

Active Fire Detection in Forests with 5 Deep Learning Techniques

Marcelo Kuchar Matte¹, José Marcato Junior², Newton Loebens³ and Hemerson Pistori⁴

^{1,2,4} Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil

^{3,4} Universidade Católica Dom Bosco, Campo Grande, Brazil

Email: ¹marcelokuchar@gmail.com

Resumo—Identifying active fires in natural and difficult-to-access environments is a real and current problem, especially in recent years. One of the simplest ways to monitor regions of difficult access is the use of satellite images in their various patterns, such as multi-spectral and hyper-spectral images. In this way, we used 5 CNNs (Faster R-CNN, ATSS, RetinaNet, VNet, and SABL) to detect active fires in satellite images of the Pantanal region in May 2020. In preliminary results, it was possible to reach values greater than 50% of mAP50 in the detections performed. Studies are still ongoing in the search for higher precisions.

I. INTRODUÇÃO

Nos últimos anos os danos ao meio ambiente causados por incêndios nos principais biomas brasileiros vêm ganhando destaque na sociedade, não sendo possível até o momento calcular o impacto gerado nas comunidades e ecossistemas naturais. Detectar esses incêndios em florestas e ambientes naturais isolados pode se mostrar de grande utilidade, permitindo assim gerar alertas adequados para os moradores da região e para os órgãos públicos responsáveis.

Jain *et al.*, em [1], apresenta 6 problemas, encontrados na literatura, relativos a incêndios em ambientes naturais, sendo eles: (1) caracterização de combustíveis e detecção de incêndio e mapeamento; (2) tempo de incêndio e mudança climática; (3) ocorrência, suscetibilidade e risco de incêndio; (4) previsão do comportamento do fogo; (5) efeitos de fogo; e (6) gerenciamento de fogo.

Uma das principais ferramentas atuais que permite a detecção de incêndios em regiões isoladas é as imagens ópticas geradas a partir de satélites. Uma das grandes vantagens do uso de imagens de satélites é a possibilidade de monitoramento periódico de regiões remotas. As imagens fornecidas por satélites colaboram com uma detecção precoce de situações onde o fogo pode ocorrer com maior gravidade e regiões com incêndios ativos, permitindo assim a prevenção de desastres de maior escala.

Buscando aprimorar a precisão da detecção de incêndios, várias técnicas de Aprendizado de Máquina (AM) são utilizadas, principalmente as técnicas de AM supervisionadas, que buscam generalizar uma expressão de conceito a partir de um conjunto de exemplos (instâncias ou padrões) previamente classificados. Entre as técnicas de AM supervisionada, um grupo de algoritmos popularmente conhecidos como Aprendizado Profundo (Deep Learning) vem se destacando, principalmente as Redes Neurais Convolucionais (CNNs). De forma

simplicista, pode-se dizer que uma CNN utiliza de neurônios artificiais, dividido em múltiplas camadas, para detectar padrões em imagens (*e.g.* linhas e formas) montando um mapa de atributos de uma determinada classe. O uso de técnicas de Aprendizado Profundo para detecção de incêndios é relativamente recente e, geralmente, os algoritmos são treinados em imagens que contêm áreas com fogo ativo, fumaça ou regiões já incendiadas [1] [2] [3] [4].

Imagens obtidas a partir de satélites combinam, além das informações do espectro do visível, tradicionalmente conhecido como RGB, informações de um conjunto variado de espectros (*e.g.* infravermelho, ultravioleta) capturados por sensores. As informações adquiridas, comumente conhecida como bandas, registram as propriedades físicas dos recursos em solo e, em geral, a posição geográfica da imagem. Diferentes materiais apresentam respostas diferentes em cada espectro, permitindo, assim, apresentar melhor acurácia nas predições de incêndios [5].

