Semantic segmentation and thermal imaging for forest fires detection and monitoring by drones

Mimoun Yandouzi[1], Mohammed Berrahal[2], Mounir Grari[3], Mohammed Boukabous[3], Omar Moussaoui[3], Mostafa Azizi[3], Kamal Ghoumid[1], Aissa Kerkour Elmiad[4]
[1]LSI Research Lab, ENSAO, Mohammed First University, Oujda, Morocco
[2]MCL Research Lab, PFS, Cadi Ayyad University, Safi, Morocco
[3]MATSI Research Lab, ESTO, Mohammed First University, Oujda, Morocco
[4]LARI Research Lab, FSO, Mohammed First University, Oujda, Morocco

Article Info

Article history:

Received Oct 4, 2023
Revised Feb 18, 2024
Accepted Mar 6, 2024

Keywords:

Deep learning
Forest fires
Mask region convolutional neural network
Segmentation
Thermal images
UAV (drones)
You only look once

ABSTRACT

Forest ecosystems play a crucial role in providing a wide range of ecological, social, and economic benefits. However, the increasing frequency and severity of forest fires pose a significant threat to the sustainability of forests and their functions, highlighting the need for early detection and swift action to mitigate damage. The combination of drones and artificial intelligence, particularly deep learning, proves to be a cost-effective solution for accurately and efficiently detecting forest fires in real-time. Deep learning-based image segmentation models can not only be employed for forest fire detection but also play a vital role in damage assessment and support reforestation efforts. Furthermore, the integration of thermal cameras on drones can significantly enhance the sensitivity in forest fire detection. This study undertakes an in-depth analysis of recent advancements in deep learning-based semantic segmentation, with a particular focus on model's mask region convolutional neural network (Mask R-CNN) and you only look once (YOLO) v5, v7, and v8 variants. Emphasis is placed on their suitability for forest fire monitoring using drones equipped with RGB and/or thermal cameras. The conducted experiments have yielded encouraging outcomes across various metrics, underscoring its significance as an invaluable asset for both fire detection and continuous monitoring endeavors.

Corresponding Author:

Mimoun Yandouzi
Engineering Sciences Laboratory (LSI), National School of Applied Sciences (ENSO)
Mohammed First University
Oujda 60000, Morocco
Email: m.yandouzi@ump.ac.ma

1. INTRODUCTION

The scourge of wildfires inflicts devastating damage worldwide, impacting not only natural ecosystems but also human communities. Apart from annihilating vegetation and wildlife across thousands of hectares, these fires release significant CO2 into the atmosphere, exacerbating global warming. This threat is particularly acute in Mediterranean regions, where hot and arid conditions foster the rapid spread of flames. Climate change renders these areas increasingly vulnerable to extended heatwaves, heightening the risk of catastrophic wildfires. Addressing this challenge requires an integrated approach involving enhanced prevention measures, sustainable land and forest management, as well as increased awareness of responsible human behavior concerning forest fires [1].

Early detection of forest fires is crucial for ecosystem preservation and public safety. Utilizing drones and deep learning has proven highly effective, overcoming the limitations of traditional methods. Advances in computer vision and deep learning allow algorithms to automatically identify fire indicators in drone-captured images, enabling swift detection of emerging fires. Drones' ability to cover challenging terrains amplifies monitoring and significantly improves forest fire detection and response efficiency [2].

Forest fire detection by drones relies on three main families of deep learning techniques, which are the result of a previous literature review [2]. Image classification categorizes drone-captured images into fire or non-fire classes. Object detection identifies flames and/or smoke in scenes, pinpointing their location accurately. Segmentation divides images, accurately outlining fire and smoke areas. These complementary deep learning approaches provide vital tools for precise fire detection, contributing to a rapid and targeted response during emergencies [3].

On the other hand, thermal images offer a unique perspective, enabling naked-eye detection of forest fires where traditional RGB cameras fall short. The heat emitted by fires creates visual contrasts in the infrared spectrum, captured by thermal cameras. This capability is crucial for disaster prevention and management [4]. To fully harness this technology, in-depth studies in computer vision are necessary, with deep learning algorithms tailored specifically for thermal image analysis. Integrating the richness of thermal data with deep learning capabilities powers advanced detection systems, significantly enhancing monitoring and swift response to forest fires. This contributes to environmental preservation and public safety.

This paper broadens our study scope after addressing image classification and object detection. We shift focus to semantic segmentation, exploring effective deep learning models. Notably, our analysis extends beyond conventional RGB images, encompassing thermal images for a comprehensive understanding.

2. BACKGROUNDS

2.1. Image segmentation

Segmentation is a fundamental technique in the field of computer vision. It involves dividing an image into meaningful regions, allowing for analysis and understanding of the various elements within it. Segmentation provides fine granularity by identifying and isolating objects, boundaries, and structures in an image, facilitating further analysis [5].

Segmentation finds numerous applications in various domains of computer vision. In medicine, for example, medical image segmentation enables the localization and delineation of anatomical structures such as organs, tumors, or lesions, aiding in diagnosis and patient monitoring [6]. In the field of autonomous driving, segmentation is used to detect pedestrians, vehicles, and obstacles on the road, contributing to safe decision-making by autonomous vehicles [7]. Segmentation is also utilized in augmented reality to realistically overlay virtual objects onto real-world images. It is useful in video surveillance for the detection of suspicious activities or facial recognition. In robotics, segmentation can enable precise object manipulation by identifying and locating objects [8].

Segmentation is very beneficial for forest fire detection and monitoring since it may offer information on the fire's size and impacted areas. It permits the discrimination of fire and non-fire regions in a picture, as well as the differentiation between different types of fires, such as controlled burns and uncontrolled wildfires [9]. Segmentation techniques include instant segmentation, semantic segmentation, and panoptic segmentation [10]:

- Instant segmentation is ideal for real-time forest fire monitoring and early detection. It can rapidly analyze pictures acquired by drones and provide real-time information about the fire's size and growth, allowing authorities to respond swiftly and avert widespread damage.
- Semantic segmentation is optimal for detecting the fire's extent and affected areas. It gives information regarding the semantic significance of the regions inside a picture, such as fire or non-fire regions, allowing authorities to make well-informed decisions regarding the best course of action (Figure 1).
- Panoptic segmentation is best suited for providing a complete representation of a picture, including the location and size of the fire, as well as the objects and areas within the image. This type of segmentation can provide extensive information on the fire, making it important for making decisions regarding forest fire response.

