YOLOv5 *vs.* YOLOv8: Performance Benchmarking in Wildfire and Smoke Detection Scenarios

Edmundo Casas [1,2], Leo Ramos [1,3,4,*], Eduardo Bendek [1,5], and Francklin Rivas-Echeverria [1,6]

[1] Kauel Inc., Houston, USA
[2] Faculty of Engineering, University of Deusto, Bilbao, Spain
[3] Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain
[4] School of Mathematical and Computational Sciences, Yachay Tech University, Urcuquí, Ecuador
[5] Jet Propulsion Laboratory, NASA, Pasadena, USA
[6] Pontificia Universidad Católica del Ecuador Sede Ibarra, Ibarra, Ecuador
Emails: edmundo.casas@kauel.com (E.C.); leo.ramos@kauel.com (L.R.); eduardo.bendek@kauel.com (E.B.); francklin.rivas@kauel.com (F.R.-E.)
*Corresponding author

***Abstract*—This paper provides a thorough analysis and comparison of the YOLOv5 and YOLOv8 models for wildfire and smoke detection, using the Foggia dataset for evaluation. The study examines the small (s), medium (m), and large (l) variants of each architecture and employs various metrics, including recall, precision, F1-Score, and mAP@50, to assess performance. Additional considerations such as training and inference times, along with the number of epochs required for optimal recall, are also evaluated to gauge the models' real-world efficiency and effectiveness. Quantitatively, YOLOv5 models generally outperform YOLOv8, with the YOLOv5s variant achieving the highest scores across all metrics. However, visual assessments reveal that YOLOv8 models exhibit similar, and in some cases superior, capabilities, particularly in detecting dark and dense smoke. Training times favor YOLOv5 models, contributing to their efficiency, and their shorter inference times offer advantages for real-time applications. While the "best model" variants confirm YOLOv5's numerical dominance, YOLOv8's "best models" also display competitive performance. Future research will explore model evaluation on diverse datasets and hyperparameter optimization to further enhance performance, adaptability, and applicability in various real-world object detection scenarios.**

***Keywords*—wildfire detection, smoke detection, computer vision, deep learning, artificial intelligence, YOLO**

I. Introduction

Wildfires are an increasing global threat with devastating consequences, including loss of life, property damage, and environmental impact [1, 2]. Annually, they burn millions of acres, releasing significant amounts of greenhouse gases that worsen air quality and climate change [3]. Therefore, timely and accurate detection is crucial for mitigating their effects and preventing further spread [4].

Historically, traditional methods like manned watchtowers, satellites, and aircraft have been mainstays in wildfire detection, despite their costs and limitations [3]. Recent technological advances, particularly neural networks, offer more accurate and real-time detection, addressing some of these limitations [2, 4, 5].

The YOLO (You Only Look Once) architecture stands out among Artificial Intelligence (AI) techniques for its role in object detection [5]. As a convolutional neural network, YOLO has redefined the field of computer vision by detecting objects in both images and videos with unparalleled speed and accuracy [6]. Its unique capabilities have not only advanced the field but also had a wide-ranging impact across various industries.

In the YOLO series, YOLOv5 has distinguished itself as a potent and efficient model, effective in fields like security, medical imaging, and autonomous vehicles [5, 6]. Its real-time object processing and precision have made it a cornerstone in computer vision, driving ongoing innovation [7]. The series recently welcomed YOLOv8, a model surpassing its predecessors in performance [8]. It shows great promise for diverse applications and specifically, our research aims to deploy it for wildfire detection, contrasting its capabilities with YOLOv5 [9].

In Ref. [10], comparisons have been conducted that include other versions of YOLO. However, our study intentionally narrows this focus to a comparison between YOLOv5 and YOLOv8, with the latter being the most recently released version by Ultralytics. While YOLOv6 and YOLOv7 do exist, they were not developed by the Ultralights' team and thus, were not included in our study.

This decision was based on the cutting-edge advancements of YOLOv8 and the proven precision and

Manuscript received October 20, 2023; revised November 29, 2023; accepted December 6, 2023; published April 10, 2024.

efficiency of YOLOv5 in the computer vision industry. By concentrating our analysis on these two specific architectures, developed by the original creators at Ultralytics, we aim to gain a deeper understanding of their respective capabilities and contributions. This focused approach is expected to yield valuable insights particularly relevant to wildfire detection and the broader field of convolutional neural network research.

