

# An FPGA Smart Camera Implementation of Segmentation Models for Drone Wildfire Imagery

Eduardo Garduño<sup>1</sup>, Jorge Ciprian-Sanchez<sup>2</sup>, Valente Vazquez-Garcia<sup>3</sup>, Miguel Gonzalez-Mendoza<sup>1</sup>, Gerardo Rodriguez-Hernandez<sup>1</sup>, Adriana Palacios-Rosas<sup>4</sup>, Lucile Rossi-Tisson<sup>5</sup>, Gilberto Ochoa-Ruiz<sup>1</sup>

<sup>1</sup> Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Mexico

<sup>2</sup> University of Postdam, Hasso-Plattner Institute, Germany

<sup>3</sup> Universidad Autónoma de Guadalajara, Maestría en Ciencias Computacionales, Mexico

<sup>4</sup> Universidad de las Americas Puebla, Departamento de Ingeniería Química, Alimentos y Ambiente, Mexico

<sup>5</sup> Università di Corsica, Laboratoire Sciences Pour l'Environnement, Campus Grimaldi, France

mgonza@tec.mx, adriana.palacios@udlap.mx, gilberto.ochoa@tec.mx

**Abstract.** Wildfires represent one of the most relevant natural disasters worldwide, due to their impact on various societal and environmental levels. Thus, a significant amount of research has been carried out to investigate and apply computer vision techniques to address this problem. One of the most promising approaches for wildfire fighting is the use of drones equipped with visible and infrared cameras for the detection, monitoring, and fire spread assessment in a remote manner but in close proximity to the affected areas. However, implementing effective computer vision algorithms on board is often prohibitive since deploying full-precision deep learning models running on GPU is not a viable option, due to their high power consumption and the limited payload a drone can handle. Thus, in this work, we posit that smart cameras, based on low-power consumption field-programmable gate arrays (FPGAs), in tandem with binarized neural networks (BNNs), represent a cost-effective alternative for implementing onboard computing on the edge.

Herein we present the implementation of a segmentation model applied to the Corsican Fire Database. We optimized an existing U-Net model for such a task and ported the model to an edge device (a Xilinx Ultra96-v2 FPGA). By pruning and quantizing the original model, we reduce the number of parameters by 90%. Furthermore, additional optimizations enabled us to increase the throughput of the original model from 8 frames per second (FPS) to 33.63 FPS without loss in the segmentation performance: our model obtained 0.912 in Matthews correlation coefficient (MCC), 0.915 in F1 score and 0.870 in Hafiane quality index (HAF), and comparable qualitative segmentation results when contrasted to the original full-precision model. The final model was integrated into a low-cost FPGA, which was used to implement a neural network accelerator.

**Keywords.** SoC FPGA, computer vision, segmentation, binarized neural networks, artificial intelligence, infrared imaging, pruning.

## 1 Introduction

A wildfire is an exceptional or extraordinary free-burning vegetation fire that may have been started maliciously, accidentally, or through natural means that could significantly affect the global carbon cycle by releasing large amounts of CO<sub>2</sub> into the atmosphere.

It has profound economic effects on people, communities, and countries, produces smoke that is harmful to health, devastates wildlife, and negatively impacts bodies of water [26].

The three main categories of remote sensing for wildfire monitoring and detection systems are ground-based systems, manned aerial vehicle-based systems, and satellite-based systems. However, they present the following technological and practical problems: ground-based have limited surveillance ranges.

Satellite-based have problems when planning routes, their spatial resolution may be low, and the information transmission may be delayed. Manned aerial vehicle-based systems are expensive and potentially dangerous due to hazardous environments and human error.

Unmanned aerial vehicles (UAVs) provide a mobile and low-cost solution using computer vision-based remote sensing systems that can perform long-time, monotonous, and repetitive tasks [31]. Drones, in particular, represent an excellent opportunity due to their easy deployment.

However, the ability to implement these fire detection systems, based on deep learning (DL), is limited by the maximum payload of the drone and the high power consumption. In this paper, we posit that a convolutional neural network (CNN) can be implemented on a hardware accelerator that can be embedded as part of a smart camera and installed on a drone for the detection of wildfires.

A review of the literature on hardware implementation for various artificial intelligence (AI) algorithms was published by Talib et al. [22] reviewing 169 different research reports published between 2009 and 2019, which focus on the implementation of hardware accelerators by using application-specific integrated circuits (ASICs), FPGAs, or GPUs. They found that most implementations were based on FPGAs, focusing

mainly on the acceleration of CNNs for object detection, letting the GPU-based implementations in second place.

Due to the diversity of applications, AI models such as CNNs need to meet various performance requirements for drones and autonomous vehicles, with the essential demands of low latency, low weight overhead, long-term battery autonomy, and low power consumption being the most pressing requirements. The complexity of the tasks that CNNs must perform continues to increase as models evolve.