The optimal segmentation technique for forest fire detection will depend on the task's specific criteria, such as the necessity for real-time monitoring, the identification of the fire's extent, or a full image representation. In our research, we will concentrate on the extent of the fire, therefore semantic segmentation will be our method of choice.

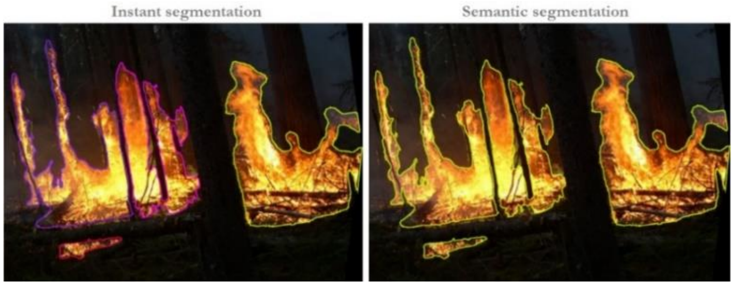

Figure 1. Instant segmentation and semantic segmentation example

2.2. Thermal images

Aerial photography or videography remains the most commonly performed tasks with drones. For this purpose, drones are equipped with one or multiple cameras, enabling them to capture their surroundings from the air. The possibilities for photographic equipment are vast, ranging from small action cameras to professional cameras with interchangeable lenses [11].

In the field of cameras, there are two main types of sensors: standard camera sensors and thermal camera sensors. Standard camera sensors, also known as visible-light cameras or RGB cameras, capture images using the visible spectrum of light with red, green, and blue channels to create full-color images. These cameras are commonly used for regular photography and videography in typical lighting conditions. On the other hand, thermal camera sensors, also called infrared cameras, detect infrared radiation emitted by objects based on their temperature, allowing visualization of heat variations in a scene [12]. Thermal cameras are valuable for applications like night vision, firefighting, search and rescue operations, industrial inspections, and identifying heat leaks in buildings. Each type of camera sensor serves specific purposes and is utilized across various industries accordingly.

The thermal sensitivity of a thermal camera measures its ability to detect subtle temperature differences, thereby enabling visualization of objects with minimal thermal variations. These cameras can operate in different ranges of the infrared spectrum, typically categorized as near-infrared (NIR), mid-infrared (MIR), and far-infrared (FIR), each having specific applications. However, thermal images may be influenced by weather conditions, absorption, and reflection of infrared radiation, as well as the distance between the object and the camera, which can impact their quality and accuracy [13].

In scenarios where conventional RGB images encounter limitations due to dense vegetation and inherent variations in lighting within forest environments, thermal imaging emerges as a promising avenue for wildfire detection (Figure 2). By virtue of their capability to capture the distinctive heat emissions associated with combustion activities, thermal images transcend the constraints of traditional visual methods. By circumventing the visual impediments posed by dense foliage, thermal images enable a more accurate detection of fire hotspots, thereby facilitating early identification and swift response in managing emergency situations related to forest fires [14].

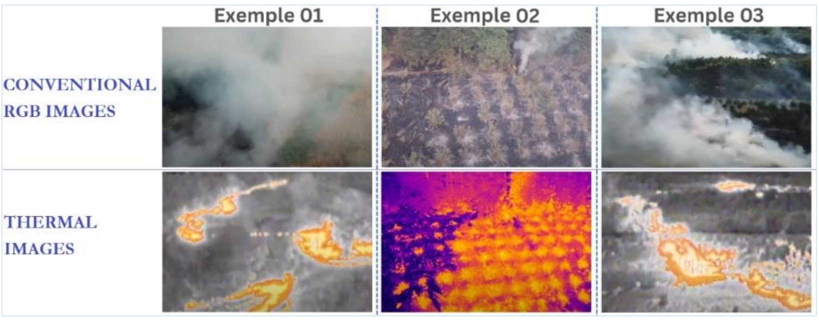

Figure 2. Advantages of thermal imaging for forest fire detection over RGB conventional images

3. RELATED WORK

Zhao *et al*. [15] introduced a groundbreaking saliency detection algorithm aimed at rapid core fire area localization and segmentation in aerial images. The suggested approach efficiently prevents feature loss resulting from direct resizing, making it suitable for data augmentation and creating the 'UAV_Fire' dataset, a standard collection of fire images from drones. Next, they introduce 'Fire_Net,' a 15-layered DCNN architecture designed to serve as both a self-learning fire feature extractor and classifier. After evaluating various architectures and several essential parameters (such as drop-out ratio, and batch size) of the DCNN model concerning its validation accuracy, the suggested design surpassed previous approaches by achieving an impressive overall accuracy of 98%. Moreover, 'Fire_Net' ensured a remarkable average processing speed of 41.5 ms per image, enabling real-time wildfire inspection.

Bulatov and Leidinger [16] used instance segmentation to accurately monitor and analyze the spatial and temporal distribution of deadwood, a significant factor in forest fires. This study highlights the impressive potential of achieving precise instance segmentation in both RGB and elevation domains, even with limited training data. By employing a high-performance Mask region convolutional neural network (Mask R-CNN) model, the researchers effectively mapped standing and fallen deadwood instances in German forests. The results were remarkable, achieving an impressive overall accuracy of 92.4%.

Tran *et al*. [17] used images collected from a forest fire in Andong, the Republic of Korea, in April 2020. Following that, they implemented two-patch-level deep-learning models. The first network, trained using the UNet++ architecture, operated at the patch-level 1. The output predictions from this initial network were then utilized as positional input for the second network, which employed UNet and refined the results based on the reference position from the first network. Subsequently, the robustness of the proposed method was demonstrated by comparing its final performance with state-of-the-art image-segmentation algorithms. Additionally, a comparative study of the loss functions was conducted. The proposed approach has certain limitations, including the need to train dual patch-level models on different locations and weather conditions for improved performance. Furthermore, its current local processing requires conversion into an online platform to enhance practicality and reduce time consumption.

