

# Detecção de fogo e fumaça em vídeos usando remoção de fundo e Redes Neurais Convolucionais

Mayla Toshimi Nagai and Bruno M N Souza

UENP - Universidade Estadual do Norte do Paraná, Brasil

Centro de Ciências Tecnológicas

Email:may.toshimi@gmail.com, brunomiguel@uenp.edu.br

**Abstract**—In open environments, when there is a low relative humidity and excessive heat, a small spark can be the trigger for large wildfires to happen. Facing the unpredictability of the fire, detecting it early is important to streamline the fire fighting and then minimize its consequences. Research with convolutional neural networks - CNN - has increased exponentially with the use of various technologies for the detection of fire. With that, this paper presents a fire and smoke detection method on videos based on removing background and then applying a CNN to detect it. As preliminary results, the proposed process was able to achieve an accuracy of 92,73% in the Fire Smoke Detection Network (FSDN) architecture and 94,88% in the Xception architecture. These results are similar to the ones found in literature showing that apply CNN after background motion subtraction is a good way to solve the fire and smoke detection problem.

**Resumo**—Em ambientes abertos, quando há baixa umidade relativa do ar e calor excessivo uma pequena faísca pode ser o gatilho para que incêndios de grandes proporções aconteçam. Diante da imprevisibilidade do surgimento do fogo, detectá-lo de forma precoce é de suma importância para agilizar o combate ao incêndio e então minimizar suas consequências. As pesquisas com as Redes Neurais Convolucionais - CNN (*Convolutional Neural Network*) - vem aumentando de forma exponencial com utilizações de diversas tecnologias para a detecção de fogo. Desta maneira, o presente trabalho apresenta um processo de detecção de fogo e fumaça em vídeos baseado na utilização de métodos de remoção de fundo e redes neurais convolucionais para a detecção. Como resultados preliminares, o processo proposto foi capaz de alcançar uma acurácia de 92,73% na arquitetura FSDN-Fire Smoke Detection Network e 94,88% utilizando a arquitetura Xception. Resultados similares aos encontrados na literatura utilizando a mesma base de Vídeos mostrando que aliar a utilização de CNN após a remoção de fundo em vídeos mostra-se uma estratégia promissora para a detecção de fogo e fumaça.

## I. INTRODUÇÃO

O fogo é um elemento dominado pelo ser humano há mais de 1,5 milhões de anos [1] melhorando vários aspectos da vida tais como: alimentação, conforto/aquecimento e locomoção. No entanto, o fogo descontrolado gera grandes impactos negativos tanto ecológicos quanto financeiros, somente no ano de 2019 foram registrados um prejuízo de cerca de U\$ 6,9 bilhões e 2.620 mortes causadas pelos incêndios [2]. Para o combate ao fogo é necessário inibir pelo menos um dos elementos do triângulo da combustão, composto de: combustível, calor e o comburente (oxigênio) [3]. Porém, quando se trata de ambientes abertos, o controle de variáveis como calor e comburente é praticamente inviável. Neste sentido, quanto

mais precoce for realizada a detecção de incêndios e a rápida ação das autoridades menor será o impacto ecológico.

Existem várias aplicações de tecnologias para a detecção de fogo precoce, como sensores térmicos, ópticos e químicos [1]. Contudo, o uso de sensores térmicos e químicos são limitados pela necessidade de proximidade para melhor reconhecimento. Além disso, em grandes áreas, o custo destes sensores são muito elevados. Desta maneira, utilizar o processamento de imagens a partir de sensores ópticos (câmeras de vigilância, por exemplo) mostra-se uma opção viável. Entretanto a identificação do fogo é complexa, uma vez que não há um padrão de forma, cor ou volume, já que este é influenciado pelo tipo de combustível e do comburente [4].