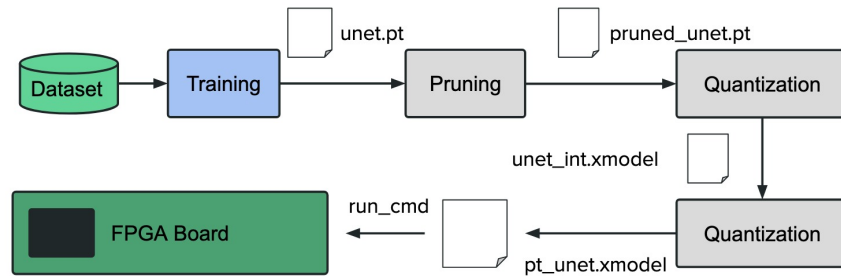
As a result, deeper networks are designed in exchange for higher computational and memory demands. In this context, the reconfiguration capabilities of FPGAs enable the creation of CNN hardware implementations that are high-performance, low-power, and configurable to fit system demands [27].

A smart camera is an embedded system for computer vision applications that has attracted great interest in various application domains, as it offers image capture and image processing capabilities in a compact system [20].

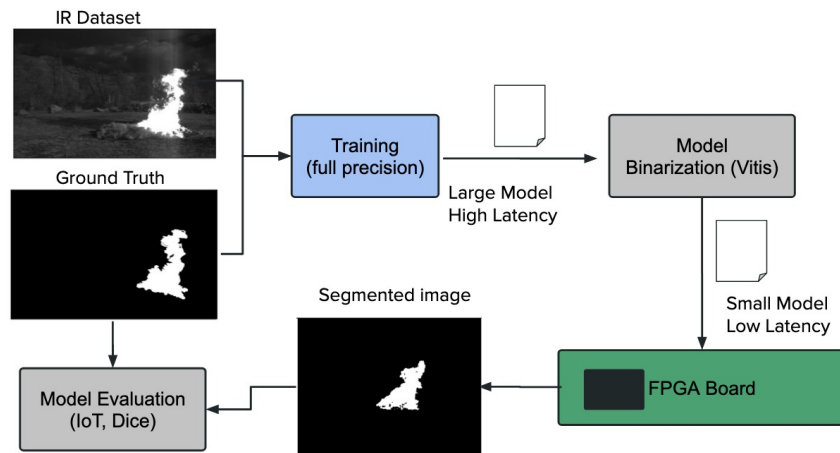
This paper describes the methodology, implementation, design cycle, and experimental protocol of porting a modified U-Net model into a Xilinx Ultra96-V2 FPGA for the wildfire semantic segmentation task for the smart camera system. The rest of the paper is organized as follows: Section 2 discusses recent works applying computer vision models for wildfire segmentation, highlighting their strengths and limitations; the second part of the section discusses related works regarding smart camera implementations in order to better contextualize our work.

Section 3 details our contribution, discussing in detail the proposed model, the dataset used for evaluating our models, and the design flow followed for optimizing the model and testing it in the target embedded FPGA board.

Section 4 discusses the results of our optimization process and provides a quantitative and qualitative comparison between full precision and the BNN model. Finally, Section 5 concludes the paper and discusses future areas of research.



**Fig. 1.** General overview of the Pytorch flow for Vitis AI. This flow allows us to optimize a given full precision model and target an embedded device such as an FPGA, consuming less power while attaining a higher throughput in terms of processed FPS



**Fig. 2.** Overall implementation flow for FPGA-based systems based on Vitis AI

## 2 State of the Art

### 2.1 Segmentation Models for Wildfire Detection and Characterization

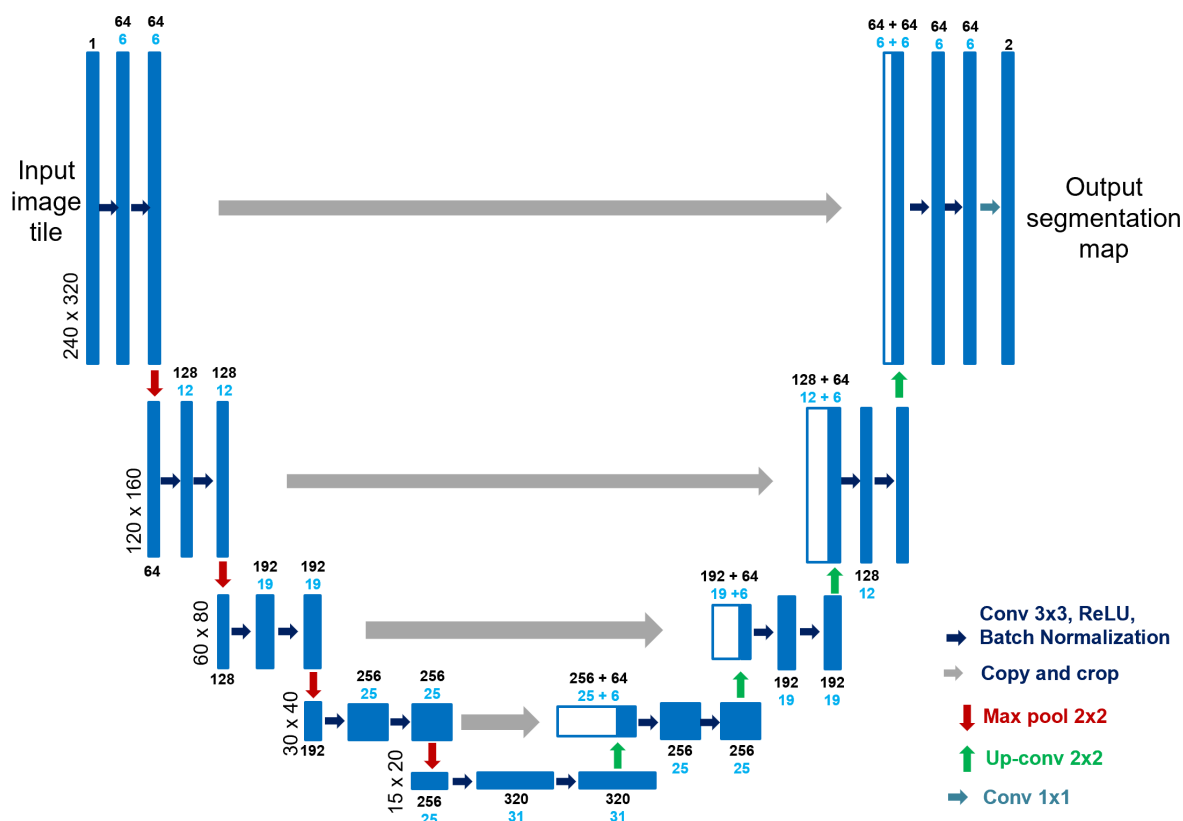
Detecting a wildfire by categorizing each pixel in an infrared image is a semantic segmentation problem; therefore, for this task, AI models have been used, such as fully convolutional networks as well as the U-Net model proposed by Ronnenberger et al. in 2015 [19], which allow precise segmentation with few training images.

