as anotações manuais. As entradas restantes correspondem a redes profundas treinadas com máscaras extraídas automaticamente usando esses conjuntos de condições. Todas as entradas são comparadas com as mesmas anotações manuais e os melhores valores para cada métrica são exibidos em destaque (negrito).

Máscara	Arquitetura	Métricas (%)				
		P	R	IoU	F	
Schroeder <i>et al.</i> (sem CNN)	—	88.1	70.2	64.1	78.1	
Schroeder <i>et al.</i>	U-Net (10c)	86.9	88.4	78.0	87.7	
	U-Net (3c)	89.5	80.5	73.6	84.8	
	U-Net-Light (3c)	89.0	78.8	71.8	83.6	
Murphy <i>et al.</i> (sem CNN)	—	76.6	96.1	74.3	85.2	
Murphy <i>et al.</i>	U-Net (10c)	74.8	96.9	73.0	84.4	
	U-Net (3c)	72.7	97.2	71.2	83.2	
	U-Net-Light (3c)	75.5	96.9	73.7	84.9	
Kumar e Roy (sem CNN)	—	82.3	68.4	59.7	74.7	
Kumar e Roy	U-Net (10c)	82.2	94.2	78.3	87.8	
	U-Net (3c)	84.5	93.4	79.8	88.8	
	U-Net-Light (3c)	78.8	96.9	76.9	86.9	
Intersecção (sem CNN)	—	92.6	57.6	55.1	71.0	
Intersecção	Schroeder <i>et al.</i>	U-Net (10c)	91.8	75.4	70.6	82.8
	Murphy <i>et al.</i>	U-Net (3c)	90.8	73.5	68.4	81.2
	Kumar e Roy	U-Net-Light (3c)	90.8	72.8	67.8	80.8
Votação (sem CNN)	—	87.7	79.1	71.2	83.2	
Votação	Schroeder <i>et al.</i>	U-Net (10c)	83.6	94.0	79.3	88.5
	Murphy <i>et al.</i>	U-Net (3c)	86.4	93.0	81.1	89.6
	Kumar e Roy	U-Net-Light (3c)	87.2	92.4	81.4	89.7

Fonte: Autoria Própria.

Como pode ser visto na Tabela 4, as condições originais propostas por Murphy *et al.*

Figura 12 – Resultados da segmentação de fogo ativo produzidos pelos conjuntos de condições de Schroeder *et al.* (2016), Murphy *et al.* (2016) e Kumar e Roy (2018), a combinação desses métodos (intersecção e votação) e as redes profundas treinadas com cada abordagem para um recorte de uma imagem do Landsat-8. Em magenta, na linha das máscaras dos algoritmos, falsos positivos gerados pelos algoritmos quando comparados com a marcação manual. Em verde, as detecções feitas pelas redes em concordância com as anotações do especialista e que não foram capturadas pela máscara da respectiva abordagem. Em vermelho, as detecções feitas pelas redes e não anotadas pelo especialista nem identificadas pelas máscara da respectiva abordagem.



Fonte: Autorial Própria.

(2016) se saem melhor que as condições de Schroeder *et al.* (2016) e Kumar e Roy (2018) sobre o *dataset* anotado manualmente, principalmente devido a sua tendência de detectar mais pixels de fogo que as outras condições, resultado em um *recall* muito alto, ao custo de uma precisão inferior (ou seja, é mais propenso a falsas detecções). As condições propostas por Kumar e Roy (2018) possuem uma precisão e *recall* inferior se comparado com as condições de Schroeder

et al. (2014) — na realidade, ele falha em detectar vários pixels de incêndio no exemplo da Figura 12, um problema que será discutido brevemente mais adiante.

Quanto às CNNs, a maioria das diferenças entre elas e os conjuntos de condições descritos na Seção 2.2 são na verdade positivas, com as redes profundas tendo um desempenho melhor que os conjuntos de condições de Schroeder *et al.* (2016) e Kumar e Roy (2018), com maior *recall* sem impactar significativamente a precisão, como pode ser visto na Tabela 4. Na Figura 12, as redes treinadas para aproximar as condições de Kumar e Roy (2018) foram capazes de detectar muitos pixels de fogo ativo que não foram identificados pelo algoritmo dos autores (destacados em verde), esse comportamento justifica o *recall* superior das redes convolucionais em relação a abordagem sem CNN.