Na literatura existem vários métodos de detecção de fogo e fumaça em imagens estáticas e vídeos, como os que utilizam de regras aplicadas em espaços de cor como: RGB [5], YCbCr [6], YUV [7], HSI [8], além de regras que utilizam-se de composição de canais de vários espaços de cor, como RGB com HSI [9]. Atualmente, o uso do aprendizado profundo (*deep learning*) para o estudo da detecção de fogo vem aumentando [10], com a utilização de diferentes arquiteturas de CNN (*Convolutional Neural Networks*) como *Faster Region Convolutional Neural Network (Faster R-CNN)* [11], CNNs baseadas na arquitetura *Inception-v3* [10], ou mesmo com o uso de redes convolucionais profundas (*Very Deep Convolutional Networks*) no caso da arquitetura VGG16 juntamente com a arquitetura *MobileNet* [12].

Em Dua et. al. [12], há o uso de uma arquitetura de rede convolucional profunda VGG16 juntamente com a MobileNet, sendo aplicadas em ambas uma base de dados com 2.515 imagens (treino 1785 imagens e teste 730 de imagens). Para o aumento da acurácia foi utilizada a técnica de transferência de aprendizagem obtendo como resultado uma melhor performance das redes convolucionais profundas, porém existe a possibilidade de *overfitting*. Em outro estudo, Priya et. al. [10] utilizaram uma arquitetura baseada na CNN *Inception-v3* juntamente com uma matriz de confusão para classificar imagens de fogo em satélites. Utilizando de uma base com 481 imagens que vão ser utilizadas para o treino e 53 imagens para o teste que obteve um resultado com 98% de precisão. Por fim em Lee et. al. [11], o estudo foi efetuado sob uma base de 12.500 imagens (sendo 6500 de fogo e 6000 de fumaça) utilizando o modelo *faster R-CNN* para a classificação de fogo e fumaça. Os autores também apontam que a utilização de uma

boa base de dados é de grande importância para a melhor acurácia.

O presente artigo apresenta uma técnica para detecção de fogo em vídeos baseado no pré-processamento dos vídeos utilizando a técnica de remoção de fundo, assim como apresentar o resultado dos testes realizados na arquitetura Xception [13] e FSDN [4] na base proposta por Foggia et.al. em [14]. A divisão do trabalho está da seguinte maneira: a segunda seção apresenta o método proposto para a identificação de fogo em vídeos utilizando recursos de extração de fundo em quadros e a aplicação de redes neurais convolucionais, a terceira seção contém os resultados obtidos e a quarta seção a conclusão do trabalho e indicação de trabalhos futuros.

## II. MÉTODO PROPOSTO

O presente trabalho visa a detecção de fogo em vídeos utilizando técnicas de remoção de fundo e CNN, para isso, as etapas ilustradas na Figura 1 são descritas a seguir. A partir da base de vídeos proposto por [14], o primeiro passo foi separar os vídeos em: fogo, não-fogo e fumaça. Após a divisão destes vídeos, o algoritmo de detecção de movimento foi aplicado para a remoção do fundo, realçando assim o fogo, movimentos de outros objetos e fumaça. Em seguida a remoção do fundo, com os quadros resultantes, foram separados e rotulados em fogo, não-fogo e fumaça. Com esta separação, as imagens foram então separadas em treino, teste e validação na proporção aproximada de 60/20/20. Assim realizou-se o treino e a validação usando duas Redes Neurais Convolucionais a Xception (proposta por [13]) e a FSDN (proposta por [4]).