Muksimova *et al*. [18] undertook a groundbreaking project, developing an innovative model with two pathways that encodes and decodes information. This model aims to detect and accurately outline wildfires and smoke in real-time images captured by UAVs. The key innovation lies in their use of a nested decoder, which includes pre-activated residual blocks and an attention-gating mechanism. This approach substantially improves the accuracy of segmentation. To enhance the model's reliability and its ability to work across different scenarios, the researchers meticulously gathered a new dataset comprising actual incidents of forest fires and smoke, ranging from small to large areas. Their efforts paid off, as experimental results clearly demonstrated the exceptional performance of their method. It outperformed existing techniques for detection and segmentation while maintaining a lightweight design.

Garcia *et al*. [19] conducted a study focusing on using various unsupervised segmentation methods to segment aerial thermal images of forest fires. In their research, they introduced an innovative approach that combines both global and local information through a multilayer level set formulation. At the global level, their method aimed to minimize an energy functional based on the widely-used Chan-Vese method, which is commonly used in image processing and computer vision tasks. The Chan-Vese model works by minimizing an energy function that considers both regions inside and outside the region of interest. To evaluate the effectiveness of their approach, they compared it with other common unsupervised segmentation methods, and the results were promising. Their method outperformed the alternatives, achieving impressive accuracy (96.2%), precision (87.1%), and intersection over union (77.5%). The evaluation involved comparing the outputs of their method with hand-drawn labels provided by the Portuguese Air Force, which served as the ground truth for the experiment. Despite its remarkable performance, their proposed method currently lacks real-time segmentation capabilities, with an inference time of 2.86 seconds per image. Nonetheless, their research represents a significant advancement in addressing the challenging task of segmenting aerial thermal images for forest fire analysis.

4. PROPOSED METHOD

4.1. Methodology and target architecture

In the context of forest fire detection and monitoring, the objective of this research is to investigate the adoption of you only look once (YOLO) and Mask R-CNN for segmentation. Particularly, we will experiment with several variations of YOLO to determine which version would be most suited for our purpose, taking accuracy and speed into account. In addition, we will investigate the segmentation capabilities of Mask R-CNN and contrast its efficacy to that of YOLO variants.

To identify the most precise and efficient model suitable for deployment on an AI-IoT-enabled device, we will test multiple iterations of YOLO and Mask R-CNN. These include YOLOv5 [20], YOLOv7 [21], YOLOv8 [22], and Mask R-CNN with ResNet50 [23]. The choice of Resnet50 as the backbone for the Mask R-CNN model was justified in a previous work [24]. Our experimentation will entail the training and evaluation of these models using a dataset comprising aerial photos captured by drones in forested environments. This dataset will feature a diverse range of fire scenarios, including controlled burns, uncontrolled wildfires, and non-fire images, thus enriching the model's versatility. We will conduct analyses on all explored models using both conventional RGB images and thermal imagery.

Once the optimal model is selected, it will be deployed on an AI-IoT-enabled device placed in a fog environment. Utilizing the RTMP protocol, the device will receive a video feed from the drone, enabling real-time forest fire detection and monitoring. The device will analyze the video feed using the deployed model to detect fire zones. When a fire is detected, the device will transmit a notification along with photographs and videos of the affected area to the cloud (authorities). The cloud will then provide a consolidated platform for authorities' surveillance and decision-making (see Figure 3).

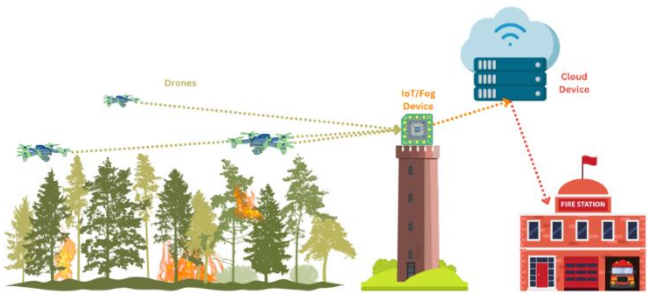

Figure 3. Proposed monitoring and detection system

Whether it's the Mask R-CNN model or the YOLO variants, for normal or thermal RGB images, all models typically operate through a five-step process to achieve accurate image segmentation (Figure 4). Firstly, input images are provided to the model (step 1). Next, the model performs object detection to identify fire regions within the image (step 2). Following this detection step, segmentation is applied individually to each bounding box encompassing these regions, breaking down each object into its distinct components (step 3). Once the boxes are segmented, the outcomes are integrated back into the overall image, yielding a comprehensive and precise representation of objects and their boundaries within the original image (step 4). Lastly, output images are generated, showcasing a detailed and accurate segmentation of all objects present in the original images, thereby enhancing the understanding and analysis of visual scenes (step 5).