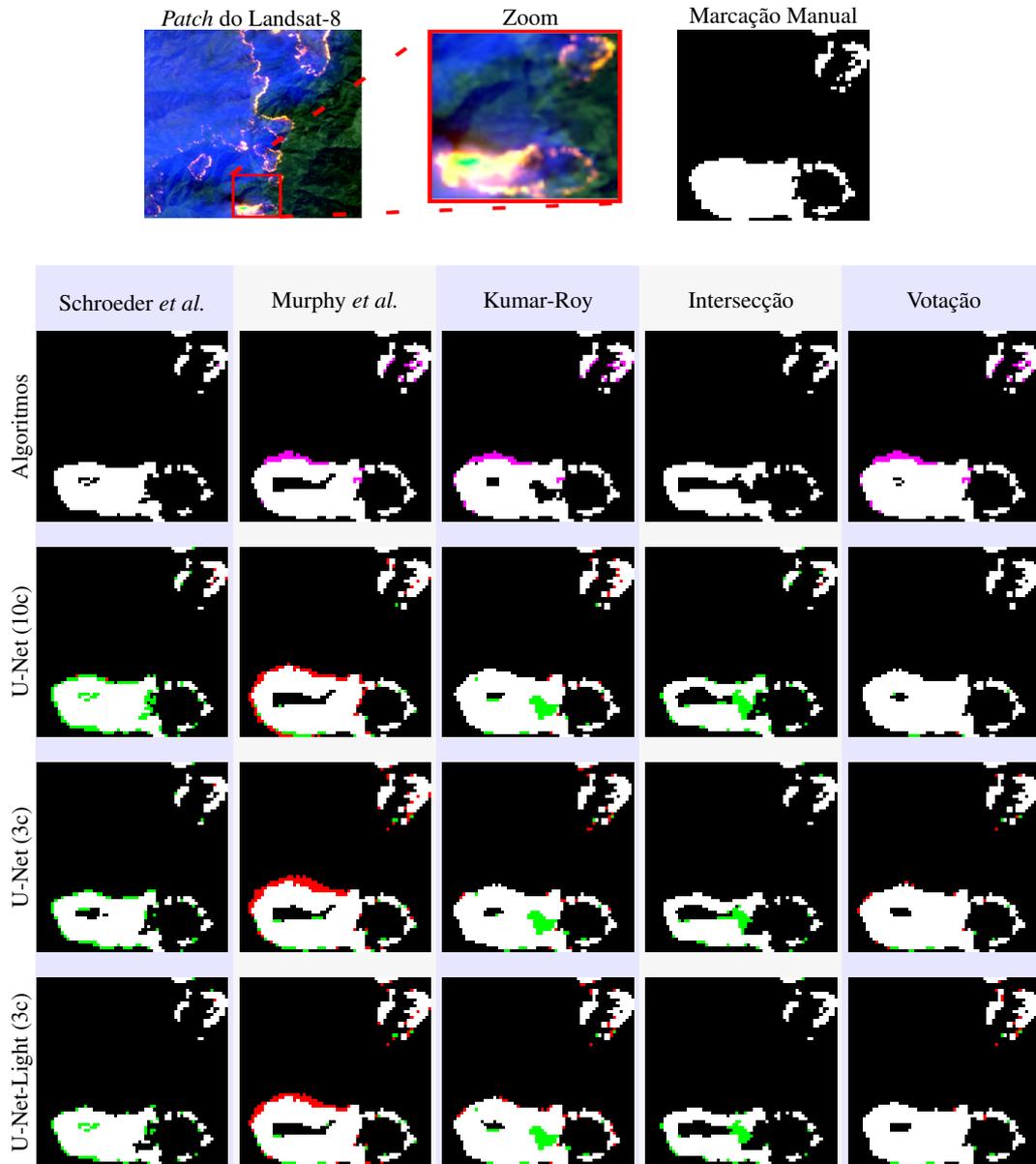
As redes profundas também mostram um maior *recall* (mas precisão inferior) se comparadas com as condições de Murphy *et al.* (2016). Na Figura 13 é possível observar a tendência em produzir falsos positivos das redes treinadas com as máscaras produzidas pelo método de Murphy *et al.* (2016). Já a interseção das condições, tanto para as três redes treinadas quanto sem nenhuma CNN, obtiveram maior precisão, mas com um *recall* consideravelmente menor, uma vez que esta combinação mantém apenas aqueles pixels que todas as três redes concordam serem pixels de fogo, ou seja, é muito restritiva. Esse rigor pode ser visto na Figura 13, onde estão destacados na cor magenta, os falsos positivos gerados pelos algoritmos da literatura, de modo que a intersecção produz poucos falsos positivos, mas em compensação gera mais falsos negativos.

Todos os conjuntos de condições tiveram problemas para detectar o núcleo do fogo, que tem intensidade muito alta, como pode ser visto na Figura 13. Apesar da maioria dos resultados não ter identificado tal área, as redes U-Net (10c) e U-Net-Light (3c) treinadas com as máscaras de Schroeder *et al.* (2016) conseguiram generalizar o problema e segmentar a região corretamente (destaque em verde na Figura 13).