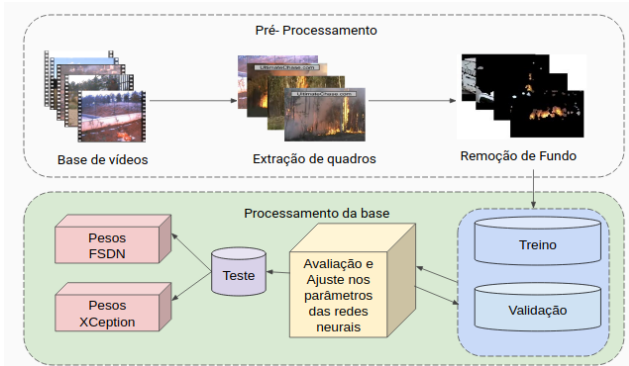


Fig. 1. Etapas para a concepção de modelo de detecção de fogo e fumaça

### A. Base de Vídeos

Inicialmente foi manuseado uma base de dados composta por 31 vídeos de câmeras fixas disponibilizados pelos autores do trabalho [14]. Apesar de ter 31 vídeos, 2 dos vídeos disponibilizados pelos autores estavam danificados, desta forma, utilizamos 29 vídeos para os testes deste trabalho. Nesta base de 62.690 quadros foram extraídos, 4.481 quadros foram escolhidos de forma sequencial, dando o espaçamento entre eles, isto se deve pela grande similaridade entre os quadros próximos. Para a realização deste estudo os quadros escolhidos

foram separados em 2.964 para o treino, 715 para teste e 802 para a validação. Para o pré-processamento, a biblioteca OpenCV v.4.2 para Python v.3.6 foram utilizadas e para as construções dos modelos de CNN o *framework* Keras 2.3 em conjunto com o Tensorflow v.2.2.

### B. Pré-processamento

Na fragmentação dos vídeos em quadros para ocorrer a realização da remoção de fundo, foi utilizada uma máscara de tamanho 9 com os desfoques gaussiano (*Gaussian Blur*) e de dilatação foram aplicadas, a margem de tolerância (quanto maior o valor, menor é a tolerância para movimentos) utilizada para a detecção do movimento foi de 0,6.

Os vídeos foram fragmentados a cada 5 iterações (5 quadros por segundo) considerando a taxa de atualização do vídeo de 50 fps (*frames per second*). Na figura 2 é possível visualizar a esquerda as imagens antes da remoção de fundo e a direita após a remoção.

Durante a remoção de fundo a estimativa de identificação do "objeto de interesse" foi realizada por meio da aplicação do modelo baseado em inferências Bayesianas e a utilização de modelos gaussianos adaptativos para a remoção de fundo [15], que consiste detecção do *pixel* com a utilização da probabilidade de ter um valor  $x_n$ , em que  $w_j$  é o valor do peso do parâmetro na  $k^{th}$  componente Gaussiano

$$p(x_n) = \sum_{j=1}^k w_j \eta(x_n; \Theta_j) \quad (1)$$

em seguida, para a detecção de fundo utilizou-se da concepção de um vetor de posição com a RGB média dos pixels de fundo,  $E$ , uma *expected chromaticity line*,  $\|E\|$ , uma aberração cromática,  $d$  e um *threshold* de brilho,  $\tau$ , para um determinado pixel de observado valor,  $I$ , uma distorção de brilho,  $a$ , e uma distorção de cor,  $c$ .

$$a = \arg \min (I - zE)^2 \quad (2)$$

$$c = \|I - aE\| \quad (3)$$

Um *pixel* só será considerado em primeiro plano se o resultado de  $a$  estiver dentro dos limites estabelecidos e  $\tau < c < 1$  [15].



Fig. 2. Imagens antes e depois da retirada de fundo

### C. Treinamento e validação da base

As CNNs utilizadas possuem a arquiteturas de deep learning FSDN-Fire Smoke Detection Network com mini-lote 100 e a *Xception* com mini-lote 16, com uma taxa de aprendizado (*Learning rate*) de 0,00001 a 0,01, com o otimizador *Stochastic Gradient Descent* (SGD) e a função de perda *categorical\_crossentropy*. A arquitetura FSDN foi criada para a classificação de 3 classes: fogo, fumaça e negativo, com 8 camadas convolucionais residuais e 2 camadas convolucionais convencionais, 22 blocos de *inception* e 1024 neurônios.