Detecção de objetos busca localizar a presença de um determinado objeto, *e.g.* padrão na imagem, delimitar sua área e predizer sua classe, *e.g.* foco de incêndio. O desempenho dos modelos de detecção de objetos, em geral, são avaliados usando o erro médio das predições em relação ao rótulo, previamente definido, do objeto e a interseção da união (IoU), uma medida de sobreposição, entre a caixa delimitadora do objeto anotado e a caixa delimitadora predita pelo modelo.

Com o objetivo de verificar o uso de 5 CNNs para detectar focos de incêndio em imagens de satélite utilizando, inicialmente, as bandas RGB, esse artigo se organiza em cinco seções que, além dessa seção de introdução, relembra, na Seção II, os trabalhos relacionados as CNNs escolhidas e cita trabalhos correlatos ao tema incêndio em ambientes naturais e detecção desses incêndios utilizando Redes Neurais Convolucionais. A Seção III apresenta os materiais e métodos utilizados nos experimentos realizados. Por fim, a Seção IV apresenta resultados e discussões dos dados obtidos. A Seção V apresenta os principais pontos do trabalho realizado bem como as conclusões, que foram subsidiadas pelos resultados dos experimentos.

II. TRABALHOS RELACIONADOS

A partir de 2012 as Redes Neurais Convolucionais (CNNs) tornaram-se dominantes nas tarefas de Visão Computacional (VC), principalmente, devido ao êxito da CNN *AlexNet* [6]

que apresentou performance superior em um concurso de classificação de imagens (*ImageNet Large Scale Visual Recognition Challenge* - ILSVRC-2012).

Em 2014, Girshick *et al.* apresentou a proposta *R-CNN* (*Regions with CNN features*) [7] que era composta de três módulos: (1) Proposta de Região, (2) Extrator de Atributos, (3) Classificador. Essa técnica, que propõem regiões candidatas com maior potencial de possuir objetos, chamada de *busca seletiva*, aplica um extrator de atributos e alimenta o classificador, em sua versão inicial, Máquina de Vetores de Suporte (SMV). Em 2015 Girshick apresenta uma extensão da *R-CNN* que deu o nome de *Fast R-CNN* [8]. A nova proposta envia a imagem original para um CNN que gera um mapa de atributos e submete esse mapa com o intuito de identificar as regiões de interesse. A terceira proposta de Girshick *et al.*, em 2016, *Faster R-CNN* [9], uma melhoria da *Fast R-CNN*, deixa de usar a busca seletiva e faz uso da RPN (Rede de Propostas de Região) para gerar as propostas de regiões. Uma RPN, em média, gera duas mil possíveis regiões de interesses por imagem, que traz embutido um custo de processamento. De forma simplificada, a *Faster R-CNN* é um detector de dois estágios, sendo o primeiro estágio usado para propor regiões onde há maior probabilidade de existir um objeto, filtrando a maioria das regiões que representam o fundo da imagem, *i.e.* regiões não-annotadas previamente como objeto. Para cada região de interesse é gerado $A = 9$ âncoras, *i.e.* caixas delimitadoras de tamanhos variados sobre a região de interesse. O segundo módulo é utilizado como um detector e classificador das âncoras propostas pelo primeiro estágio.

Em 2017, Lin *et al.* apresentou a implementação do detector de um estágio chamado *RetinaNet* [10]. Um detector de um estágio aplica a fase de detecção diretamente nas âncoras inicialmente geradas, *i.e.* não realiza a fase de seleção das âncoras com maior probabilidade de conter um objeto. A *RetinaNet* é uma combinação de uma rede *backbone* e duas sub-redes de tarefas específicas, onde a rede *backbone* é uma implementação de uma rede FPN (Rede em Pirâmide de Atributos) baseada na arquitetura *ResNet*, enquanto a FPN gera, a partir da imagem de entrada, um mapa de atributos em várias escalas com conexões laterais e cada um dos níveis da pirâmide pode ser usado para detectar objetos. À saída do *backbone* são anexadas duas sub-redes, uma de classificação das âncoras e outra para regressão das caixas que delimitam tais âncoras. Ambas sub-redes compartilham uma mesma estrutura, porém seus parâmetros são separados. A *RetinaNet* implementou a função de custo *Perda Focal*, que é uma atualização da função de custo *Entropia Cruzada*. A *Perda Focal* busca otimizar o problema de desequilíbrio de classes atribuindo mais peso a âncoras difíceis de serem classificadas, introduzindo, assim, um fator de modulação que reduz a perda em âncoras fáceis de serem classificadas e aumenta a importância de falsos negativos e falsos positivos, permitindo, dessa forma, que a rede seja mais precisa.