O melhor desempenho geral — incluindo os conjuntos de condições originais — foi obtido pelo esquema de votação, que resultou no maior *recall* sem grandes impactos na precisão, com a U-Net-Light (3c) sendo a abordagem que produziu resultados mais próximos da segmentação manual, mas sem diferença significativa para a U-Net (3c). É importante notar que as redes foram treinadas usando apenas amostras do conjunto de dados segmentados automaticamente, portanto, o desempenho obtido não pode ser causado por algum viés aprendido com os *patches* anotados manualmente, que só foram vistos nos testes.

Além dos resultados quantitativos discutidos até o momento, uma série de casos indivi-

Figura 13 – Ampliação em uma região da imagem, bem como os resultados produzidos pelos métodos da literatura e pelas CNNs para segmentação de fogo ativo. Em magenta, na linha das máscaras dos algoritmos handcrafted tradicionais na literatura, falsos positivos gerados pelos algoritmos quando comparados com a marcação manual. Em verde, as detecções feitas pelas redes em concordância com as anotações do especialista e que não foram capturadas pelos algoritmos originais. Em vermelho, as detecções feitas pelas redes e não anotadas pelo especialista nem identificadas pelos algoritmos originais.

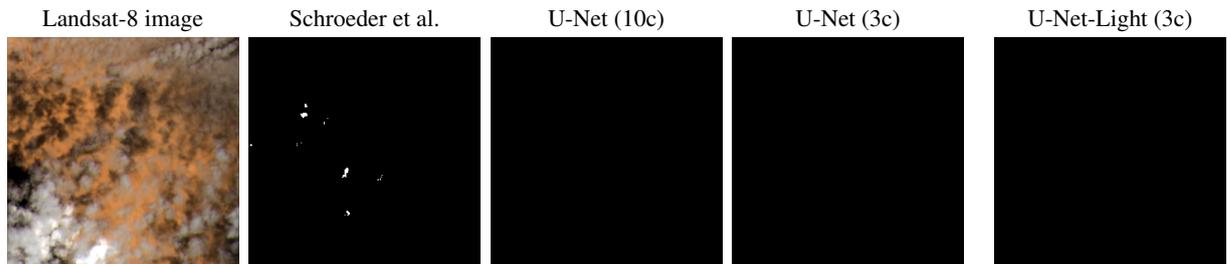


Fonte: Autoria Própria.

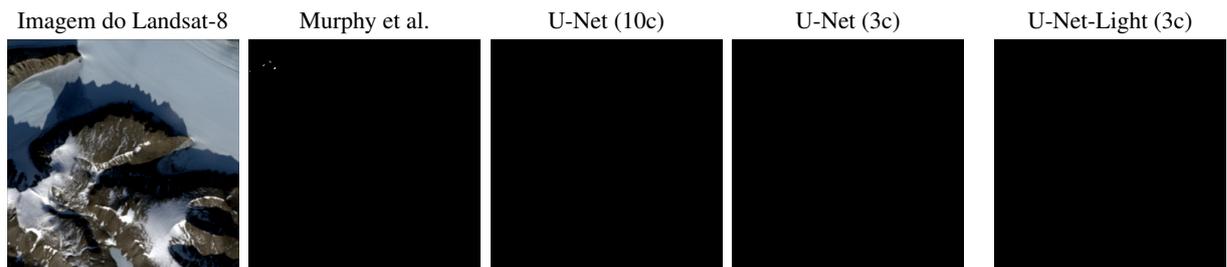
duais em que o desempenho ou *recall* obtidos por uma abordagem foram particularmente baixos foram analisados. Um exemplo é a segmentação para as condições originais de Kumar e Roy (2018) exibida na Figura 12, que perdeu muitos pixels de fogo. Foi observado que esses pixels foram de fato detectados como pixels de fogo, mas posteriormente foram descartados após serem considerados pixels de água, devido a valores muito similares nas bandas 2 a 5. Outros casos são

apresentados na Figura 14, que mostra casos de falha para cada um dos conjuntos de condições. Na primeira linha, Figura 14(a), as condições de Schroeder *et al.* (2016) produziram detecções falsas na região do deserto do Sahara. Na segunda e terceira linha, Figura 14(b, c), as condições de Murphy *et al.* (2016) e Kumar e Roy (2018) produziram detecções falsas em algumas regiões da Groenlândia.