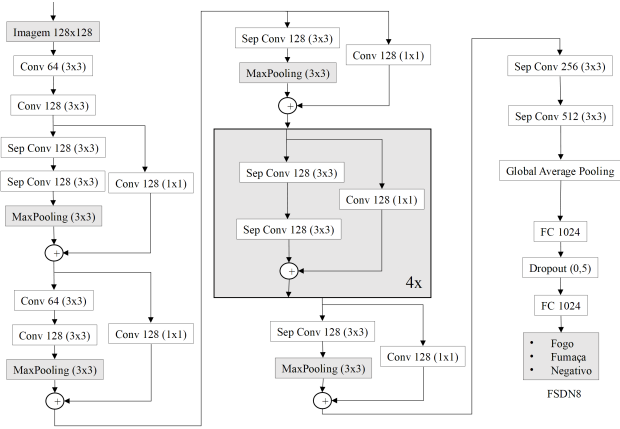


Fig. 3. Arquitetura FSDN

Já a arquitetura *Xception* possui 36 camadas convolucionais na forma de pilha linear de camadas de convoluções [13], que foram pré-treinadas com a base de dados *ImageNet*.

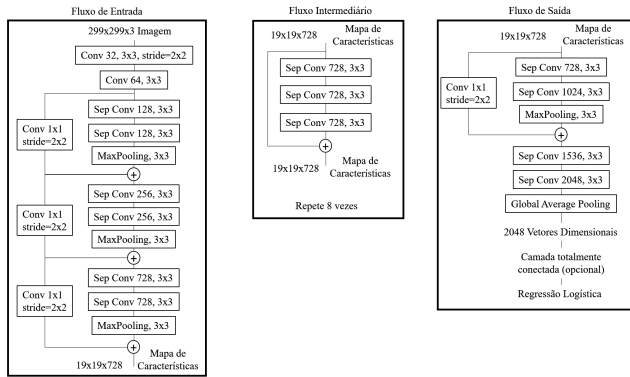


Fig. 4. Arquitetura Xception

## III. RESULTADOS E DISCUSSÃO

A partir do pré-processamento em que foram retiradas o fundo e realizada a parametrização das CNN, será efetuado o treino e validação da base de quadros com 30 épocas cada. Pode-se verificar na Figura 5, que os treinos geraram maior acurácia na arquitetura *Xception* com 94,87% e 88,69% na FSDN, com uma perda (*loss*) de 33,90% na arquitetura FSDN e 19,73% na Xception. Durante a validação como mostra a

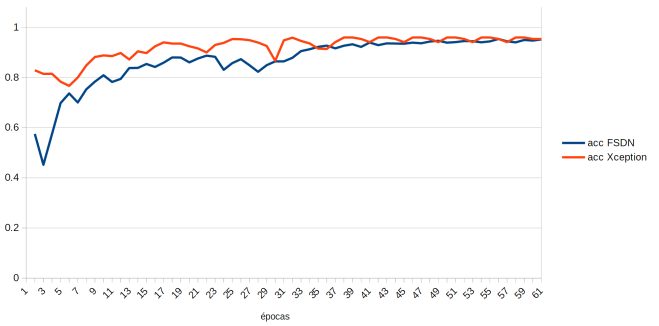


Fig. 5. Acurácia nas arquiteturas *Xception* e FSDN durante o treino

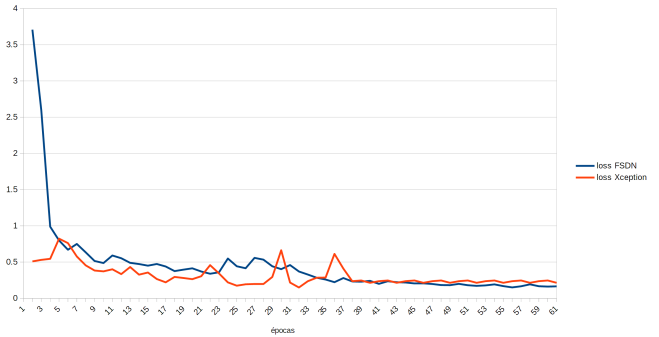


Fig. 6. Perda (*loss*) nas arquiteturas *Xception* e FSDN durante o treino

Tabela I, a acurácia da FSDN foi maior com 92,73% durante a época 29 e a *Xception* com 86,28% durante a época 25.