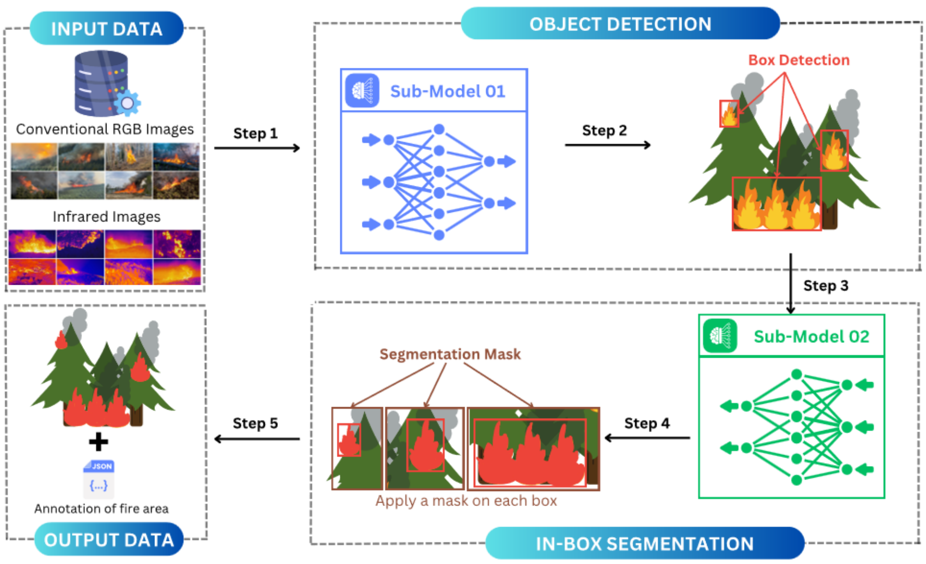

Figure 4. Methodology for experimental research

4.2. Datasets collection and preprocessing

The primary phase of our proposed method involves assembling a dataset of images depicting instances of forest fires. This dataset comprises two distinct collections: one containing RGB images, and the other containing thermal images. The first dataset, which includes after data augmentation a total of 4236 labeled images, incorporates photographs taken from ground-level cameras as well as aerial drones. These ground-level cameras have been employed to capture intricate real-time portrayals of forest fires, some of which were taken within our campus. In contrast, the aerial drone images offer a wider perspective, capturing expansive forested areas to facilitate monitoring of the fire's scope and magnitude. Additionally, we have incorporated images from publicly accessible datasets, such as those found in online image repositories and websites, to diversify and enhance the dataset with authentic depictions of real-world forest fire scenarios. For the second dataset, we combined the limited available thermal images sourced online with images simulating fires on our university campus. The augmented dataset for this second collection comprises a total of 1250 images. It's important to highlight our thorough measures to ensure a comprehensive range of images. These images encompass diverse fire types, varying intensities of fires, and an array of environmental settings where these incidents occurred. This strategic approach was taken to bolster the model's ability to accurately detect fires under a multitude of conditions, thus leading to an enhancement in its overall performance and generalization capacity. We streamlined the labeling process effectively by harnessing the power of Darwin, a robust training data platform widely recognized for its ability to create impactful AI solutions [25]. Packed with an array of exciting features, this platform simplifies tasks related to training data, dataset management, and model training. This venture entailed a significant investment of time and effort from our project team members. To optimize the workflow, we distributed the dataset among team members, enabling a smooth and swift resolution of the task.

Our method involves preprocessing and augmenting collected datasets. Preprocessing includes resizing images to 640x640 and converting them to JPG format. This is essential for successful image segmentation. Augmentation adds distortions like rotations and flips to enhance dataset variety and system robustness. Resizing and augmenting data increase segmentation success chances.

4.3. Deep learning models for segmentation

We will examine two classes of deep learning models used for segmentation: YOLO, encompassing versions v5, v7, and v8, as well as Mask R-CNN: YOLO, specifically in versions YOLO v5, v7, and v8, has ushered in a significant evolution within the realm of computer vision. Originally designed for object detection, these iterations have now transcended their initial purpose by incorporating image segmentation capabilities. This fusion of semantic segmentation represents a pivotal advancement in the field, seamlessly combining object detection with image segmentation. YOLO v5, v7, and v8 have successfully unified the extraction of information related to detected objects and their precise contours. This breakthrough unlocks a myriad of possibilities, from enhancing automated surveillance systems to enabling advanced environmental analysis. These developments underscore the unwavering dedication of the computer vision community to continually push the boundaries of innovation [26].

Mask R-CNN stands as a pivotal milestone in the realm of computer vision, representing a historic breakthrough by seamlessly amalgamating object detection and semantic segmentation. Emerged in 2017, Mask R-CNN evolved from its predecessor, faster R-CNN, by introducing a segmentation branch to the network, enabling precise identification and delineation of detected objects. The underlying principal hinges on harnessing convolutional neural networks to first detect object bounding boxes, then refining these detections by generating binary masks that intricately outline individual objects. This fusion of object detection and segmentation revolutionized applications like multi-instance object detection, object recognition, and contextual understanding. The evolution of Mask R-CNN has paved the way for ongoing enhancements in neural network architectures for computer vision, driving significant progress in how machines comprehend and interact with the visual world around them [27].

4.4. Models training and evaluation

In our proposed method, we proceed with training segmentation models on preprocessed and augmented datasets. The datasets are divided into train (70%), validation (20%), and test (10%) subsets. Various models including YOLOv5, v7, and v8, as well as Mask R-CNN, are employed. We opt for Mask R-CNN due to its efficiency and precision in handling complex images, learning high-level features with rapid inference speed [23]. YOLO, particularly its newer versions, is also recognized for its speed and accuracy [28]. The training process utilizes labeled data containing fire bounding boxes and class information. Mask R-CNN relies on tensorflow record (TFRecord) files for input, whereas YOLO uses TXT annotations and YAML config files. The primary objective here is to establish a dependable and highly accurate model for forest fire detection. Subsequently, the models' performance is assessed using the test set (10%) from the collected dataset. This evaluation phase is crucial as it gauges each model's performance and inference speed on previously unseen data, providing insights into their overall efficiency in detecting diverse forest fire types.

5. RESULTS AND DISCUSSIONS

5.1. Hardware characteristics

In the model training process, we used TensorFlow v2.13.0 [29], an open-source data analysis and machine learning software library, on an HPC system with powerful hardware specifications, including 2x Intel Gold 6148 (2.4 GHz/20 cores) CPUs and 2x NVIDIA Tesla V100 graphics cards, each with 32GB of RAM. These specifications were necessary to provide us with the required computational power to train our deep learning models effectively. To further enhance the model's performance and prevent overfitting, we employed transfer learning techniques and data augmentation during training.

The simulated fires on the university campus were filmed by a DJI Mavic Air drone and the team members' mobile phone camera. For thermal image acquisition, we utilized a Fluke Ti90 camera from Fluke Corporation, renowned for its measurement and testing instruments. The Fluke Ti90 offers an impressive temperature range of -20°C to 250°C (-4°F to 482°F) and features a high-quality LCD screen for real-time thermal image display, aiding in thermal issue identification. Its infrared optics with a 35° x 26° field of view enable precise targeting of specific measurement areas, while its thermal sensitivity of 0.15°C at 30°C (0.27°F at 86°F) allows for detecting subtle temperature variations. Additionally, it offers advanced thermal imaging features such as visual and thermal image overlay for in-depth data analysis. For future deployments, we're considering the Raspberry Pi 4 for its popularity, cost-effectiveness, and machine learning capabilities, featuring a quad-core ARM Cortex-A72 processor, 4 GB of RAM, and various connectivity options like ethernet, Wi-Fi, and Bluetooth.