For the specific task of fire segmentation, artificial intelligence models have already been implemented to solve this problem with visible images of fire [2], the fusion of visible and infrared

images of fire [8], and visible images of fire and smoke [18]. Akhloufi et al. [2] proposed Deep-Fire, a semantic segmentation model based on the U-Net architecture.

The authors trained and evaluated their model using the Corsican Fire Database [25]. With an F1 score ranging from 64.2% to 99% on the test set, Akhloufi et al. claimed successful results using the Dice similarity coefficient as the loss function for the model.

Ciprián-Sánchez et al. [8] evaluated thirty-six different DL combinations of the U-Net-based Akhloufi architecture [2], the FusionNet-based Choi architecture [6], and the VGG16-based Frizzi architecture [9], the Dice [16], Focal Tversky



**Fig. 3.** Proposed architecture. The original U-Net architecture has been extended by introducing batch normalization layers and fewer filters in the deepest layers to reduce training and inference times. The numbers on black on the top of the blocks represent the original filter sizes, whereas the blue one (below) represents the filter size after the optimization process

[1], Unified Focal [30] losses, and the visible and near-infrared (NIR) images of the Corsican Fire Database [25] and fused visible-NIR images produced by the methods by Li et al. [15] and Ciprián-Sánchez et al. [7].

After evaluating these models, the combination with the best results was Akhloufi + Dice + visible with a 0.9323 F1 score, also known as the Dice coefficient.

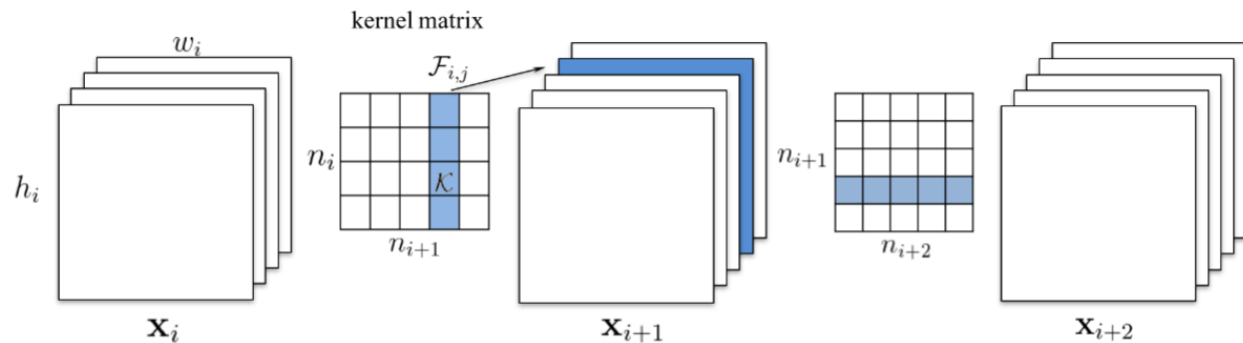
Although these works have highlighted the potential of using AI in this domain, many of these algorithms are incapable of operating in real-time, as they inherently suffer from very high inference times and are prohibitive as they require many computing resources, which impedes their usability on drone missions and thus we posit that new paradigms are needed for their successful

deployment, particularly in terms of inference time (FPS) and power consumption.

## 2.2 Smart Camera Implementations for Computer Vision

Smart cameras are devices that process, analyze, and extract data from the images they capture. Different video processing algorithms are used for the extraction.

Smart cameras have been employed in a variety of applications, including human gesture recognition [29], surveillance [4], smart traffic signal optimization systems [23], and a fire detection system based on conventional image processing methods [10].



**Fig. 4.** When a filter is pruned, the matching feature map and associated kernels in the following layer are removed. Retrieved from: Hao Li et al. [14]

We propose a DL implementation capable of performing a precise segmentation that can be used as a first step in wildfire characterization and risk assessment systems.

