KutralNet: A Portable Deep Learning Model for Fire Recognition

Angel Ayala [1] **Bruno Fernandes** [1] **Francisco Cruz** [2] **David Macêdo** [3] **Adriano L. I. Oliveira** [3] **Cleber Zanchettin** [3]

Abstract

Most of the automatic fire alarm systems detect the fire presence through sensors like thermal, smoke, or flame. The image approach is promising since it does not need specific sensors and can be easily embedded in different devices. However, in deep learning methods, a high computational cost is associated. In this work, we propose a new deep learning architecture that requires fewer floating-point operations (flops) for fire recognition. Additionally, we use modern techniques such as inverted residual block, convolutions like depth-wise, and octave, to reduce the model's computational cost. The proposed methods are evaluated on FireNet and FiSmo datasets. The experiments show that our model keeps high accuracy, and a portable version presents 71% fewer parameters than FireNet, while still presenting competitive accuracy and AUROC performance. The obtained results are promising for the implementation of the model in a mobile device, considering the reduced number of flops and parameters.

1. Introduction

The fire alarm systems are a combination of sensors and machine learning algorithms to identify patterns of warning. The use of images to fire recognition is a new promising approach (Ayala et al., 2019), based on the excellent results that deep learning models are obtaining in image processing applications (Chollet, 2017) allowing to avoid the use of special sensors to perform the recognition.

The deep learning (DL) approach (LeCun et al., 2015) has proved to be suitable for automating the feature acquisition in multiple levels of abstraction from complex data in machine learning tasks. Considering that, the use of computer vision to fire recognition reduces the necessity of specific sensors, being suitable for the inlay to portable, remote, and mobile devices. However, DL approaches have some challenges: (i) the required computational resources; (ii) the model's computation complexity and size; and (iii) the quantity of data needed for its training, among others. Some works have addressed portable models of deep learning using more efficient ways to compute the convolutions (Sandler et al., 2018). The authors proposed the inverted residual block with point-wise and depth-wise convolutions to simplify the dimensionality of the signal processing. Another efficient approach is the octave convolution (Chen et al., 2019), which reduces the spatial redundancy of the signal, processed as high and low spatial frequency,

In this work, we propose a new deep learning model to fire recognition called KutralNet[1], which comprises five layers and require 92% fewer floating-point operations (flops) for processing in comparison with previous approaches. This model is used as a baseline to build portable models that compare the efficiency in signal processing of the inverted residual block, the depth-wise convolution, and octave convolution approaches. Our best portable model, KutralNet Mobile Octave, presents a competitive validation accuracy and AUROC performance despite using 71% fewer numbers of parameters in comparison with FireNet. We compare the proposed models with state-of-the-art approaches to fire recognition over the FireNet and FiSmo datasets.

2. Proposal and Experimental Setup

In this section, we present the KutralNet model to fire recognition and the datasets implemented for training, validation, and test of the models.

2.1. Kutralnet Architectures

The **KutralNet's baseline** model is inspired by FireNet (Jadon et al., 2019), OctFiResNet (Ayala et al., 2019), and the modified ResNet50 (Sharma et al., 2017) models. Our KutralNet is defined for processing 84x84 pixels images

[1]Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil [2]School of Information Technology, Deakin University, Geelong, Australia [3]Centro de Informática, Universidade Federal de Pernambuco, Recife, Brasil. Correspondence to: Angel Ayala <aaam@ecomp.poli.br>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

[1]We took inspiration from Mapuche language or Mapudungun where kütral means fire.

Table 1. The computational cost of each implemented model represented as parameters and flops. The fewer parameters, the less on-disk size, is required. Moreover, as fewer the number of flops, less is the computational cost for processing.

Model$_{(InputSize)}$	**Parameters**	**Flops**
ResNet50$_{(224^2)}$	31.91M	4.13G
OctFiResNet$_{(96^2)}$	956.23K	928.95M
FireNet$_{(64^2)}$	646.82K	-
KutralNet$_{(84^2)}$	138.91K	76.85M
KutralNet Mobile$_{(84^2)}$	173.09K	43.27M
KutralNet Octave$_{(84^2)}$	125.73K	29.98M
KutralNet Mobile Octave$_{(84^2)}$	185.25K	24.59M

with RGB channels in a lightweight five-block layers configuration. All the portable models are based in this architecture, which on the top of the network is composed of a LeakyReLU activation, passing through a global average pooling layer directly to a fully connected layer with two neurons on the exit. The **KutralNet Mobile** model, is inspired by MobileNetV2 (Sandler et al., 2018) and presents the implementation of the inverted residual block. In this approach, from the second block, the KutralNet convolution blocks are replaced with the inverted residual block, in which each block contains point-wise and depth-wise convolution with shortcut connections in some cases. For the **KutralNet Octave** model, all the KutralNet vanilla convolution are replaced with octave convolution with an α parameter of 0.5. The **KutralNet Mobile Octave** model, is the combination of the MobileNetV2 block and the octave convolution. It is the same KutralNet Mobile but replacing the vanilla convolution with the octave convolution combined with depth-wise convolution form. The models' summary is in Table 1. More implementation details are available at the work's repository.

2.2. Datasets

For the training, validation, and test of the models, two datasets are used in this work. The first one is called FireNet as the model (Jadon et al., 2019) and contains training and test subsets, with 2425 and 871 images, respectively. The second one is the FiSmo dataset (Cazzolato et al., 2017), which has been recently published with a total of 6063 images. Additionally, we have used a contained subset of FiSmo (FiSmoB) comprised of 1968 images equally balanced between the fire and no-fire label. An augmented version of FiSmo (FiSmoA) is also used, adding 485 black images labeled as no-fire, in order to check out the models' response. In addition to the balanced FiSmo version, we have also used an augmented version of this subset (FiSmoBA), which replaces 98 no-fire images for black images.