Zhang *et al.*, em 2020, propôs um novo método de definir se as âncoras são classificadas como objeto ou fundo, chamado de ATSS (Seleção de amostra de treinamento adaptativo) [11].

A ATSS modifica a forma de definir como as âncoras são classificadas após a saída da FPN. A ATSS propõem que em vez de usar um valor fixo para o limite do IoU, que defini se a âncora contém ou não um objeto, usar um modelo de divisão automático de acordo com as características estatísticas dos objetos.

Ainda em 2020, Wang *et al.* propôs a CNN SABL (Localização de limite com reconhecimento lateral) [12] que implementa uma modificação na regressão da caixa delimitadora convencional. A partir do mapa de atributos gerado pela FPN, para cada proposta gerada, o SABL aprende a prever e classificar um caixa delimitadora baseada na âncora correspondente.

Em [13] é apresentada a proposta *VarifocalNet* (VfNet) realizada por Zhang *et al.*. A VfNet tem como base a CNN FCOS (detecção de objetos de um estágio totalmente convolucional) + ATSS descrita em [11]. A rede FCOS tem como pontos principais ser um detector de um estágio e não utilizar âncoras. A FCOS é construída sobre uma FPN e usa extrator de atributos, *e.g.* ResNet50, gerando, assim, uma pirâmide de atributos e utiliza o método ATSS para definir automaticamente se uma amostra é positiva ou negativa. Comparado com a FCOS + ATSS, há três novos componentes na VfNet: A função de custo Varifocal; o novo modelo em estrela de representar a caixa delimitadora; o refinamento da caixa delimitadora. A função de custo Varifocal, uma atualização da função de custo Perda Focal, trata os exemplos positivos e negativos assimetricamente, de forma que o peso não diminui da mesma forma para os exemplos positivos e negativos.

Nos últimos anos, o estudo de imagens de satélites utilizando AM, e principalmente CNNs, para detecção de incêndios em florestas e ambientes naturais, vem crescendo, como pode ser visto em [14], [15], [16], [17], [18], [19]. Jain *et al.* [1] apresenta as principais técnicas de AM e Aprendizado Profundo utilizados até o ano de 2020. O mesmo autor, em [5], apresenta o uso de CNNs em imagens hiper-espectrais de satélites, permitindo assim uma predição mais precisa.

III. MATERIAIS E MÉTODOS

Para a realização dos experimentos descritos neste trabalho, foi utilizado um conjunto de imagens do satélite CBERS-4A. A seleção e organização desse conjunto de imagens foi realizada e primeiramente apresentada em [20]. Foram escolhidas imagens onde foram detectados focos de incêndios no ano de 2020, no bioma do Pantanal. As imagens originais foram convertidas para imagens de resolução 256×256 pixels. Ainda, neste trabalho inicial, somente as bandas do visível (RGB) são consideradas. As imagens foram previamente anotadas utilizando o sistema *roboflow* (www.roboflow.com), ferramenta de anotação em imagens que disponibiliza acesso há varias funções de forma gratuita. O *roboflow* permite exportar os dados em vários formatos, entre eles o 'COCO', padrão utilizado nos experimentos descritos neste trabalho.

Foi aplicado o modelo de anotação, conhecido como 'caixas delimitadora', de tamanhos variados, onde foram anotados os pontos de incêndios visíveis nas imagens escolhidas, como



Figura 1: Exemplo de identificação e anotação, via *RoboFlow*, dos pontos de incêndio na imagem de Satélite.