To conduct our study, we use the Foggia dataset [11], featuring 8,974 images tailored for wildfire detection. We carry out the comparison using three different variants (small, medium and, large) of both YOLOv5 and YOLOv8, training them for a fixed number of epochs and assessing their “best models”. Performance metrics include recall, precision, f1, and mAP.

Our study aims to evaluate the pros and cons of YOLOv5 and YOLOv8 in wildfire detection. Through this research, we seek to provide a comprehensive assessment of these existing detection systems, highlighting their strengths and limitations in the context of environmental monitoring. The insights gained from this study are intended to inform future research in the field of computer vision as applied to environmental challenges. By thoroughly analyzing these architectures, we aim to contribute valuable knowledge to the field, aiding in the refinement and application of these technologies for enhanced environmental protection and disaster mitigation efforts.

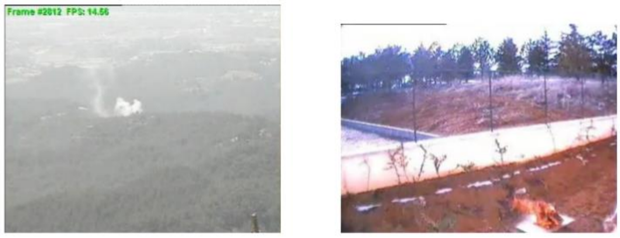
II. Materials and Methods

A. Dataset

The work presented in this paper leverages the Foggia dataset [11], a robust set of purpose-built images for smoke and wildfire detection. It was originally created in 2015 by a team of researchers with the aim of providing a representative set of visual data for the development of fire and smoke detection systems. The dataset initially comprised 31 videos, all sourced from the web, featuring real-life fire scenarios. These videos offer a range of resolutions from 320×240 to 800×600 pixels.

The dataset is organized into two distinct sections. The first part, encompassing the initial 14 videos, predominantly features scenes with fire, while the latter part, comprising the remaining 17 videos, focuses on non-fire scenarios. This segmentation is essential for training detection systems to discern between actual fire scenes and those that might be mistakenly identified as such. The Foggia dataset has been instrumental in various studies [12–14], demonstrating its effectiveness in training fire and smoke detection systems and highlighting its significance in the field.

Over the years, the Foggia dataset has been the basis for various versions and adaptations, evolving to meet the changing requirements of wildfire detection research. The version we utilized in this study includes a total of 8,974 images extracted from these videos. These images encompass 3,731 instances of fire and 6,791 instances featuring smoke, thereby offering a diverse range of scenarios for thorough analysis. Fig. 1 illustrates this variety.



(a) Smoke (b) Fire

Fig. 1. Sample images from the Foggia dataset.

This dataset has been carefully curated and was last updated in February 2023, ensuring its relevance and applicability to contemporary wildfire detection challenges. It is accessible through Roboflow, a platform revered for its provision of high-quality, machine learning compatible datasets.

For our study, the dataset was partitioned into 70% (6,300 images) for training, 20% (1,800 images) for validation, and 10% (899 images) for testing. This distribution ensures a comprehensive model training and evaluation process, aiding in fine-tuning performance and providing a robust benchmark for assessing the models’ effectiveness in real-world scenarios.

B. Architectures under Study

1) YOLOv5

YOLOv5, standing for You Only Look Once version 5, is a state-of-the-art object detection algorithm noted for its simplicity, accuracy, and reliability [5]. Released by Ultralytics on June 25th, 2020 [15], the model is composed of four core parts: input, backbone, neck, and head, as depicted in Fig. 2.

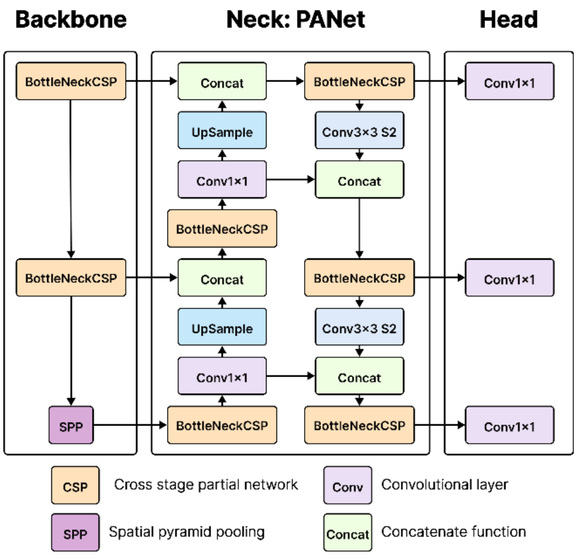


Fig. 2. YOLOv5 architecture [16].