5.2. Metrics for segmentation evaluation

The presence of a suitable evaluation metric holds significant importance in discerning the most optimal model throughout the training phase. Utilizing distinct metrics becomes imperative when evaluating deep learning models [30]. For image segmentation:

- True positive (TP): pixels that the model accurately classifies as being members of the target object class.
- True negative (TN): pixels that the model correctly recognizes as not being part of the target object class.
- False positive (FP): pixels that the model erroneously categorizes as being part of the target object class, even when they are not.
- False negative (FN): pixels that belong to the target object class but are inaccurately labeled as not belonging to it by the model.

It's important to note that in segmentation, the primary focus is on TP and FP, as they are used to calculate precision and recall, which are metrics evaluating the quality of the model's segmentation performance.

Precision: precision refers to the accuracy of accurately identifying and labeling only the relevant pixels as belonging to the target object or class [31]. Precision is defined as (1):

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

Recall: recall, or true positive rate, is the ratio of correctly identified positive pixels (true positives) to the total number of actual positive pixels, indicating the model's ability to capture and identify the entirety of the target object class [31]. It's calculated as (2):

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

F1 score: the F1 score is a harmonic metric that combines precision and recall, thereby providing a balanced trade-off between the model's ability to accurately identify positive instances and its ability to avoid false negative classifications [31]. The F1 score is calculated as (3):

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

Inference time: the inference time is the time it takes for a deep learning model to analyze a single image and provide a prediction, specifically the elapsed time between inputting the image into the model and outputting the prediction [31].

Average precision (AP): AP is a metric that assesses the quality of object delineation and classification in image segmentation. Its formula combines precision and recall at various similarity thresholds between predictions and annotations [32]. The formula for average precision is (4):

$$AP = \sum_{k=0}^{k=n-1}[Recalls(k) - Recalls(k+1) * Precisions(k)] \quad (4)$$

Where recalls(n)=0, precisions(n)=1, and n is number of thresholds

Mean average precision (mAP): mAP evaluates a segmentation model's performance holistically by computing the AP across various object classes. This metric is commonly presented at multiple confidence thresholds, such as 0.5 and 0.95, to offer a comprehensive understanding of the model's accuracy across different levels of confidence [32].

$$mAP = \frac{1}{N}\sum_{i=1}^{N} APi \quad (5)$$

Where APi is the AP of class i and N is the number of classes.

Intersection over union (IoU): IoU also known as the Jaccard Index, is a crucial metric in image segmentation that gauges the degree of overlap between a predicted object and its corresponding ground truth annotation. This is achieved by dividing the intersection area of the predicted region and the ground truth region by their combined area. IoU offers insight into how well a model's predicted segmentation aligns with the real object boundaries, making it a pivotal measure of segmentation accuracy and spatial precision [32].

$$IoU = \frac{(Intersection\ Area)}{(Union\ Area)} \quad (6)$$

Where the "intersection area" is the overlapping region between the predicted and ground truth segments, and the "union area" is the combined area of both the predicted and ground truth segments.

5.3. Experimental results

In Table 1 and Figure 5, we present the results obtained for various metrics, including accuracy, precision, recall, Intersection over Union, mAP at confidence thresholds of 0.5 and 0.95, as well as the inference time for each model, both for standard RGB and thermal images. To maximize effectiveness, we trained all of the models for seventy epochs.

Table 1. Obtained results for the implemented models

Deep learning model	Stage	IoU %	mAP @0.5 %	mAP @0.95 %	Inference s/Image	Precision %	Recall %	F1-score %
Mask R-CNN real images	Boxing	91.02	97.93	97.53	~0.028	99.13	98.89	99.01
	Segmentation	92.14	97.91	96.42	~0.0046			
Mask R-CNN ifrared images	Boxing	91.11	98.52	98.12	~0.028	99.17	99.07	99.12
	Segmentation	92.32	98.48	94.56	~0.0046			
YOLOv5 real images	Boxing	88.94	95.43	92.04	~0.0051	98.18	97.95	98.06
	Segmentation	89.02	95.25	84.20	~0.0039			
YOLOv5 ifrared images	Boxing	89.04	95.49	88.23	~0.0051	98.23	97.86	98.04
	Segmentation	88.91	92.37	83.01	~0.0039			
YOLOv7 real images	Boxing	89.17	95.10	95.96	~0.0027	98.35	98.17	98.26
	Segmentation	89.14	94.90	85.00	~0.0037			
YOLOv7 infrared images	Boxing	89.18	95.42	93.63	~0.0027	98.31	98.06	98.18
	Segmentation	88.72	95.23	88.83	~0.0037			
YOLOv8 real images	Boxing	89.36	96.50	96.28	~0.0011	98.47	98.22	98.34
	Segmentation	90.54	96.37	89.61	~0.0027			
YOLOv8 infrared images	Boxing	89.41	96.46	96.19	~0.0011	98.45	98.19	98.32
	Segmentation	90.16	96.36	93.31	~0.0027			

The primary observation to emphasize is that all the models analyzed showed the potential to be trained for forest fire segmentation, regardless of whether utilizing RGB or thermal images. However, there was variance in the convergence time, with YOLO models notably demonstrating faster training, reaching convergence in fewer than ten epochs. Despite employing a substantially smaller number of images, the results achieved with thermal images were nearly comparable to those obtained through RGB training for all the models examined. Specifically, the thermal image dataset comprised less than one-third of the images used for RGB training.

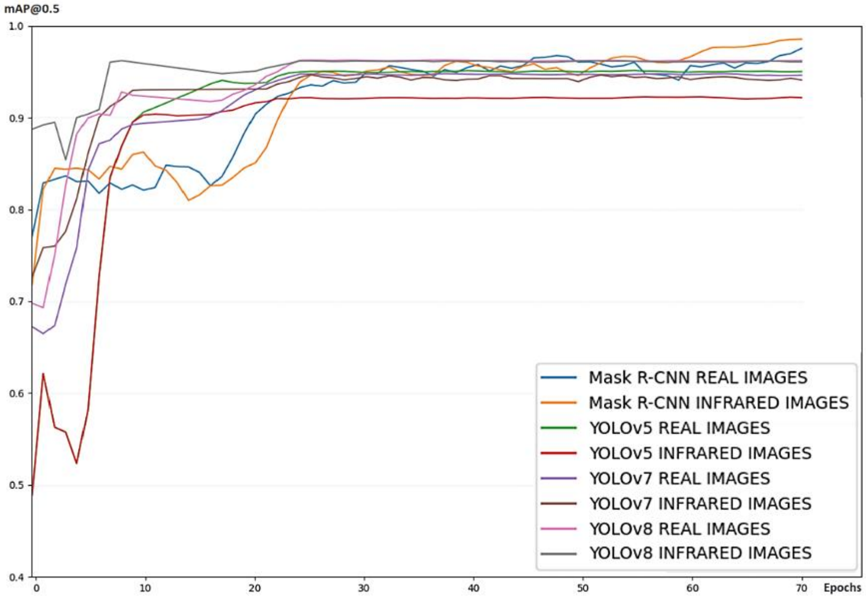

Figure 5. Achieved mAP@0.5 over epochs on validation set for the implemented models