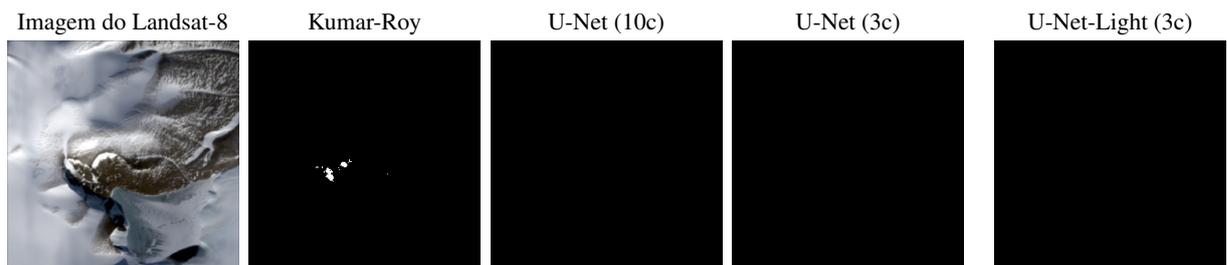
Figura 14 – Erros de classificação: (a) região do deserto do Sahara, WRS: 189/046, data: 2020/09/18; (b) e (c) região da Groenlândia, WRS: 003/006, data: 2020/09/11; e (d) cidade de Milão, Itália, WRS: 193/029, data: 2020/09/14, onde todas as condições originais e também as arquiteturas CNN reportaram erros de classificação em um ambiente urbano.



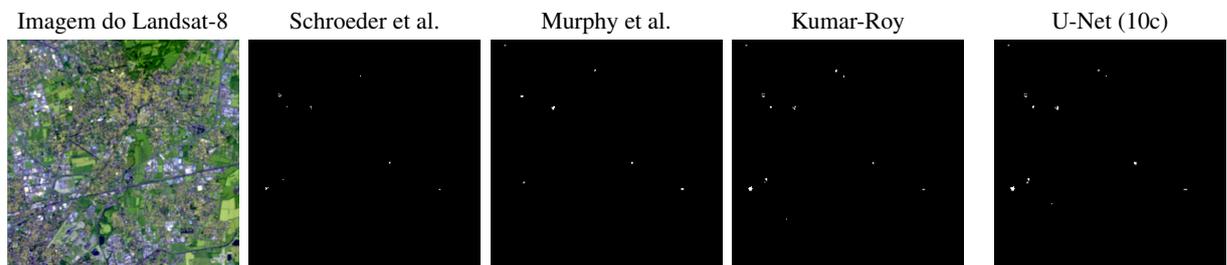
(a) Segmentações das condições de Schroeder *et al.* (2016) e das arquiteturas U-Net treinadas com essas condições.



(b) Segmentações das condições de Murphy *et al.* (2016) e das arquiteturas U-Net treinadas com essas condições.



(c) Segmentações das condições de Kumar e Roy (2018) e das arquiteturas U-Net com essas condições.



(d) Segmentações das condições de Schroeder *et al.* (2016), Murphy *et al.* (2016), Kumar e Roy (2018) e uma arquitetura U-Net.

Fonte: Autoria Própria.

Ao analisar esses casos em mais detalhes, observando cada etapa e componente das

condições, é possível notar que a maioria dos erros não foi causada por problemas fundamentais com as próprias condições, mas por valores de refletância que aparecem muito próximos dos limiares de uma regra ou outra. Por exemplo, alguns erros poderiam ser evitados se o limiar 0.2 de uma condição de fogo inequívoca do método proposto por Schroeder *et al.* (2016) (Equação 1) fosse alterado para 0.194, entretanto isso também pode fazer com que incêndios reais não sejam detectados. Isso destaca um dos desafios enfrentados ao criar conjuntos de regras à mão: os limiares geralmente são arredondados para uma ou duas casas decimais e apenas as relações mais importantes entre as bandas são consideradas. Técnicas de aprendizado de máquina são capazes de codificar regras complexas, com pesos, coeficientes e limiares mais precisos; embora isso também possa significar que as regras e relações aprendidas podem ser difíceis de entender e serem explicadas para um ser humano. Em todo caso, as redes profundas foram capazes de evitar erros nas amostras exibidas nas primeiras três linhas da Figura 14. Um caso que não foi propriamente aproximado pelas CNNs é exibido na última linha (d) da Figura 14 — todas as abordagens produziram detecções falsas em ambientes urbanos. A partir desses exemplos, acreditamos que esses tipos de erros persistentes devem ser tratados de forma diferente, considerando a análise temporal conforme descrito por Schroeder *et al.* (2016), uma vez que é muito improvável que um fogo ativo permaneça ativo no mesmo local por vários meses.

4.3 DISCUSSÃO

Nas Seções 4.1 e 4.2 foram discutidos resultados específicos sobre, respectivamente, quão bem os modelos de aprendizagem profunda poderiam aproximar os resultados produzidos por algoritmos com limiares definidos manualmente e como as CNNs e os conjuntos de condições da literatura se comportavam quando comparados a um especialista humano. As principais conclusões foram que as CNNs podem de fato aproximar algoritmos que utilizam limiares definidos manualmente, mesmo com uma arquitetura reduzida e utilizando apenas 3 canais de dados dos sensores. Além disso, com CNNs que utilizam a combinação dos métodos, é possível obter melhor desempenho em comparação com os conjuntos originais de condições, e as situações em que as CNNs diferem dos algoritmos originais podem na verdade ser positivas, permitindo regras e relações mais complexas, embora com a desvantagem de ser difícil de entender ou descrever para um humano.