The backbone relies on the CSP-Darknet53 convolutional network and uses the Cross Stage Partial (CSP) strategy to facilitate information flow while mitigating redundant gradients and vanishing gradient issues [17]. Features are extracted and mapped from the input image via this backbone.

The neck of YOLOv5 incorporates a variant of Spatial Pyramid Pooling (SPP) and integrates BottleNeckCSP into the Path Aggregation Network (PANet) [18]. This setup enhances the receptive field and focuses on crucial contextual features. PANet is further optimized using the CSPNet strategy for improved pixel localization.

The head section, following the trend of earlier versions, consists of three convolutional layers that predict bounding boxes, scores, and object classes [17]. Despite some ongoing challenges such as feature map redundancy and target misses in certain scenarios [19], YOLOv5 remains a leading solution in the object detection landscape.

This model is available in various versions, ranging from lighter models to more robust ones. In this study, we will analyze three of these versions, as detailed in Table I. Each variant differs in terms of depth, width, number of parameters, and layers, offering a spectrum of choices from faster inference speeds to greater object detection accuracy [20, 21].

TABLE I. DETAILS OF THE STUDIED YOLOV5 VARIANTS

Model	Layers	Size (pixels)	Params (M)	FLOPs (B)
YOLOv5s	214	640	7.2	16.5
YOLOv5m	291	640	21.2	49
YOLOv5l	368	640	46.5	109.1

2) YOLOv8

YOLOv8, the newest version of the You Only Look Once series, was launched by Ultralytics on January 10th, 2023 [15]. Building on the foundational YOLOv5, YOLOv8 introduces several key improvements, as depicted in Fig. 3, that make it an attractive option for future computer vision applications.

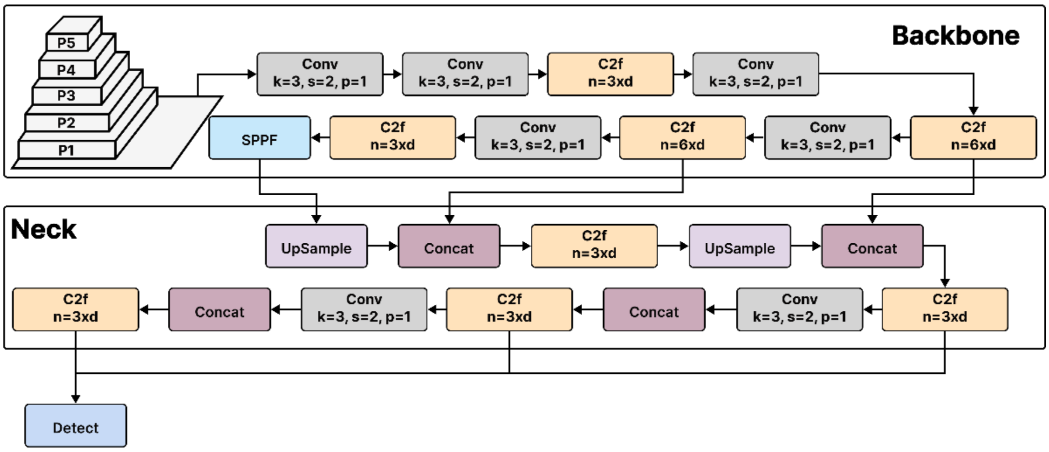


Fig. 3. YOLOv8 architecture [16].

One pivotal update in YOLOv8 is the shift to anchor-free detection, veering away from the anchor-box techniques of its predecessors [9]. This modification enables direct prediction of object centers, streamlining the Non-Maximum Suppression (NMS) process and alleviating issues related to anchor boxes, such as lack of generalization and difficulty in handling irregular shapes.

In terms of convolutional design, YOLOv8 replaces the initial 6×6 convolution in the stem with a more efficient 3×3 and updates the core building block by swapping C3 for C2f [22]. Features are also concatenated directly in the model's neck without enforcing identical channel dimensions, which cuts down on the parameter count and tensor size. An innovative feature named Spatial Pyramid Pooling Feature (SPPF) is introduced, enhancing the model's ability to manage objects of diverse scales [22, 23].

YOLOv8 also incorporates online image augmentation during training, such as mosaic augmentation. This feature allows the model to learn objects in new positions and under different conditions, including partial occlusions and varying backgrounds [24].

While still lacking an official research paper, YOLOv8's feature-rich design and vibrant community support underscore its cutting-edge nature and potential in the field of computer vision.