The Mask R-CNN model excels in terms of image segmentation performance in both scenarios. For RGB images, it achieves an mAP@0.5 of 97.93% and an mAP@0.95 of 97.53% for object detection "Fire", along with an mAP@0.5 of 97.91% and an mAP@0.95 of 96.42% for fire segmentation, accompanied by an impressive F1 score of 99.01%. In the case of thermal images, the results are equally impressive, with an mAP@0.5 of 98.52% and an mAP@0.95 of 98.12% for object detection "Fire," as well as an mAP@0.5 of 98.48% and an mAP@0.95 of 94.56% for fire segmentation, and an outstanding F1 score of 99.12%. Mask R-CNN stands out as the top-performing model among all the models examined. However, it also demands the highest computational resources, requiring approximately 0.028 seconds per frame for object detection and around 0.0046 seconds per frame for the segmentation operation. This translates to an overall processing time of 0.0346 seconds per frame for the entire image analysis pipeline.

All variants of the YOLO model delivered slightly lower performance compared to the Mask R-CNN model, but offered significantly improved inference times. These results remain highly convincing, particularly those achieved by the YOLOv8 model, boasting an F1 score of around 98.3% for both RGB and thermal images. The model achieved a mAP@0.95 of 96.28% and 89.61% for object detection and segmentation, respectively, with RGB images, and 96.19% and 93.31% with infrared images. YOLOv8's performance is quite commendable, especially considering its rapid total inference time of 3.8 ms/image, making it an ideal choice for real-time detection.

Across the majority of models under examination, the object detection phase consistently exhibits longer processing times compared to the segmentation stage, where the primary objective is to identify the fire's outline within the designated area. Notably, the segmentation time remains relatively consistent among different models and is largely unaffected by the choice of RGB or infrared imagery. Noteworthy advancements in both performance and speed within the object detection phase, particularly with the YOLOv8 model, have been observed. Figure 6 showcases several examples of normal and thermal RGB forest fire image segmentation carried out by the YOLOv8 model.

While the speed of segmentation remains a crucial factor for achieving real-time processing capabilities on board a drone, our focus in this discussion is primarily directed towards the metric known as mAP@0.95. This choice stems from our broader exploration of segmentation's utility, which extends beyond merely fire detection. Instead, we aim to harness segmentation for comprehensive purposes, such as assessing post-fire damage and contributing to reforestation initiatives. In scenarios like these, the drone must operate at elevated altitudes to effectively survey expansive regions. Consequently, the segmentation process must attain a near-perfect level of precision to faithfully represent various aspects, including the extent of burnt areas and the count of trees burned. This precision is paramount to the accuracy of our mission, as it enables us to gather invaluable data for efficient post-fire rehabilitation and reforestation efforts.

When assessing the use of infrared imagery, the results obtained are promising, especially considering the scarcity of training images and the inherent challenges of thermography. It's important to recognize that thermal imaging cameras don't consistently capture, process, and display temperature data uniformly. As a result, the variations in color seen in thermal images primarily arise from differences in camera settings, color palette choices, and temperature variations within the scene being depicted.

Figure 6. Forest fire segmentation - examples using YOLOv8

6. CONCLUSION

The combination of drones and deep learning proves cost-effective for the precise and efficient real-time detection of forest fires. Deep learning-based image segmentation models not only play a pivotal role in fire detection but also in thorough environmental analysis. This advanced technological approach facilitates ecosystem monitoring, thus contributing to nature preservation and damage assessment from wildfires. Moreover, the integration of thermal cameras on drones can significantly enhance forest fire detection by boosting thermal sensitivity. This study explores recent advancements in deep learning-based semantic segmentation, with a specific focus on the Mask R-CNN and YOLO versions 5, 7, and 8 models. It highlights their significance in the context of forest fire monitoring, particularly when deployed on drones equipped with both RGB and thermal cameras. All models under scrutiny exhibit positive performance across various metrics, thereby establishing themselves as promising tools for semantic segmentation of forest fire images. While Mask R-CNN is slower than all YOLO variants, it excels in terms of image segmentation performance in both RGB and thermal image scenarios, achieving a remarkable F1 score exceeding 99%, along with an mAP@0.5 of nearly 97.5% for "Fire" object detection step and 98% for segmentation within the bounding boxes. YOLO models offer commendable performance coupled with exceptional inference speeds, rendering them optimal selections for real-time detection tasks. Notably, YOLOv8 boasts an impressive overall inference time of just 3.8 milliseconds per image. For the use of infrared imagery, the results obtained are quite promising, especially when considering the scarcity of training images and the inherent challenges of thermography.

As future work, we intend to design and implement a comprehensive architecture for the detection and monitoring of forest fires. This architecture will rely on synergy between IoT sensor networks, drone networks, and advanced deep learning algorithms.

ACKNOWLEDGEMENTS

This contribution is a component of the "SDF-RCSF: Low-Cost, Real-Time Forest Fire Detection System Using Wireless Sensor Networks" project, sponsored by Mohammed First University. The computational assets utilized were sourced from HPC-MARWAN, which is made available through the National Center for Scientific and Technical Research (CNRST), Rabat, Morocco.

REFERENCES

[1] M. Grari *et al.*, "IoT-Based Approach for Wildfire Monitoring and Detection," *International Conference on Advanced Intelligent Systems for Sustainable Development*, pp. 205–213, 2023, doi: 10.1007/978-3-031-35251-5_19.
[2] M. Yandouzi *et al.*, "Review on forest fires detection and prediction using deep learning and drones," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 12, pp. 4565–4576, 2022.
[3] M. Yandouzi *et al.*, "A Lightweight Deep Learning Model for Forest Fires Detection and Monitoring," *Proceedings of the 3rd International Conference on Electronic Engineering and Renewable Energy Systems*, pp. 697–705, 2023, doi: 10.1007/978-981-19-6223-3_71
[4] E. A. Sekehravani, E. Babulak, and M. Masoodi, "Flying object tracking and classification of military versus nonmilitary aircraft," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1394–1403, Aug. 2020, doi: 10.11591/EEI.V9I4.1843.