No estudo realizado anteriormente por Foggia et. al. [14], a metodologia empregada baseou-se em múltiplos sistemas especialistas (MSE) com regras em espaços de cores, a base utilizada consiste em 62.690 quadros. Neste estudo é realizada um pré-processamento com a retirada de fundo dos quadros, para assim realizar a validação a partir de três sistemas especialistas.

O primeiro sistema consiste na validação de cor no espaço de cor YUV, já que este é bem utilizado por separar a luminância e a crominância. Já o segundo é baseado no formato do primeiro plano, comparando dois quadros consecutivos para a comparação da diferença de formato do fogo. O terceiro sistema é baseado no movimento dos quadros na qual é utilizada o modelo saco de palavras. Em seguida, as decisões tomadas pelos sistemas são combinadas em um MSE com seus respectivos pesos, gerando assim uma acurácia de 93,53%.

Nos estudos realizados anteriormente com a utilização das CNN sem a utilização da remoção de fundo é possível verificar a existência de um *overfitting*, como dito em Dua et.al. [12], ou um aumento na complexidade do algoritmo, como o ocorrido em Priya et. al [10]. Sob o mesmo ponto, o estudo atual realizado visa o reconhecimento de cenários de fogo em vídeos, realizando no pré-processamento a remoção de fundo, o que acarretou em um resultado preliminar significativo. Contudo, foi proposta uma base de dados relativamente menor, com 4.481 quadros juntamente com o uso de CNNs, obtendo um

TABELA I  
RESULTADOS DAS ARQUITETURAS UTILIZADAS DURANTE A VALIDAÇÃO

Arquitetura	Época	Acurácia	F1_Score	Loss	MAE
FSDN	4	0,8685	0,8702	0,5935	0,2572
	12	0,8727	0,8723	0,1114	0,0844
	20	0,8041	0,7947	0,1097	0,1405
	<b>29</b>	<b>0,9272</b>	<b>0,9272</b>	<b>0,4574</b>	<b>0,0730</b>
XCception	6	0,7748	0,7744	0,7682	0,1823
	14	0,8339	0,8335	0,6072	0,1301
	21	0,8276	0,8292	0,8182	0,1197
	<b>25</b>	<b>0,8628</b>	<b>0,8615</b>	<b>0,6184</b>	<b>0,1139</b>

resultado similar (acurácia de 92,73% na arquitetura *Xception* e 86,28% na FSDN), com uma menor amostragem e processamento, o que demonstra que a etapa de pré-processamento remoção de fundo auxilia diretamente na detecção de fogo.

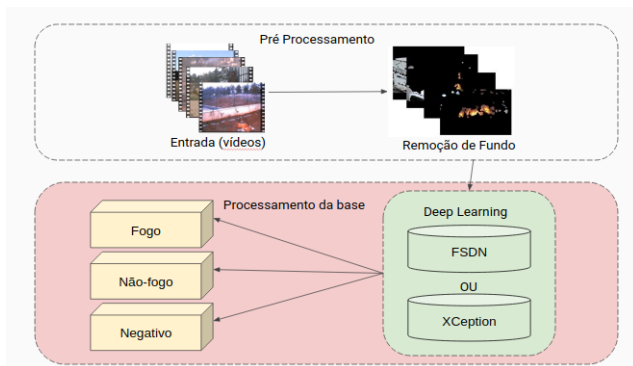


Fig. 7. Método de detecção de Fogo e Fumaça proposto

#### IV. CONCLUSÃO

A detecção do fogo é complexa já que possui muitas variáveis complexas como mudança de formato, cor, tamanho, entre outros, para isto é necessária uma ampla base de dados a serem analisados e uma boa remoção de fundo a fim de aumentar a acurácia da detecção. Foi utilizada nesse artigo um método de remoção de fundo baseado em inferências Bayesianas junto com a mudança no espectro de cor. Tendo como resultado preliminar um valor significativo de acurácia de 92,73% e um F1 de 92,73% com a arquitetura FSDN e acurácia de 86,28% e F1 de 86,15% com a arquitetura *Xception* na realização do treino, foi possível concluir que a remoção de fundo como etapa de pré-processamento juntamente com a parametrização das redes neurais foram de grande importância para o alcance deste resultado.