FPGAs are excellent choices for creating smart cameras because they offer significant processing capabilities while maintaining a low power consumption, which makes them good candidates for particular edge tasks creating efficient hardware accelerators capable of high throughput [27], and maintaining a high degree of flexibility and reconfigurability.

The disadvantage of FPGAs is that developers need to be skilled in hardware design to accomplish these goals. The design process frequently takes longer with FPGAs than with CPU and GPU systems.

To address such issues, FPGA vendors and other academic and industrial tool developers have introduced several computer-aided design (CAD) tools for training and optimizing DL models and mapping such models into the reconfigurable fabric.

Convolutional neural networks provide high-accuracy results for computer vision tasks, and their applications could be benefited from being implemented in edge devices such as FPGAs. Still, for applications such as smart cameras, limited use of hardware resources and power requirements are of the utmost importance.

Therefore, to implement models that generally require a large number of computational resources, large storage capabilities for the model

parameters, and the use of high energy consuming hardware [28, 27], such as GPUs, it is necessary to use model optimization techniques such as pruning and quantization [3] for the compression of the model, to implement it in devices such as an FPGA while achieving high inference speed.

### 3 Proposed Method

The implementation of a BNN for the segmentation of wildfire images was done using the Xilinx tool Vitis AI because each operation of the model is mapped into a hardware-accelerated microinstruction, in which a series of sequential micro-instructions can represent the whole DL model, while a scheduler is in charge of managing the hardware calls and data flow. This enables the customization of the HW accelerator while considering the resources of the FPGA.

In the particular context of our application, Vitis AI is indeed the best choice as we target a small FPGA device (Xilinx Ultra96-V2) for deep embedded image processing.

#### 3.1 General Overview of the Optimization Approach

Fig. 1 depicts the Vitis AI Pytorch flow followed in this paper. The design process begins by training a segmentation with NIR images from the Corsican Fire Database [25] and their corresponding ground truths for fire region segmentation.

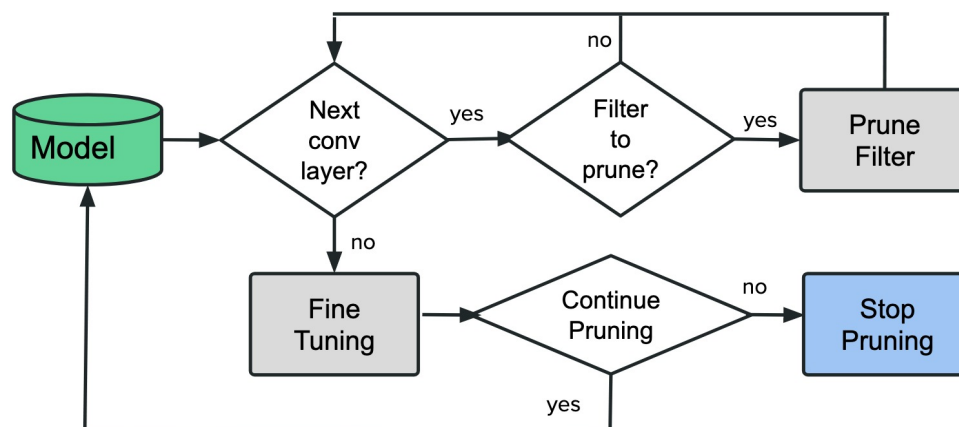


Fig. 5. Schematic flow for optimizing a model in Vitis AI

Subsequently, a pruning process to reduce the number of filters in the convolution layers using the Pytorch framework is performed. Then, both the original and the pruned models are saved in pt files.

The next module is in charge of changing the numerical representation of the DL model by performing an 8-bit quantization using the Vitis AI quantizer module, producing an xmodel file.

Finally, the quantized model is compiled, producing an xmodel file containing all the instructions needed by the DPU to execute the model. After the model has been compiled, it can be loaded on the target FPGA board and tested.

In our work, this model is a U-Net model modified to accommodate the needs of our application. The rest of this section will detail the implementation of such an optimized segmentation model.

### 3.2 Dataset: Corsican Fire Database

In this paper, we employ the NIR images from the Corsican Fire Database, first introduced by Toulouse et al. [25].

For fire region segmentation tasks, this dataset includes 640 pairs of visible and NIR fire images along with the matching ground truths created manually by experts in the field.

A representative NIR image from the Corsican Fire Database is shown in the top left corner of Fig. 2, along with its corresponding ground truth.

### 3.3 Segmentation Model Training

The proposed architecture for this paper is a modified version of a U-Net model [19] with the number of filters from the deepest layers reduced to reduce training and inference times.