É importante notar que as arquiteturas de CNN descritas no presente trabalho são as

que relataram resultados mais promissores, entretanto, muitas outras variações foram testadas, tais como arquiteturas sem *batch normalization* (que foi incapaz de gerar resultados próximos dos desejados, provavelmente devido ao problema de desaparecimento do gradiente), variações com mais e também com menos camadas convolucionais ou filtros por camada. É preciso enfatizar que o foco principal do trabalho não foi encontrar arquiteturas de CNN extremamente otimizadas, mas sim analisar a viabilidade da abordagem para detecção do fogo ativo. Assim, os resultados obtidos provavelmente podem ser melhorados de diferentes formas: pesquisas futuras podem explorar o uso de super-resolução de imagens (GARGIULO *et al.*, 2019; MA *et al.*, 2019); *ensembles* de arquiteturas de redes neurais convolucionais (MINETTO *et al.*, 2019); e informação espaço-temporal (BOSCHETTI *et al.*, 2017) para diferenciar entre pixels com fogo persistente como fábricas e telhados para regiões de incêndios florestais. Além disso, abordagens orientadas ao aprendizado de máquina podem ser facilmente adaptadas para outros sensores (MATEO-GARCÍA *et al.*, 2020) (e.g. Sentinel-2), utilizando a técnica de transferência de aprendizado, potencialmente sem exigir conjuntos de condições projetadas especificamente para um dado sensor.

5 CONCLUSÕES

O presente trabalho abordou o problema de detecção de fogo ativo utilizando técnicas de aprendizado profundo. Foi utilizado um *dataset* com cobertura global, contendo 146.214 recortes de imagens extraídas do satélite Landsat-8, e também as máscaras produzidas por três algoritmos bem estabelecidos na literatura (SCHROEDER *et al.*, 2016; MURPHY *et al.*, 2016; KUMAR; ROY, 2018) e também a combinação desses algoritmos (intersecção e votação). Além disso, um segundo conjunto de dados também foi disponibilizado, contendo 9.044 recortes, com anotações manuais de fogo ativo feitas por um especialista humano. Esses dados foram utilizados para analisar como uma rede neural convolucional pode ser utilizada para aproximar as saídas produzidas pelos métodos tradicionais, bem como a combinação desses métodos.

O estudo realizado mostrou que a construção de um conjunto de condições para reconhecimento de fogo ativo é uma tarefa complexa, e pequenas variações nos limiares podem causar alterações nos resultados, gerando falsos positivos ou negativos. Em contrapartida, técnicas de aprendizado profundo são capazes de codificar regras mais complexas, mas ao custo de relações que são difíceis de entender e serem descritas para um ser humano.

Trabalhos futuros devem focar em explorar métodos alternativos para detecção de fogo ativo, explorando a hipótese que modelos treinados sobre imagens de um satélite podem se comportar bem para outros satélites. Também é possível desenvolver meios de combinar resultados de detecção de múltiplos satélites, com diferentes respostas, resoluções espaciais e ciclos de captura; visando melhorar a confiabilidade e produzir estimativas com maior resolução no espaço e no tempo.

REFERÊNCIAS

ARAÚJO, Fernando Moreira de; FERREIRA, Laerte Guimarães; ARANTES, Arielle Elias. Distribution Patterns of Burned Areas in the Brazilian Biomes: An Analysis Based on Satellite Data for the 2002-2010 Period. **Remote Sensing**, v. 4, n. 7, p. 1929–1946, 2012. ISSN 20724292.

BOSCHETTI, Mirco; Busetto, Lorenzo; MANFRON, Giacinto; LABORTE, Alice; ASILO, Sonia; PAZHANIVELAN, Sellaperumal; NELSON, Andrew. Phenorice: A method for automatic extraction of spatio-temporal information on rice crops using satellite data time series. **Remote Sensing of Environment (Elsevier)**, v. 194, p. 347 – 365, 2017. ISSN 0034-4257.

CARDIL, Adrián; MONEDERO, Santiago; RAMÍREZ, Joaquín; SILVA, Carlos Alberto. Assessing and reinitializing wildland fire simulations through satellite active fire data. **Journal of Environmental Management (Elsevier)**, v. 231, p. 996–1003, 2019. ISSN 0301-4797.

CERTINI, Giacomo. Effects of fire on properties of forest soils: A review. **Oecologia**, v. 143, n. 1, p. 1–10, 2005. ISSN 00298549.

CHAPELLE, Olivier; SCHLKOPF, Bernhard; ZIEN, Alexander. **Semi-Supervised Learning**. 1st. ed. [S.l.]: The MIT Press, 2010. ISBN 0262514125.

CHEN, Dong; LOBODA, Tatiana V.; HALL, Joanne V. A systematic evaluation of influence of image selection process on remote sensing- based burn severity indices in north american boreal forest and tundra ecosystems. **ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)**, v. 159, p. 63 – 77, 2020. ISSN 0924-2716.