Tabela I: Taxa de aprendizado para o modelo de treinamento.

Modelo	Taxa de aprendizado
Faster R-CNN	0.01
ATSS	0.01
RetinaNet	0.001
VfNet	0.01
SABL	0.001

exemplifica a Figura 1. O conjunto utilizado possui 576 imagens, tendo sido dividida em conjuntos de treino e teste na proporção de 70% e 30%, respectivamente, onde 10% das imagens de treinamento foram utilizadas como conjunto de validação. Foi utilizado a validação cruzada estratificada em 5 dobras.

O *framework MMDetection* [21] foi aplicado para auxiliar a execução das redes definidas. O *MMDetection* é uma ferramenta de código aberto baseada em *PyTorch* que traz um conjunto de pré-configurações para um grande conjunto de CNNs. Na execução dos experimentos, salvo os parâmetros descritos nas Tabela I e Tabela II, foram empregadas as configurações padrão disponibilizadas pelo *MMDetection* para as CNNs *Faster R-CNN*, *RetinaNet*, *ATSS*, *VfNet* e *SABL*.

A Tabela I apresenta as taxas de aprendizado definidas para cada uma das arquiteturas. A CNN *RetinaNet* e *SABL* foram inicialmente treinadas na taxa 0,01, porém, foi necessário modificar a taxa para 0,001 devido a incapacidade das duas arquiteturas aprenderem com o valor inicialmente estabelecido. Após testes empíricos, um comportamento 'asintomático' foi constatado em relação à função de custo em, aproximadamente, 10 épocas. Dessa forma, nos experimentos descritos nesse trabalho, foram definidas 50 épocas como valor padrão de execução, de forma a minimizar o risco de encerrar a execução do algoritmo antes de se atingir um valor ótimo ou razoável para a função de custo.

IV. RESULTADOS E DISCUSSÕES

A partir dos experimentos realizados foi produzida a Tabela III com a média das 5 dobras executadas, contendo as informações do mAP50, do desvio padrão (DP) e da mediana

Tabela II: Valores de hiperparâmetro usados nos experimentos para todas as redes.

	Hiperparâmetro	Valor
#1	Épocas de Treinamento	50
#2	Tamanho do Patch	Tamanhos Variados
#3	Tamanho do Batch	50
#4	Aumento de Dados	Nenhum
#5	Divisão do Conjunto de Teste	Aleatório com 5 rep
#6	Tamanho do Conjunto de Teste	30%
#7	Divisão de Validação	10%

dos resultados das execuções para cada uma das arquiteturas utilizadas. Apesar dos dados ainda serem preliminares, já foi possível conseguir um mAP50 médio acima de 50% para 4 das 5 arquiteturas utilizadas.

Tabela III: Para cada CNN, a média (mAP > 0.50) e o desvio padrão (DP) do resultado das 5 dobras estratificadas na validação cruzada e sua mediana.

CNN	mAP50	DP	MEDIANA
Faster	0,57	0,04	0,56
ATSS	0,54	0,03	0,55
RetinaNet	0,45	0,07	0,48
VfNet	0,53	0,25	0,52
SABL	0,54	0,04	0,54

O histograma apresentado na Figura 2 traz a distribuição do número de focos anotados em cada uma das imagens de satélite selecionadas para compor a base de imagens. Empiricamente, é possível perceber que a medida que o fogo se espalha pela região, múltiplos focos são encontrados por imagem. Tal distribuição dos focos pela imagem acabou gerando um complicador no momento de definir as caixas delimitadoras anotadas. Esse complicador está na dificuldade de distinguir os focos de incêndios ativos em grandes áreas recém queimadas. Outro complicador encontrado é a fumaça que cobre parte das regiões, onde, muitas vezes, há vários Kilômetros de foco de incêndio, dificultando assim a detecção nas regiões com foco de incêndio ativo, como pode ser visto na Figura 1.