[5] R. Yogamangalam and B. Karthikeyan, “Segmentation techniques comparison in image processing,” *International Journal of Engineering and Technology*, vol. 5, no. 1, pp. 307–313, 2013.

[6] A. I. Sapitri, S. Nurmaini, Sukemi, M. N. Rachmatullah, and A. Darmawahyuni, “Segmentation atrioventricular septal defect by using convolutional neural networks based on U-NET architecture,” *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 553–562, Sep. 2021, doi: 10.11591/IJAI.V10.I3.PP553-562.

[7] C. Hwang, J. Jeong, and H. Jung, “Pine wilt disease spreading prevention system using semantic segmentation,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2666–2673, Jun. 2021, doi: 10.11591/IJECE.V11I3.PP2666-2673.

[8] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, “Review the state-of-the-art technologies of semantic segmentation based on deep learning,” *Neurocomputing*, vol. 493, pp. 626–646, Jul. 2022, doi: 10.1016/J.NEUCOM.2022.01.005.

[9] A. Kherraki, M. Maqbool, and R. El Ouazzani, “Traffic Scene Semantic Segmentation by Using Several Deep Convolutional Neural Networks,” *2021 3rd IEEE Middle East and North Africa COMMunications Conference (MENACOMM)*, pp. 1–6, Dec. 2021, doi: 10.1109/MENACOMM50742.2021.9678270.

[10] A. Al Mamun, E. P. Ping, and J. Hossen, “An efficient encode-decode deep learning network for lane markings instant segmentation,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, pp. 4982–4990, Dec. 2021, doi: 10.11591/IJECE.V11I6.PP4982-4990.

[11] K. AL-Dosari, Z. Hunaiti, and W. Balachandran, “Systematic Review on Civilian Drones in Safety and Security Applications,” *Drones*, vol. 7, no. 3, Mar. 2023, doi: 10.3390/DRONES7030210.

[12] W. Lai *et al.*, “MEFNET: Multi-expert fusion network for RGB-Thermal semantic segmentation,” *Engineering Applications of Artificial Intelligence*, vol. 125, Oct. 2023, doi: 10.1016/J.ENGAPPAI.2023.106638.

[13] A. Tammana, M. P. Amogh, B. Gagan, M. Anuradha, and H. R. Vanamala, “Thermal Image Processing and Analysis for Surveillance UAVs,” *Lecture Notes in Networks and Systems*, vol. 190, pp. 577–585, 2021, doi: 10.1007/978-981-16-0882-7_50.

[14] Y. Diez, S. Kentsch, M. Fukuda, M. L. L. Caceres, K. Moritake, and M. Cabezas, “Deep learning in forestry using uav-acquired rgb data: A practical review,” *Remote Sensing*, vol. 13, no. 14, Jul. 2021, doi: 10.3390/RS13142837.

[15] Y. Zhao, J. Ma, X. Li, and J. Zhang, “Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery,” *Sensors*, vol. 18, no. 3, p. 712, Feb. 2018, doi: 10.3390/S18030712.

[16] D. Bulatov and F. Leidinger, “Instance segmentation of deadwood objects in combined optical and elevation data using convolutional neural networks,” *Proceedings, Earth Resources and Environmental Remote Sensing/GIS Applications XII*, vol. 11863, pp. 299–308, Sep. 2021, doi: 10.1117/12.2599837.

[17] D. Q. Tran, M. Park, D. Jung, and S. Park, “Damage-Map Estimation Using UAV Images and Deep Learning Algorithms for Disaster Management System,” *Remote Sensing*, vol. 12, no. 24, p. 4169, Dec. 2020, doi: 10.3390/RS12244169.

[18] S. Muksimova, S. Mardieva, and Y. I. Cho, “Deep Encoder–Decoder Network-Based Wildfire Segmentation Using Drone Images in Real-Time,” *Remote Sensing*, vol. 14, no. 24, 2022, doi: 10.3390/rs14246302.

[19] T. Garcia, R. Ribeiro, A. Bernardino, T. Garcia, R. Ribeiro, and A. Bernardino, “Wildfire aerial thermal image segmentation using unsupervised methods: a multilayer level set approach,” *International Journal of Wildland Fire*, vol. 32, no. 3, pp. 435–447, Mar. 2023, doi: 10.1071/WF22136.

[20] D. Qi, W. Tan, Q. Yao, and J. Liu, “YOLO5Face: Why Reinventing a Face Detector,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13805 LNCS, pp. 228–244, 2023, doi: 10.1007/978-3-031-25072-9_15.

[21] C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.

[22] G. Jocher, A. Chaurasia, and J. Qiu, “YOLOv8,” 2023, available online: https://github.com/ultralytics/ultralytics, accessed Mar. 30, 2023.

[23] L. Cai, T. Long, Y. Dai, and Y. Huang, “Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis,” *IEEE Access*, vol. 8, pp. 44400–44409, 2020, doi: 10.1109/ACCESS.2020.2976432.

[24] M. Yandouzi *et al.*, “Investigation of Combining Deep Learning Object Recognition with Drones for Forest Fire Detection and Monitoring,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, p. 2023, doi: 10.14569/IJACSA.2023.0140342.

[25] “Darwin,” available online: https://darwin.v7labs.com/datasets, accessed Sep. 05, 2023.

[26] A. Islam, S. R. S. Raisa, and N. H. Khan, “Enhanced Leafy Vegetable Analysis: Image Classification and Disease Instance Segmentation Using Deep Learning Techniques,” *SSRN (preprint)*, doi: 10.2139/SSRN.4470131.

[27] P. Bharati and A. Pramanik, “Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey,” *Advances in Intelligent Systems and Computing*, vol. 999, pp. 657–668, 2020, doi: 10.1007/978-981-13-9042-5_56.