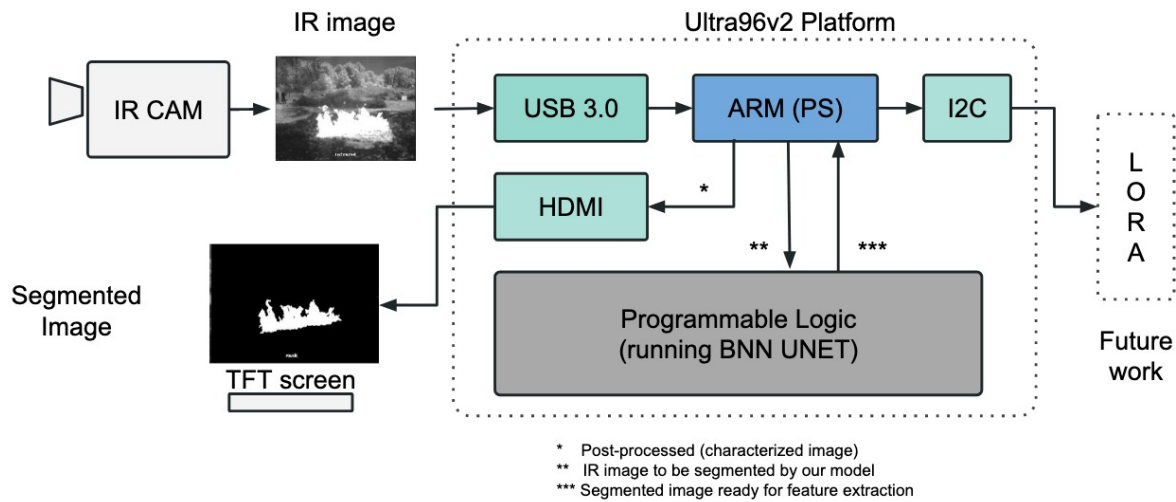
Furthermore, we add batch normalization layers [12] after every convolutional layer. The final architecture is shown in Fig. 3; the numbers in black are the number of filters before pruning, and the numbers in blue are the number of filters after pruning.

As depicted in Fig. 2, every image in the training set was resized to a width of 320 and a height of 240 pixels for training. For the training of the proposed model, the dataset was divided into 80% for training and 20% for testing.

The model was trained with a learning rate of 0.0001 for 350 epochs with a batch size of 5 using cross-entropy loss and Adam optimizer.

### 3.4 Optimization

**Pruning.** The pruning method (contained in the binarization block of Fig. 2) employed in the present paper is based on the work of [14] in which, as shown in Fig. 4, when a filter is pruned, the corresponding feature map is removed and the kernels of the input feature maps for the next



**Fig. 6.** Proposed solution model for implementing a smart camera for wildfire detection. Our current implementation processes images from external memory or an infrared (IR) camera; communication capabilities have not yet been implemented

convolution that correspond to the output feature maps of the pruned filters are also removed.

Fig. 5 briefly explains the pruning process used for this paper, with which it was possible to reduce the number of filters in each convolutional layer by approximately 90%.

In Fig. 3, we can see the final architecture of the model, the numbers in blue being the number of filters after the pruning process.

**Quantization.** The model was quantized using the Vitis AI quantizer module; resulting in a CNN model with all its values represented with only 8 bits. That is, the floating-point checkpoint model is converted into a fixed-point integer checkpoint.

After confirming there was no significant degradation in the model's performance, the quantized model was compiled with the Vitis compiler, which creates a xmodel file with all the instructions required by the DPU to execute the model.

### 3.5 Proposed FPGA-Based Smart Camera System

Fig. 6 shows the system implementation for the smart camera solution of wildfire detection. The processing system (PS) controls every step of the application's life cycle, including retrieving images from the camera, feeding them to the programmable logic (PL) section of the SoC (hardware accelerator implementing the proposed model), and processing the segmented image. An IR camera is attached to the Ultra96 board using a USB port in the SoC.

The PS block (an ARM processor) processes the input picture before feeding it to the PL section, which runs the binarized U-Net model mapped into the reconfigurable fabric. The image is processed and then passed back into the PS block for feature extraction.

If a complete IIoT solution is implemented, these features may be used for viewing on a TFT screen or communicated via a communication protocol (i.e., LORA) to a cloud.

These capabilities are not yet implemented here and are left for future work. In order to make the picture more straightforward, the