A Figura 3 apresenta as informações sobre a avaliação mAP50 das execuções, mostrando as variações que ocorreram durante as 5 dobras executadas. É possível perceber que nos experimentos realizados, considerando as anotações previamente realizadas manualmente, que a arquitetura *Faster R-CNN* apresentou os melhores resultados de predição. A Figura 4 apresenta o RMSE (Raiz Do Erro Quadrático Médio) das execuções realizadas para cada uma das arquiteturas.

A Figura 5 apresenta os dados com a predição em relação ao número de caixas delimitadoras anotadas por imagem, para a melhor execução das 5 dobras de cada arquitetura, além de apresentar o RMSE, MAE e r^2 relativo a execução.

Na Figura 6 pode-se visualizar 4 exemplos das predições realizadas, sendo a caixa delimitadora em vermelho a anotação realizada manualmente sobre um foco de incêndio. A caixa delimitadora verde é a predição realizada pela arquitetura.

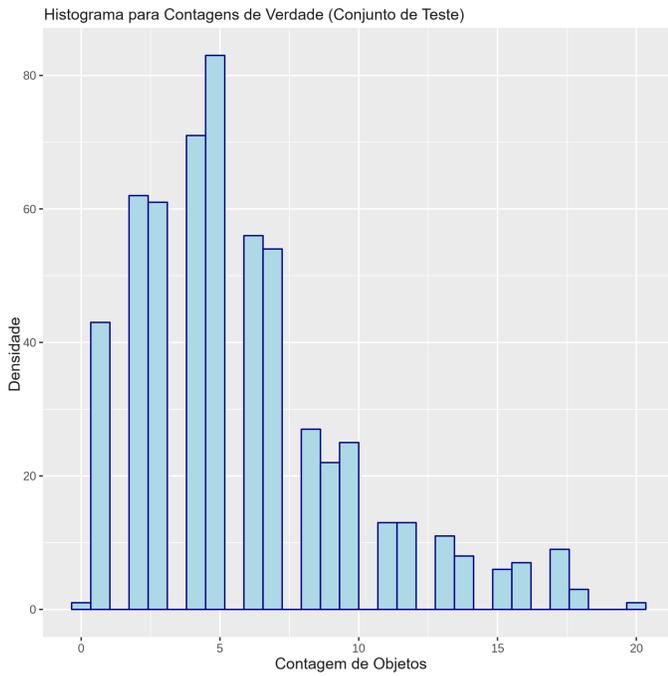


Figura 2: Histograma com o número de caixas delimitadoras anotadas por imagem.

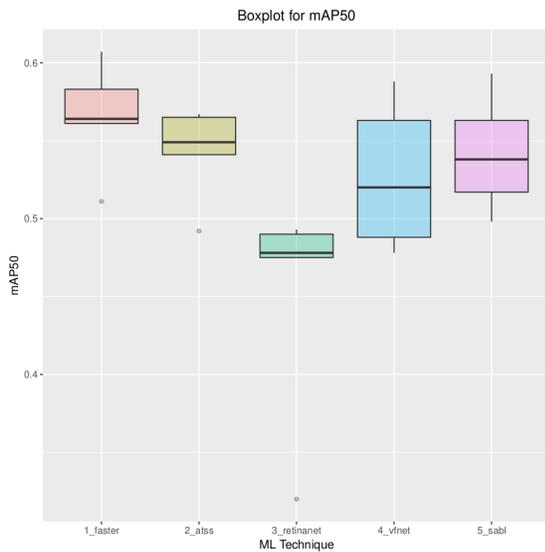


Figura 3: Boxplot da mAP50 dos 5 arquiteturas.