Com este estudo preliminar foi possível localizar os parâmetros ideais para o treino das redes neurais, possibilitando a realização de novos testes com o maior número de épocas, causando uma exaustão do treino e consolidando um modelo. Nos trabalhos futuros é necessário buscar novas formas da extração de fundo, utilizando outros modelos de CNNs a fim de aumentar a acurácia de detecção, podendo ser aplicadas em cenários de vídeos retirados a partir de veículos aéreos não tripulados.

#### AGRADECIMENTOS

Agradecemos a Fundação Araucária pelo apoio financeiro e a *NVIDIA* por fornecer uma GPU.

#### REFERENCES

- [1] C. F. de Castro, G. Serra, J. Parola, J. Reis, L. Lourenço, and S. Correia, "Combate a incêndios florestais," *Escola Nacional de Bombeiros*, vol. 13, 2003.
- [2] INPE. (2019) Monitoramento dos focos ativos por países. [Online]. Available: [http://queimadas.dgi.inpe.br/queimadas/portal-static/estatisticas\\_paises/](http://queimadas.dgi.inpe.br/queimadas/portal-static/estatisticas_paises/)
- [3] L. F. L. Fraga, "Verificação de adequação do projeto de uma tubulação de incêndio de uma refinaria de petróleo conforme as normas da petrobras," 2010.
- [4] B. M. N. de Souza, "Detecção e localização de fogo em imagens digitais usando técnicas de aprendizado de máquina," Ph.D. dissertation, Pontifícia Universidade Católica do Paraná, 2019. [Online]. Available: [https://www.ppgia.pucpr.br/pt/arquivos/doutorado/teses/2019/066\\_Tese\\_Bruno.pdf](https://www.ppgia.pucpr.br/pt/arquivos/doutorado/teses/2019/066_Tese_Bruno.pdf)
- [5] T. Celik, H. Demirel, and H. Ozkaramanli, "Automatic fire detection in video sequences," in *2006 14th European Signal Processing Conference. IEEE*, 2006, pp. 1–5.
- [6] T. Celik and H. Demirel, "Fire detection in video sequences using a generic color model," *Fire safety journal*, vol. 44, no. 2, pp. 147–158, 2009.
- [7] S. Rudz, K. Chetehouna, A. Hafiane, O. Sero-Guillaume, and H. Laurent, "On the evaluation of segmentation methods for wildland fire," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2009, pp. 12–23.
- [8] W.-B. Horng, J.-W. Peng, and C.-Y. Chen, "A new image-based real-time flame detection method using color analysis," in *Proceedings. 2005 IEEE Networking, Sensing and Control*, 2005. IEEE, 2005, pp. 100–105.
- [9] B. C. Ko, K.-H. Cheong, and J.-Y. Nam, "Fire detection based on vision sensor and support vector machines," *Fire Safety Journal*, vol. 44, no. 3, pp. 322–329, 2009.
- [10] R. S. priya and K. Vani, "Deep learning based forest fire classification and detection in satellite images," in *2019 11th International Conference on Advanced Computing (ICoAC)*, 2019, pp. 61–65.
- [11] Y. Lee, D. Im, and J. Shim, "Data labeling research for deep learning based fire detection system," in *2019 International Conference on Systems of Collaboration Big Data, Internet of Things Security (SysCoBioTS)*, 2019, pp. 1–4.
- [12] M. Dua, M. Kumar, G. Singh Charan, and P. Sagar Ravi, "An improved approach for fire detection using deep learning models," in *2020 International Conference on Industry 4.0 Technology (I4Tech)*, 2020, pp. 171–175.
- [13] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 1251–1258.
- [14] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE TRANSACTIONS on circuits and systems for video technology*, vol. 25, no. 9, pp. 1545–1556, 2015.
- [15] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Video-based surveillance systems*. Springer, 2002, pp. 135–144.