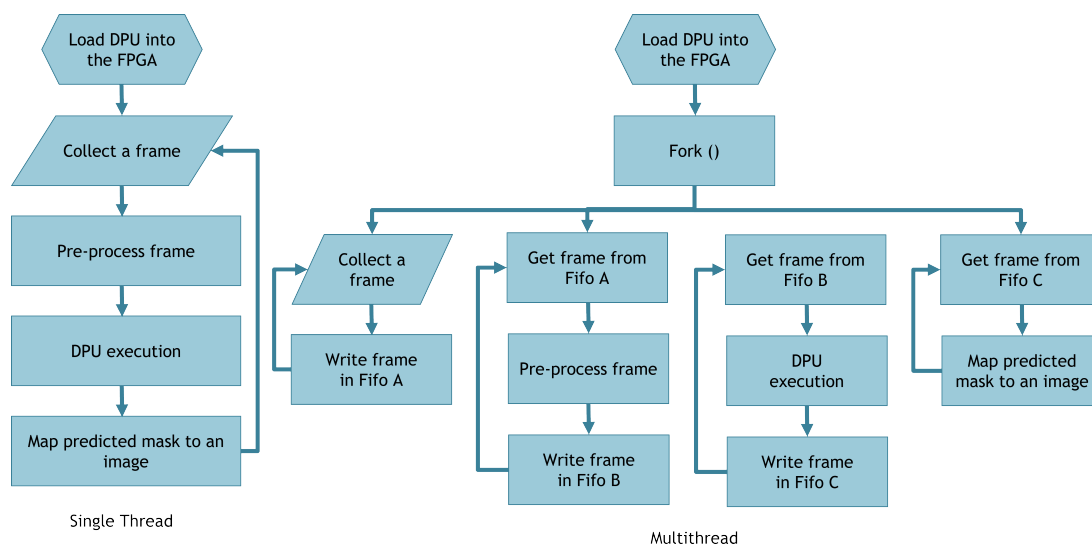


Fig. 7. Flow chart for single and multi-threading inference approaches

Table 1. Segmentation comparison for the different model implementations

Model	MCC	F1 score	Hafiane
Proposed Model Original (Validation)	<b>0.964</b>	0.964	<b>0.946</b>
Proposed Model Original (Test)	0.933	0.934	0.902
Proposed Model Pruned (Validation)	0.964	<b>0.965</b>	0.941
Proposed Model Pruned (Test)	0.924	0.926	0.877
Proposed Model FPGA (Validation)	0.932	0.933	0.899
Proposed Model FPGA (Test)	0.912	0.915	0.870

AXI connection, which is not illustrated here, is used for all communication between the PS, PL, and peripherals.

In our experiments, the overall performance of the model implemented using single-thread execution was not satisfactory, as we obtained only a throughput of 15.77 FPS, even after the pruning and quantization of the model. Therefore, we explored the use of a multi-thread approach supported by the Ultra89-v2 board.

The use of this functionality enabled us to attain a higher performance. The main limitation of the single-threaded approach is the bottleneck introduced by the DPU when performing inference in the FPFA, as it introduces a significant latency.

This problem arises from the use of queues for exchanging information among the different

threads. In Fig. 7, we provide a flow chart comparing both software implementations.

## 4 Results and Discussion

In the subsequent section, we will discuss the results obtained from implementing the U-Net model for segmenting images of the Corsican Fire Database, comparing both the original full-precision model and the optimized model running on the FPGA platform.

We will also compare our results with previous works in the state-of-the-art based on a number of metrics used in the literature, which will be described in the next subsection.

After this, quantitative and qualitative results will be provided, based on these metrics, followed by a discussion of the obtained results.



**Table 2.** Comparison of the proposed model (full-precision and FPGA implementation) with other models in the state-of-the-art

Model	MCC	F1 score	Hafiane
Akhloufi + Dice + NIR	0.910	0.915	0.890
Akhloufi + Focal Tversky + NIR	0.914	0.916	0.889
Akhloufi + Mixed focal + NIR	0.828	0.843	0.802
Proposed Model Original (Test)	<b>0.933</b>	<b>0.934</b>	<b>0.902</b>
Proposed Model FPGA (Test)	0.912	0.915	0.870

## 4.1 Comparison Metrics

### 4.1.1 Matthews Correlation Coefficient

First proposed by Matthews [17], it measures the correlation of the true classes with their predicted labels [5]. The MCC represents the geometric mean of the regression coefficient and its dual, and is defined as follows [24]:

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{N_- \cdot N_+ \cdot P_- \cdot P_+}}, \quad (1)$$

where:

$$\begin{aligned} N_- &= \text{TN} + \text{FN}, \\ N_+ &= \text{TN} + \text{FP}, \\ P_- &= \text{TP} + \text{FN}, \\ P_+ &= \text{TP} + \text{FP}, \end{aligned}$$

where TP is the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives.

### 4.1.2 F1 score

Also known as the Dice coefficient or overlap index [21], the F1 score is the harmonic mean of the precision Pr and recall Re. The F1 score is defined as the harmonic mean of Pr and Re as follows:

$$F1 = 2 \times \frac{\text{Pr} \times \text{Re}}{\text{Pr} + \text{Re}}. \quad (2)$$

### 4.1.3 Hafiane Quality Index

Proposed by Hafiane et al. [11] for fire segmentation evaluation, it measures the overlap between the ground truth and the segmentation results, penalizing as well the over- and under-segmentation [11]. First, the authors define a matching index  $M$  as follows [24]:

$$M = \alpha \sum_{j=1}^{NR^S} \frac{\text{Card}(R_{i^*}^{GT} \cap R_j^S) \times \text{Card}(R_j^S)}{\text{Card}(R_{i^*}^{GT} \cup R_j^S)}, \quad (3)$$

where  $\alpha = 1/\text{Card}(I^S)$  and  $NR^S$  is the number of connected regions in the segmentation result  $I^S$ .  $R_j^S$  represents one of the said regions, and  $R_{i^*}^{GT}$  is the region in the reference image  $I^{GT}$  that has the most significant overlapping surface with the  $R_j^S$  region. Next, Hafiane et al. define an additional index  $\eta$  to take into account the over- and under-segmentation as follows [24]:

$$\eta = \begin{cases} NR^{GT}/NR^S & \text{if } NR^S \geq NR^{GT}, \\ \log(1 + NR^S/NR^{GT}) & \text{otherwise.} \end{cases} \quad (4)$$

Finally, the Hafiane quality index is defined as follows:











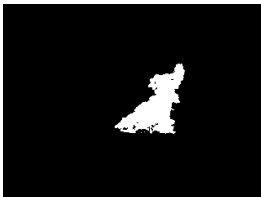

$$HAF = \frac{M + m \times \eta}{1 + m}, \quad (5)$$

where  $m$  is a weighting factor set to 0.5.