Neste exemplo são mostradas as predições realizadas pela *Faster R-CNN*, rede que apresentou os melhores resultados nos experimentos realizados. Na imagem (a) pode-se perceber que uma das predições ocorreu sobre uma região com nuvens e duas caixas delimitadoras anotadas não tiveram seus pares correspondentes. Na imagem (b) duas caixas preditas ocorreram sobre áreas com nuvens e novamente duas caixas delimitadoras anotadas não tiveram seus pares correspondentes. Na Imagem (c) não foi realizada nenhuma predição sobre a imagem e, por fim, na Imagem (d) ocorreu a detecção de todas as caixas

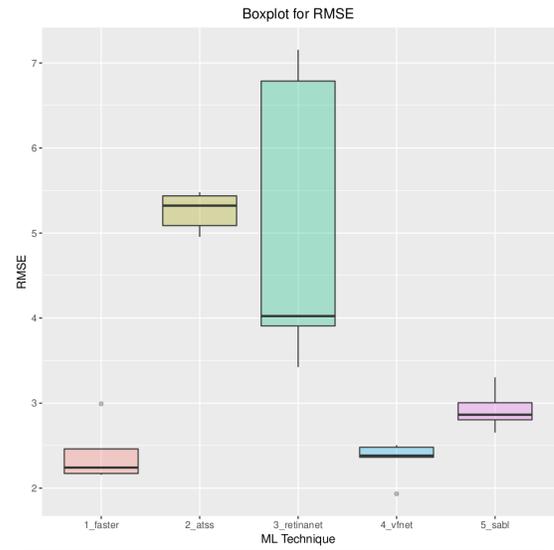


Figura 4: RMSE das execuções das 5 arquiteturas.

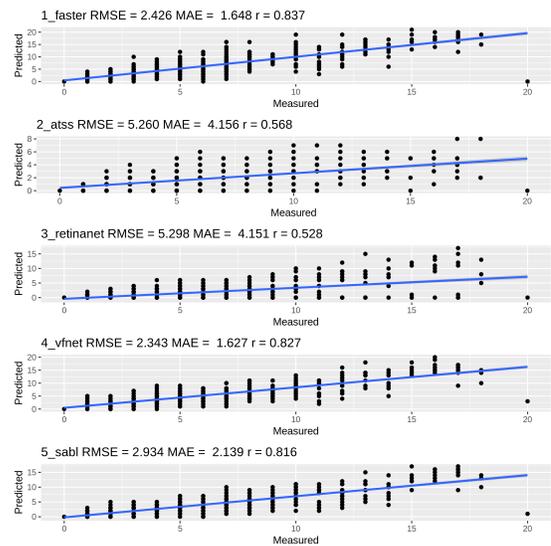


Figura 5: Gráficos XY para prever e medir os focos de incêndios decorrentes da melhor execução para cada arquitetura. As linhas de regressão são mostradas em azul.

delimitadoras anotadas.

V. CONCLUSÃO

^{1 2} Apesar dos estudos estarem em seu início, já é possível determinar a viabilidade do uso de técnicas de visão computacional, especificamente técnicas de aprendizagem profunda, para detectar focos de incêndios ativos em ambientes naturais e de difíceis acesso. Quatro das cinco técnicas apresentaram mAP50 superior a 50% em suas primeiras execuções. Espera-se que na medida em que novos parâmetros forem testados e

¹link Vídeo Apresentação: <https://youtu.be/fp5dMwae6wA>

²link Código: encurtador.com.br/HL068

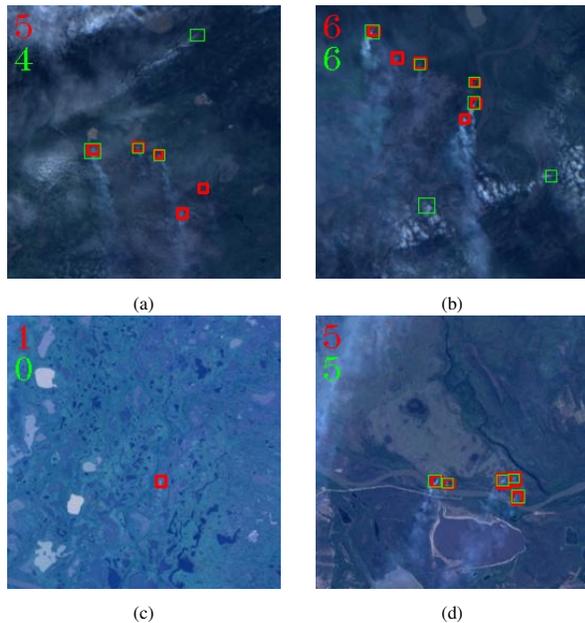


Figura 6: Quatro imagens exemplificando o resultado da predição da CNN *Faster R-CNN*, modelo que apresentou melhores resultados em mAP50, sendo em vermelho as caixas delimitadoras anotadas manualmente e em verde a predição realizada pelo modelo.