Table 2. Mean performance values for testing accuracy and AUROC index of each model. The datasets FireNet, FiSmo, and FiSmoA, were used for training and validation of the models.

Model	**Test Acc**	**AUROC**
FireNet	64.27%	**0.96**
KutralNet	**78.26%**	0.92
OctFiResNet	75.92%	0.87
ResNet50	70.26%	0.90

Table 3. Mean performance values for testing accuracy and AUROC index of each portable model. The datasets FiSmo, FiSmoB, and FiSmoBA, were used for training and validation of the models.

Model	**Test Acc**	**AUROC**
KutralNet	76.01%	0.86
KutralNet Mobile	71.99%	0.85
KutralNet Octave	73.90%	0.85
KutralNet Mobile Octave	**79.49%**	**0.90**

3. Results

The training of all the models was performed during 100 epochs to choose the model with the best validation accuracy. Our first experiment aimed to prove the effectiveness to fire recognition of our baseline model. The first comparison has been performed over the FireNet dataset, with a validation accuracy of 93.83%, 96.02%, 95.34%, and 98.22% for FireNet, KutralNet, OctFiResNet, and ResNet50 respectively. The test results are present in Table 2.

With our KutralNet baseline architecture, the next experimentation was to reduce its computational cost. In the first place, the training was performed over FiSmo, with a validation accuracy of 88.62%, 85.99%, 87.55%, and 87.39% is achieved by KutralNet, KutralNet Mobile, KutralNet Octave, and KutralNet Mobile Octave model respectively. The test results are present in Table 3.

4. Conclusions

Our baseline proposed as KutralNet accomplishes an interesting performance compared with previous deep models for fire recognition. KutralNet reduces 85% the parameters number and 92% the number of operations required, in comparison to the OctFiResNet model.

The portable version KutralNet Mobile Octave can achieve good performance even if trained with different datasets for the fire and no-fire classification task, requiring only of 24.6M flops with 185.3K parameters.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazilian research agencies.

References

Ayala, A., Lima, E., Fernandes, B., Bezerra, B. L. D., and Cruz, F. Lightweight and efficient octave convolutional neural network for fire recognition. In *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pp. 87–92, 2019. doi: 10.1109/LA-CCI47412.2019.9037059.

Barmpoutis, P., Dimitropoulos, K., and Grammalidis, N. Real time video fire detection using spatio-temporal consistency energy. In *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 365–370, 2013. doi: 10.1109/AVSS.2013.6636667.

Cazzolato, M. T., Avalhais, L. P., Chino, D. Y., Ramos, J. S., de Souza, J. A., Rodrigues-Jr, J. F., and Traina, A. Fismo: A compilation of datasets from emergency situations for fire and smoke analysis. In *Brazilian Symposium on Databases - SBBD*, pp. 213–223, 2017.

Chen, Y., Fan, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., Yan, S., and Feng, J. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. *arXiv preprint arXiv:1904.05049*, 2019.

Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Deng, Y. Deep learning on mobile devices: A review. In Agaian, S. S., Asari, V. K., and DelMarco, S. P. (eds.), *Mobile Multimedia/Image Processing, Security, and Applications 2019*, volume 10993, pp. 52 – 66. International Society for Optics and Photonics, SPIE, 2019. doi: 10.1117/12.2518469.

Dimitropoulos, K., Barmpoutis, P., and Grammalidis, N. Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(2):339–351, 2015. ISSN 1051-8215. doi: 10.1109/TCSVT.2014.2339592.

Grammalidis, N., Çetin, E., Dimitropoulos, K., Tsalakanidou, F., Kose, K., Gunay, O., Gouverneur, B., Torri, D., Kuruoglu, E., Tozzi, S., Benazza, A., Chaabane, F., Kosucu, B., and Ersoy, C. A multi-sensor network for the protection of cultural heritage. In *2011 19th European Signal Processing Conference*, pp. 889–893, 2011.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

Jadon, A., Omama, M., Varshney, A., Ansari, M. S., and Sharma, R. FireNet: A specialized lightweight fire & smoke detection model for real-time iot applications. *CoRR*, abs/1905.11922, 2019.

Kim, Y., Kim, A., and Jeong, H. RGB color model based the fire detection algorithm in video sequences on wireless sensor network. *International Journal of Distributed Sensor Networks*, 10(4):923609, 2014. doi: 10.1155/2014/923609.

LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature*, 521(7553):436–444, may 2015. ISSN 0028-0836. doi: 10.1038/nature14539.

Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., and Baik, S. W. Efficient deep CNN-based fire detection and localization in video surveillance applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(7):1419–1434, 2018. ISSN 21682232. doi: 10.1109/TSMC.2018.2830099.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, June 2018. doi: 10.1109/CVPR.2018.00474.

Sharma, J., Granmo, O., Goodwin, M., and Fidje, J. T. Deep convolutional neural networks for fire detection in images. In *Engineering Applications of Neural Networks*, pp. 183–193, Cham, 2017. Springer International Publishing. ISBN 978-3-319-65172-9.

Xu, B., Wang, N., Chen, T., and Li, M. Empirical evaluation of rectified activations in convolutional network. *CoRR*, abs/1505.00853, 2015.