## 4.2 Quantitative Results

Table 1 shows the results obtained by the final implementation of the optimized model in the FPGA using MCC, HAF, and F1 score. It can be observed the pruned model presented a slight drop in performance (3% in MCC) whereas the FPGA

**Table 3.** Qualitative visual comparison of the segmented images produced by three model configurations: original (full-precision), pruned and quantized (FPGA implementation)

Image	Example 1	Example 2	Example 3
Ground truth			
Original model			
Pruned model			
FPGA model			

model presented a slightly higher drop (of about 5% both in MCC and F1 score) of performance for all metrics. This slight degradation is expected given the heavy optimization undergone by the model when passing from 64-bit to 8-bit data representation.

However, the gain in throughput (and thus inference time) is significant: the full precision model runs at 8 FPS in a GPU, consuming a large amount of power, whereas our model can attain up to 33.64 FPS in the selected FPGA when running in multi-threaded mode (15.77 FPS for the single-threaded mode), for a fraction of the power consumption. Table 2 provides a comparison with other models in the literature.

A recent and thorough comparison of the state-of-the-art carried out by Ciprián-Sánchez et al. [8] compared different architectures, image types, and loss functions on the Corsican Fire Database. Here, we compared the best model from this study (by Akhloufi et al. [2] with various losses) using the base metrics (i.e., MCC, HAF, and F1 score).

From the table, it can be observed that the original model outperforms this previous work by about 2% (0.933 MCC), whereas the FPGA implemented model attains a similar performance to the best configuration obtained by Akhloufi (0.912 vs 0.910 MCC), using a much smaller footprint.

### 4.3 Qualitative Results

Table 3 provides a qualitative comparison of the different models compared in Table 1.

It shows the original images of the Corsican Fire Database and the segmentation results using the original model before the optimization process, after the pruning method, and finally, the final model used in the FPGA.

It can be observed that for the 3 examples provided, both the pruned model and the FPGA implementation yielded practically the same results as the full-precision model, albeit at a much higher frame rate (33 FPS vs the 8 the U-Net running on a V100 GPU).

Such results can be used in the smart camera for higher image processing tasks in real-time, such as fire spread prediction by using the processing section (ARM processor) of the Ultra96-v2 platform.

## 5 Conclusions

In the present paper, we implement and analyze the performance of a smart camera system based on an FPGA accelerator.

A modified version of the U-Net architecture was used, to which optimization methods such as quantization and pruning were applied, effectively reducing the inference time and, at the same time, obtaining good results in the wildfire segmentation task. The frame rate obtained in the segmentation task was 33.63 FPS.

It is believed that there is still some potential to improve the speed of inference by using other strategies, such as the conversion of CNN models to spiking neural networks (SNN), whose conversion has been shown to reduce inference times by reducing the number of operations performed [13].

Finally, given the results obtained, heavy computational tasks are believed to benefit from the accelerators implemented in FPGAs for their use in real-time applications such as wildfire surveillance using drones.

## Acknowledgments

The authors wish to acknowledge the Mexican Council for Science and Technology (CONACYT) for the support in terms of postgraduate scholarships in this project, and the Data Science Hub at Tecnológico de Monterrey for their support on this project. This work was supported in part by the SEP CONACYT ANUIES ECOS NORD project 315597.

## References

1. **Abraham, N., Khan, N. M. (2019).** A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 683–687. DOI: 10.1109/ISBI.2019.8759329.
2. **Akhloufi, M. A., Booto-Tokime, R., Elassady, H. (2018).** Wildland fires detection and segmentation using deep learning. *Pattern Recognition and Tracking XXIX*, Vol. 10649, pp. 106490B. DOI: 10.1117/12.2304936.
3. **Berthelie, A., Chateau, T., Duffner, S., Garcia, C., Blanc, C. (2021).** Deep model compression and architecture optimization for embedded systems: A survey. *Journal of Signal Processing Systems*, Vol. 93, No. 8, pp. 863–878. DOI: 10.1007/s11265-020-01596-1.
4. **Bramberger, M., Doblender, A., Maier, A., Rinner, B., Schwabach, H. (2006).** Distributed embedded smart cameras for surveillance applications. *Computer*, Vol. 39, No. 2, pp. 68–75. DOI: 10.1109/MC.2006.55.
5. **Chicco, D., Tötsch, N., Jurman, G. (2021).** The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, Vol. 14, No. 1, pp. 13. DOI: 10.1186/s13040-021-00244-z.