seja realizado um refinamento no conjunto de imagens anotadas, sejam encontrados resultados ainda mais promissores.

AGRADECIMENTOS

Este trabalho foi financiado pela Universidade Católica Dom Bosco, pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do estado de Mato Grosso do Sul (FUNDECT).

REFERÊNCIAS

- [1] P. Jain, S. C. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," *Environmental Reviews*, vol. 28, no. 4, pp. 478–505, 2020.
- [2] R. Ba, C. Chen, J. Yuan, W. Song, and S. Lo, "Smokenet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention," *Remote Sensing*, vol. 11, no. 14, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/14/1702>
- [3] G. H. de Almeida Pereira, A. M. Fusioka, B. T. Nassu, and R. Minetto, "Active fire detection in landsat-8 imagery: a large-scale dataset and a deep-learning study," 2021.
- [4] S. Jana and S. K. Shome, "Hybrid ensemble based machine learning for smart building fire detection using multi modal sensor data," *Fire Technology*, vol. 59, no. 2, pp. 473–496, 2023.
- [5] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, 2021.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

- [8] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
- [9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.
- [10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [11] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9759–9768.
- [12] J. Wang, W. Zhang, Y. Cao, K. Chen, J. Pang, T. Gong, J. Shi, C. C. Loy, and D. Lin, "Side-aware boundary localization for more precise object detection," in *European Conference on Computer Vision*. Springer, 2020, pp. 403–419.
- [13] H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf, "Varifocalnet: An iou-aware dense object detector," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8514–8523.
- [14] D. Hamilton, E. Levandovsky, and N. Hamilton, "Mapping burn extent of large wildland fires from satellite imagery using machine learning trained from localized hyperspatial imagery," *Remote Sensing*, vol. 12, no. 24, p. 4097, 2020.
- [15] X. Hu, Y. Ban, and A. Nascetti, "Uni-temporal multispectral imagery for burned area mapping with deep learning," *Remote Sensing*, vol. 13, no. 8, p. 1509, 2021.
- [16] V. Khryashchev and R. Larionov, "Wildfire segmentation on satellite images using deep learning," in *2020 Moscow Workshop on Electronic and Networking Technologies (MWENT)*, 2020, pp. 1–5.
- [17] P. Pednekar, A. Srivastava, and A. Jadhav, "Fire detection using transfer learning and pre-trained model," in *2023 3rd International Conference on Intelligent Technologies (CONIT)*. IEEE, 2023, pp. 1–4.
- [18] M. M. Pinto, R. M. Trigo, I. F. Trigo, and C. C. DaCamara, "A practical method for high-resolution burned area monitoring using sentinel-2 and viirs," *Remote Sensing*, vol. 13, no. 9, p. 1608, 2021.
- [19] R. Singh, S. Sharma, S. Sharma, and S. Kaushik, "Real-time fire detection system based on cnn using tensorflow and opencv," *Journal of Data Acquisition and Processing*, vol. 38, no. 2, p. 723, 2023.
- [20] L. Higa, J. Marcatto Junior, T. Rodrigues, P. Zamboni, R. Silva, L. Almeida, V. Liesenberg, F. Roque, R. Libonati, W. N. Gonçalves *et al.*, "Active fire mapping on brazilian pantanal based on deep learning and cbers 04a imagery," *Remote Sensing*, vol. 14, no. 3, p. 688, 2022.
- [21] K. Chen, J. Wang, and *et al.*, "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.