CHINCHOR, Nancy; SUNDHEIM, Beth. Muc-5 evaluation metrics. *In: Conference on Message Understanding*. USA: Association for Computational Linguistics, 1993. (MUC5 '93), p. 69–78. ISBN 1558603360.

DUARTE, D.; NEX, F.; KERLE, N.; VOSSSELMAN, G. Satellite image classification of building damages using airborne and satellite image samples in a Deep Learning approach. **ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences**, IV-2, n. 2, p. 89–96, may 2018. ISSN 2194-9050. Disponível em: <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2/89/2018/>.

EVERINGHAM, Mark; GOOL, Luc; WILLIAMS, Christopher K.; WINN, John; ZISSERMAN, Andrew. The Pascal Visual Object Classes (VOC) Challenge. **International Journal of Computer Vision (IJCV)**, Kluwer Academic Publishers, USA, v. 88, n. 2, p. 303–338, jun. 2010. ISSN 0920-5691.

FLANNIGAN, M. D.; HAAR, T. H. Vonder. Forest fire monitoring using NOAA satellite AVHRR. **Canadian Journal of Forest Research**, v. 16, n. 5, p. 975–982, 1986.

FRITSCH, Jannik; KÜHNEL, Tobias; GEIGER, Andreas. A new performance measure and evaluation benchmark for road detection algorithms. *In: International IEEE Conference on Intelligent Transportation Systems*. [S.l.: s.n.], 2013. v. 1, n. 1, p. 1693–1700.

GARCIA, Alberto; ORTS-ESCOLANO, Sergio; OPREA, Sergiu; VILLENA-MARTINEZ, Victor; MARTINEZ-GONZALEZ, Pablo; GARCIA-RODRIGUEZ, Jose. A survey on deep learning techniques for image and video semantic segmentation. **Applied Soft Computing (Elsevier)**, v. 70, p. 41 – 65, 2018. ISSN 1568-4946.

GARGIULO, Massimiliano; DELL'AGLIO, Domenico Antonio Giuseppe; IODICE, Antonio; RICCIO, Daniele; RUELLO, Giuseppe. A CNN-Based Super-Resolution Technique for Active Fire Detection on Sentinel-2 Data. *In: PhotonIcs Electromagnetics Research Symposium (Spring)*. [S.l.: s.n.], 2019. abs/1906.10413, p. 418–426.

GIGLIO, Louis; DESCLOITRES, Jacques; JUSTICE, Christopher O.; KAUFMAN, Yoram J. An enhanced contextual fire detection algorithm for modis. **Remote Sensing of Environment (Elsevier)**, v. 87, n. 2, p. 273 – 282, 2003. ISSN 0034-4257.

GIGLIO, Louis; SCHROEDER, Wilfrid; JUSTICE, Christopher O. The collection 6 MODIS active fire detection algorithm and fire products. **Remote Sensing of Environment (Elsevier)**, v. 178, p. 31–41, 2016. ISSN 0034-4257.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: The MIT Press, 2016. ISBN 0262035618.

JI, Yimin; STOCKER, Erich. Seasonal, intraseasonal, and interannual variability of global land fires and their effects on atmospheric aerosol distribution. **Journal of Geophysical Research: Atmospheres**, v. 107, n. D23, p. ACH 10–1–ACH 10–11, 2002.

KAUFMAN, Yoram J.; JUSTICE, Christopher O.; FLYNN, Luke P.; KENDALL, Jackie D.; PRINS, Elaine M.; GIGLIO, Louis; WARD, Darold E.; MENZEL, W. Paul; SETZER, Alberto W. Potential global fire monitoring from eos-modis. **Journal of Geophysical Research: Atmospheres**, v. 103, n. D24, p. 32215–32238, 1998.

KHRYASHCHEV, Vladimir; LARIONOV, Roman. Wildfire Segmentation on Satellite Images using Deep Learning. **Moscow Workshop on Electronic and Networking Technologies, MWENT 2020 - Proceedings, IEEE**, p. 1–5, 2020.

KONDRATYEV, K. Ya.; DYACHENKO, L.N.; BINENKO, V.I.; CHERNENKO, A.P. Detection of Small Fires and Mapping of Large Forest Fires by Infrared Imagery. *In: International Symposium on Remote Sensing of Environment. [S.l.: s.n.]*, 1972. v. 1, n. 1, p. 1297–1303.

KUMAR, S. S.; ROY, D. P. Global operational land imager Landsat-8 reflectance-based active fire detection algorithm. **International Journal of Digital Earth**, Taylor & Francis, v. 11, n. 2, p. 154–178, 2018.

LANGFORD, Zachary; KUMAR, Jitendra; HOFFMAN, Forrest. Wildfire mapping in interior alaska using deep neural networks on imbalanced datasets. **IEEE International Conference on Data Mining Workshops, ICDMW**, IEEE, v. 2018-Novem, p. 770–778, 2019. ISSN 23759259.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 5 2015.