6. **Choi, H. S., Jeon, M., Song, K., Kang, M. (2021).** Semantic fire segmentation model based on convolutional neural network for outdoor image. *Fire Technology*, Vol. 57, pp. 3005–3019. DOI: 10.1007/s10694-020-01080-z.
7. **Ciprián-Sánchez, J. F., Ochoa-Ruiz, G., Gonzalez-Mendoza, M., Rossi, L. (2021).** FIRE-GAN: a novel deep learning-based infrared-visible fusion method for wildfire imagery. *Neural Computing and Applications*, Vol. 35, No. 25, pp. 18201–18213. DOI: 10.1007/s00521-021-06691-3.
8. **Ciprián-Sánchez, J. F., Ochoa-Ruiz, G., Rossi, L., Morandini, F. (2021).** Assessing the impact of the loss function, architecture and image type for deep learning-based wildfire segmentation. *Applied Sciences*, Vol. 11, No. 15, pp. 7046. DOI: 10.3390/app11157046.
9. **Frizzi, S., Bouchouicha, M., Ginoux, J., Moreau, E., Sayadi, M. (2021).** Convolutional neural network for smoke and fire semantic segmentation. *IET Image Processing*, Vol. 15, No. 3, pp. 634–647. DOI: 10.1049/ipr2.12046.
10. **Gomes, P., Santana, P., Barata, J. (2014).** A vision-based approach to fire detection. *International Journal of Advanced Robotic Systems*, Vol. 11, No. 9, pp. 149. DOI: 10.5772/58821.
11. **Hafiane, A., Chabrier, S., Rosenberger, C., Laurent, H. (2007).** A new supervised evaluation criterion for region based segmentation methods. *Advanced Concepts for Intelligent Vision Systems*, Springer Berlin Heidelberg, pp. 439–448. DOI: 10.1007/978-3-540-74607-2\_40.
12. **Ioffe, S., Szegedy, C. (2015).** Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Vol. 37. DOI: 10.48550/arXiv.1502.03167.
13. **Ju, X., Fang, B., Yan, R., Xu, X., Tang, H. (2020).** An FPGA implementation of deep spiking neural networks for low-power and fast classification. *Neural Computation*, Vol. 32, pp. 182–204. DOI: 10.1162/neco\_a.01245.
14. **Li, H., Kadav, A., Durdanovic, I., Samet, H., Graf, H. P. (2017).** Pruning filters for efficient convnets. DOI: 10.48550/ARXIV.1608.08710.
15. **Li, H., Wu, X. J., Kittler, J. (2018).** Infrared and visible image fusion using a deep learning framework. *24th International Conference on Pattern Recognition*, pp. 2705–2710. DOI: 10.1109/ICPR.2018.8546006.
16. **Ma, J. (2020).** Segmentation loss odyssey. DOI: 10.48550/arXiv.2005.13449.
17. **Matthews, B. W. (1975).** Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, Vol. 405, No. 2, pp. 442–451. DOI: 10.1016/0005-2795(75)90109-9.
18. **Perrolas, G., Niknejad, M., Ribeiro, R., Bernardino, A. (2022).** Scalable fire and smoke segmentation from aerial images using convolutional neural networks and quad-tree search. *Sensors*, Vol. 22, No. 5, pp. 1701. DOI: 10.3390/s22051701.
19. **Ronneberger, O., Fischer, P., Brox, T. (2015).** U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241.
20. **Shi, Y., Real, F. D. (2009).** Smart cameras: Fundamentals and classification. In *Smart cameras*. Springer, pp. 19–34.
21. **Taha, A. A., Hanbury, A. (2015).** Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, Vol. 15, No. 1, pp. 29. DOI: 10.1186/s12880-015-0068-x.
22. **Talib, M. A., Majzoub, S., Nasir, Q., Jamal, D. (2021).** A systematic literature review on hardware implementation of artificial intelligence algorithms. *The Journal of*

- Supercomputing, Vol. 77, pp. 1897–1938. DOI: 10.1007/s11227-020-03325-8.
23. **Tchuitcheu, W. C., Bobda, C., Pantho, M. J. H. (2020).** Internet of smart-cameras for traffic lights optimization in smart cities. *Internet of Things*, Vol. 11, pp. 100207. DOI: 10.1016/j.iot.2020.100207.
  24. **Toulouse, T., Rossi, L., Akhloufi, M., Celik, T., Maldague, X. (2015).** Benchmarking of wildland fire colour segmentation algorithms. *IET Image Processing*, Vol. 9, No. 12, pp. 1064–1072. DOI: 10.1049/iet-ipr.2014.0935.
  25. **Toulouse, T., Rossi, L., Campana, A., Celik, T., Akhloufi, M. A. (2017).** Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Safety Journal*, Vol. 92, pp. 188–194. DOI: 10.1016/j.firesaf.2017.06.012.
  26. **United Nations Environment Programme (2022).** Spreading like wildfire the rising threat of extraordinary landscape fires. pp. 8–12.
  27. **Venieris, S. I., Leonidas-Kouris, A. S., Savvas-Bouganis, C. (2018).** Toolflows for mapping convolutional neural networks on FPGAs. *ACM Computing Surveys*, Vol. 51, No. 3, pp. 1–39. DOI: 10.1145/3186332.
  28. **Véstias, M. P. (2019).** A survey of convolutional neural networks on edge with reconfigurable computing. *Algorithms*, Vol. 12, No. 8, pp. 154. DOI: 10.3390/a12080154.
  29. **Wolf, W., Ozer, B., Lv, T. (2002).** Smart cameras as embedded systems. *Computer*, Vol. 35, No. 9, pp. 48–53. DOI: 10.1109/MC.2002.1033027.
  30. **Yeung, M., Sala, E., Schönlieb, C. B., Rundo, L. (2021).** Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. DOI: 10.48550/ARXIV.2102.04525.
  31. **Yuan, C., Zhang, Y., Liu, Z. (2015).** A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques. *Canadian Journal of Forest Research*, Vol. 45, No. 7, pp. 783–792. DOI: 10.1139/cjfr-2014-0347.

*Article received on 05/06/2023; accepted on 21/09/2023.  
Corresponding author is Miguel Gonzalez-Mendoza.*