[28] L. Cao, X. Zheng, and L. Fang, “The Semantic Segmentation of Standing Tree Images Based on the Yolo V7 Deep Learning Algorithm,” *Electronics*, vol. 12, no. 4, p. 929, Feb. 2023, doi: 10.3390/ELECTRONICS12040929.

[29] “API Documentation | TensorFlow v2.13.0.,” available online: https://www.tensorflow.org/api_docs, accessed Sep. 05, 2023.

[30] M. Berrahal *et al.*, “Investigating the effectiveness of deep learning approaches for deep fake detection,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3853–3860, Dec. 2023, doi: 10.11591/EEI.V12I6.6221

[31] L. Zhao, L. Zhi, C. Zhao, and W. Zheng, “Fire-YOLO: A Small Target Object Detection Method for Fire Inspection,” *Sustainability*, vol. 14, no. 9, p. 4930, Apr. 2022, doi: 10.3390/SU14094930.

[32] “Average Precision | Hasty.ai., ” available online: https://hasty.ai/docs/mp-wiki/metrics/average-precision, accessed: Jan. 31, 2024.

BIOGRAPHIES OF AUTHORS

Mimoun Yandouzi Ph.D. in Computer Science at Mohammed First University in Oujda, Morocco, where he is conducting research on the use of computer vision and deep learning techniques for the analysis of drone data, particularly in the case of forest fire detection. He holds a degree in Computer Engineering from the School of Mineral Industry in Rabat, Morocco (2001). Furthermore, he holds several certifications in artificial intelligence, computer vision, cloud computing, big data, and data mining. He also acted as a reviewer for several international conferences. He is currently employed at Mohammed First University as a professor at the ENSA Engineering School. He can be contacted at email: m.yandouzi@ump.ac.ma.

Mohammed Berrahal is a Professor at the Polydisciplinary Faculty of Safi (PFS), Cadi Ayyad University, in Safi, Morocco. His research focuses on the applications of deep learning in security and law enforcement. He obtained his M.Sc. in internet of things from the National School of Computer Science and Systems Analysis (ENSIAS), Mohammed V University in Rabat, Morocco, in 2018, and his B.Sc. in Computer Engineering from the School of Technology (ESTO), Mohammed I University in Oujda, Morocco, in 2016. He is also recognized for his expertise in artificial intelligence, 3D modeling, and programming, holding certifications in these areas. His contributions extend to serving as a reviewer for various international conferences and journals. He can be contacted at email: m.berrahal@ump.ac.ma.

Mounir Grari has a Ph.D. in Computer Engineering at Mohammed First University in Oujda, Morocco, where he is conducting research on the use of the internet of things and machine learning in the detection and monitoring of forest fires. He holds an Engineering degree in Computer Science from EMI, University Mohammed 5 in Rabat, Morocco (2002). Furthermore, he is certified in artificial intelligence, 3D modeling, and programming. Additionally, he has served as a reviewer for a number of international conferences and journals. And is currently employed at Mohammed First University as Secretary General of the College of Technology. He can be contacted at email: m.grari@ump.ac.ma.

Mohammed Boukabous is a Ph.D. candidate in Computer Engineering at Mohammed First University in Oujda, Morocco, where he is conducting research in security intelligence using deep learning algorithms in exchanged messages. He holds an M.Sc. degree in internet of things from Sidi Mohamed Ben Abdellah University in Fez, Morocco (2019), as well as a B.Sc. degree in Computer Engineering from Mohammed First University (2016). Furthermore, he holds several certifications in natural language processing, artificial intelligence, security intelligence, big data, and cybersecurity. Additionally, he served as a reviewer for various international conferences. He is currently employed at Mohammed First University as an administrative. He can be contacted at email: m.boukabous@ump.ac.ma.

Omar Moussaoui is an Associate Professor at the Higher School of Technology (ESTO) of Mohammed First University, Oujda – Morocco. He has been a member of the Computer Science Department of ESTO since 2013. He is currently director of the MATSI research laboratory. Omar completed his Ph.D. in computer science at the University of Cergy-Pontoise France in 2006. His research interests lie in the fields of IoT, wireless networks and security. He has actively collaborated with researchers in several other computer science disciplines. He participated in several scientific & organizing committees of national and international conferences. He served as reviewer for numerous international journals. He is an instructor for CISCO Networking Academy on CCNA Routing & Switching and CCNA Security. He can be contacted at email: o.moussaoui@ump.ac.ma.

Mostafa Azizi received a State Engineer degree in Automation and Industrial Computing from the Engineering School EMI of Rabat, Morocco in 1993, then a Master degree in Automation and Industrial Computing from the Faculty of Sciences of Oujda, Morocco in 1995, and a Ph.D. degree in Computer Science from the University of Montreal, Canada in 2001. He earned also tens of online certifications in Programming, Networking, AI, Computer Security. He is currently a Professor at the ESTO, University Mohammed First of Oujda. His research interests include Security and Networking, AI, Software Engineering, IoT, and Embedded Systems. His research findings with his team are published in over 100 peer-reviewed communications and papers. He also served as PC member and reviewer in several international conferences and journals. He can be contacted at email: azizi.mos@ump.ac.ma.

Kamal Ghoumid earned his Ph.D. from the Institut FEMTO-ST of Franche-Comté University (Besançon, France) and Télécom SudParis of Institut Polytechnique de Paris (Evry, France) in 2008, and subsequently obtained the "Habilitation à Diriger des Recherches (HDR)" diploma from Sorbonne University (Pierre-et-Marie-Curie Univ., France). He currently holds the position of Professor at the National School of Applied Sciences (ENSAO), Mohammed Premier University (Oujda, Morocco). Initially focusing on integrated components for optical telecommunications systems and Radio-over-Fiber applications, his research trajectory has evolved to encompass digital communications, the Internet of Things, and wireless communications, complemented by extensive experience in antennas and propagation research. He can be contacted at email: k.ghoumid@ump.ac.ma.

Aissa Kerkour Elmiad has been serving as a distinguished professor of Computer Science at the University of Computer Science, Mohamed I University – Oujda since 2018. He earned his doctorate in Computer Science from UMP1-Oujda, solidifying his expertise in the field. He participated in several scientific & organizing committees of national and international conferences. His research interests revolve around leveraging artificial intelligence in various domains including optimization, data mining, high performance computing, logistics, and healthcare. He can be contacted at email: mid.kerkour@gmail.com.