LEE, Thomas F.; TAG, Paul M. Improved Detection of Hotspots using the AVHRR 3.7-um Channel. **Bulletin of the American Meteorological Society**, American Meteorological Society, Boston MA, USA, v. 71, n. 12, p. 1722 – 1730, 1990.

LIN, Tsung-Yi; MAIRE, Michael; BELONGIE, Serge; HAYS, James; PERONA, Pietro; RAMANAN, Deva; DOLLÁR, Piotr; ZITNICK, C. Lawrence. Microsoft COCO: Common Objects in Context. *In: FLEET, David; PAJDLA, Tomas; SCHIELE, Bernt; TUYTELAARS, Tinne (Ed.). European Conference on Computer Vision (ECCV)*. Cham: Springer, 2014. p. 740–755. ISBN 978-3-319-10602-1.

LITJENS, Geert; KOOI, Thijs; BEJNORDI, Babak Ehteshami; SETIO, Arnaud Arindra Adiyoso; CIOMPI, Francesco; GHAFOORIAN, Mohsen; van der Laak, Jeroen A.W.M.; van Ginneken, Bram; SÁNCHEZ, Clara I. A survey on deep learning in medical image analysis. **Medical Image Analysis (Elsevier)**, v. 42, p. 60 – 88, 2017. ISSN 1361-8415.

LIZUNDIA-LOIOLA, Joshua; PETTINARI, M. Lucrecia; CHUVIECO, Emilio. Temporal Anomalies in Burned Area Trends: Satellite Estimations of the Amazonian 2019 Fire Crisis. **Remote Sensing**, v. 12, n. 1, p. 151, 2020. ISSN 2072-4292.

LOYD, Charlie. **Landsat 8 Bands**. NASA, 2013. Acesso em: 03 de Agosto de 2021. Disponível em: <https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands>.

LYU, Yecheng; BAI, Lin; HUANG, Xinming. Road Segmentation using CNN and Distributed LSTM. *In: IEEE International Symposium on Circuits and Systems. [S.l.: s.n.]*, 2019. v. 1, n. 1, p. 1–5.

- MA, Lei; LIU, Yu; ZHANG, Xueliang; YE, Yuanxin; YIN, Gaofei; JOHNSON, Brian Alan. Deep learning in remote sensing applications: A meta-analysis and review. **ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)**, v. 152, p. 166 – 177, 2019. ISSN 0924-2716.
- MAIER, Stefan W.; RUSSELL-SMITH, Jeremy; EDWARDS, Andrew C.; YATES, Cameron. Sensitivity of the modis fire detection algorithm (mod14) in the savanna region of the northern territory, australia. **ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)**, v. 76, p. 11 – 16, 2013. ISSN 0924-2716. Terrestrial 3D modelling.
- MALAMBO, Lonesome; HEATWOLE, Conrad D. Automated training sample definition for seasonal burned area mapping. **ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)**, v. 160, p. 107 – 123, 2020. ISSN 0924-2716.
- MATEO-GARCÍA, Gonzalo; LAPARRA, Valero; LÓPEZ-PUIGDOLLERS, Dan; GÓMEZ-CHOVA, Luis. Transferring deep learning models for cloud detection between landsat-8 and proba-v. **ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)**, v. 160, p. 1 – 17, 2020. ISSN 0924-2716.
- MATSON, Michael; HOLBEN, Brent. Satellite detection of tropical burning in brazil. **International Journal of Remote Sensing**, Taylor & Francis, v. 8, n. 3, p. 509–516, 1987.
- MINETTO, Rodrigo; SEGUNDO, Maurício Pamplona; SARKAR, Sudeep. Hydra: An ensemble of convolutional neural networks for geospatial land classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 57, n. 9, p. 6530–6541, 2019.
- MORISSETTE, Jeffrey T.; GIGLIO, Louis; CSISZAR, Ivan; JUSTICE, Christopher O. Validation of the modis active fire product over southern africa with aster data. **International Journal of Remote Sensing**, Taylor & Francis, v. 26, n. 19, p. 4239–4264, 2005.
- MURPHY, Sam W.; DE SOUZA FILHO, Carlos Roberto; WRIGHT, Rob; SABATINO, Giovanni; CORREA PABON, Rosa. Hotmap: Global hot target detection at moderate spatial resolution. **Remote Sensing of Environment (Elsevier)**, v. 177, p. 78 – 88, 2016. ISSN 0034-4257.
- PAOLETTI, Mercedes Eugenia; HAUT, Juan Mario; PLAZA, Javier; PLAZA, Antonio. Deep learning classifiers for hyperspectral imaging: A review. **ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)**, v. 158, p. 279 – 317, 2019. ISSN 0924-2716.
- PEREIRA, Gabriel Henrique de Almeida; FUSIOKA, Andre Minoro; NASSU, Bogdan Tomoyuki; MINETTO, Rodrigo. 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>.

PETERSSON, Henrik; GUSTAFSSON, David; BERGSTROM, David. Hyperspectral image analysis using deep learning — a review. *In: International Conference on Image Processing Theory, Tools and Applications*. [S.l.: s.n.], 2016. v. 1, n. 1, p. 1–6.

PINHEIRO, Pedro O O.; COLLOBERT, Ronan; DOLLAR, Piotr. Learning to Segment Object Candidates. *In: CORTES, C.; LAWRENCE, N.; LEE, D.; SUGIYAMA, M.; GARNETT, R. (Ed.). Advances in Neural Information Processing Systems (NIPS)*. [S.l.: s.n.], 2015. v. 28, p. 1990–1998.

PRIYA, R. Shanmuga; VANI, K. Deep learning based forest fire classification and detection in satellite images. **Proceedings of the 11th International Conference on Advanced Computing, ICoAC 2019**, p. 61–65, 2019.

RONNEBERGER, O.; PFISCHER; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *In: Medical Image Computing and Computer-Assisted Intervention*. [S.l.]: Springer, 2015. (LNCS, v. 9351), p. 234–241.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning Representations by Back-propagating Errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986.

SCHROEDER, Wilfrid; OLIVA, Patricia; GIGLIO, Louis; CSISZAR, Ivan A. The new viirs 375m active fire detection data product: Algorithm description and initial assessment. **Remote Sensing of Environment (Elsevier)**, v. 143, p. 85 – 96, 2014. ISSN 0034-4257.

SCHROEDER, Wilfrid; OLIVA, Patricia; GIGLIO, Louis; QUAYLE, Brad; LORENZ, Eckehard; MORELLI, Fabiano. Active fire detection using Landsat-8/OLI data. **Remote Sensing of Environment (Elsevier)**, v. 185, p. 210 – 220, 2016. ISSN 0034-4257.

TOAN, Nguyen Thanh; Thanh Cong, Phan; Viet Hung, Nguyen Quoc; JO, Jun. A deep learning approach for early wildfire detection from hyperspectral satellite images. *In: 2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA)*. IEEE, 2019. p. 38–45. ISBN 978-1-7281-3118-4. Disponível em: <https://ieeexplore.ieee.org/document/8932740/>.

TREKIN, Alexey; NOVIKOV, German; POTAPOV, Georgy; IGNATIEV, Vladimir; BURNAEV, Evgeny. Satellite imagery analysis for operational damage assessment in emergency situations. **Lecture Notes in Business Information Processing**, v. 320, p. 347–358, 2018. ISSN 18651348.

USGS. **Landsat Collection 1 Level-1 Quality Assessment Band**. Acesso em: 03 de Agosto de 2021. Disponível em: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-1-level-1-quality-assessment-band?qt-science_support_page_related_con=0#qt-science_support_page_related_con.

USGS. **Using the USGS Landsat Level-1 Data Product**. Acesso em: 03 de Agosto de 2021. Disponível em: <https://www.usgs.gov/core-science-systems/nli/landsat/using-usgs-landsat-level-1-data-product>.

WANG, Zhihao; CHEN, Jian; HOI, Steven C.H. Deep learning for image super-resolution: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 1, n. 1, p. 1–22, 2020.

YALOVEHA, Vladyslav; HLAVCHEVA, Daria; PODOROZHNIAK, Andrii. Usage of Convolutional Neural Network for multispectral image processing applied to the problem of detecting fire hazardous forest areas. **Advanced Information Systems**, v. 3, n. 1, p. 116–120, may 2019. ISSN 2522-9052. Disponível em: <http://ais.khpi.edu.ua/article/view/2522-9052.2019.1.19>.

YAO, Guangle; LEI, Tao; ZHONG, Jiandan. A review of convolutional-neural-network-based action recognition. **Pattern Recognition Letters (Elsevier)**, v. 118, p. 14 – 22, 2019. ISSN 0167-8655.

YUAN, Qiangqiang; SHEN, Huanfeng; LI, Tongwen; LI, Zhiwei; LI, Shuwen; JIANG, Yun; XU, Hongzhang; TAN, Weiwei; YANG, Qianqian; WANG, Jiwen; GAO, Jianhao; ZHANG, Liangpei. Deep learning in environmental remote sensing: Achievements and challenges. **Remote Sensing of Environment (Elsevier)**, v. 241, p. 111716, 2020. ISSN 0034-4257.

ZHU, Xiao Xiang; TUIA, Devis; MOU, Lichao; XIA, Gui-Song; ZHANG, Liangpei; XU, Feng; FRAUNDORFER, Friedrich. Deep learning in remote sensing: A comprehensive review and list of resources. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 4, p. 8–36, 2017.