Similar to what we observed with the YOLOv5 model, YOLOv8 also comes in different versions, each tailored to balance between detection accuracy and computational efficiency. Our study focuses on analyzing three of these versions, as specified in Table II. These variants of YOLOv8, like those of YOLOv5, vary in terms of depth, width, number of parameters, and layers [25, 26], allowing for a range of applications from resource-limited environments to scenarios requiring high precision in object detection.

TABLE II. DETAILS OF THE STUDIED YOLOV8 VARIANTS

Model	Layers	Size (pixels)	Params (M)	FLOPs (B)
YOLOv8s	225	640	11.2	28.6
YOLOv8m	295	640	25.9	78.9
YOLOv8l	365	640	43.7	165.2

C. Training and Evaluation

In our study, we implement a systematic methodology for training, evaluating, and comparing the deep learning architectures. Each model is trained for a predefined 200 epochs, a duration chosen based on recommendations from the official repositories of the architectures, which suggest testing with epoch numbers close to 200 and 300.

The performance of each model is monitored using a separate validation dataset. The ‘best model’ is selected based on achieving the highest recall on the validation set, emphasizing the importance of minimizing false negatives in wildfire detection. Following the training phase, these chosen ‘best models’ are evaluated on a dedicated test set to assess their generalization capabilities.

1) Performance metrics

a) Precision

Precision is a fundamental metric in object detection. It measures the accuracy of the model’s positive predictions [16, 27, 28]. Mathematically, it’s defined as shown in Eq. (1), where TP is the count of correctly predicted positive instances, and FP is the number of false positives. High precision indicates fewer false positives.

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

b) Recall

Recall signifies the percentage of actual positive instances correctly identified by the model [29]. In the realm of object detection, it essentially gauges the model’s ability to accurately detect and capture object instances [27, 28]. A higher value of recall corresponds to a lower count of missed detections. Recall is mathematically defined as shown in Eq. (2), where TP stands for the number of correctly predicted positive instances, and FN refers to instances erroneously identified as negative.

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

c) F1-score

The F1-score combines precision and recall to offer an overview of the model’s accuracy [29]. It is a calculated as the harmonic mean of these two metrics. The F1-score, as per Eq. (3), represents the ratio of the product of precision and recall to their sum, multiplied by 2.

$$F1 = 2 \times (Precision \times Recall)/(Precision + Recall) \tag{3}$$

An elevated F1-Score suggests a superior balance between precision and recall, illustrating the model’s aptitude for accurate object detection while keeping false positives and false negatives to a minimum [27, 29].

d) Mean average precision

The Mean Average Precision (mAP) gauges the balance between precision and recall. It achieves this by computing the Average Precision (AP) per class and subsequently averaging across all classes [28]. AP calculates precision at varying recall levels, essentially computing the area under the precision-recall curve [30].

Mathematically, it can be formulated as shown in Eq (4), with precision defining the precision at a specific recall level ®. A higher mAP indicates superior object detection performance, factoring in both precision and recall. Higher mAP indicates superior object detection performance, factoring in both precision and recall [16, 27].

$$AP = a \int precision(r)dr \tag{4}$$

D. Implementation Details

We conducted the training on a High-Performance Computing (HPC) cluster using two Nvidia A100 SXM4 40GB GPUs, 32 CPU cores, and 64GB of system memory. The batch size used for training was set to 64, and we kept all the original hyperparameters for each model in order to establish a baseline for comparison.

The implementations of the YOLOv5 and YOLOv8 models were obtained from their respective official Ultralytics repositories, ensuring authenticity and consistency with the original designs. Python, in conjunction with the PyTorch framework, was employed as the programming language for this study, leveraging its robustness and flexibility in handling complex neural network architectures.

III. Result and Discussion

Table III provides the training times for each variant of the YOLOv5 and YOLOv8 models. Smaller models, YOLOv5s and YOLOv8s, have the shortest training duration, recorded at 1.073 and 1.269 h respectively. As expected, smaller models, with fewer parameters, generally demand less computational resources and time for training.

The larger YOLOv5l and YOLOv8l models require 1.951 and 2.275 h, respectively, for training, making them the longest to train. This is mainly because these models are more complex, demanding more computational resources. Notably, YOLOv8 models consistently have longer training times compared to YOLOv5 variants, possibly indicating greater complexity or additional features needing optimization.

TABLE III. Training Time for Each Model Addressed

Model	Time (h)
YOLOv5s	1.073
YOLOv5m	1.470
YOLOv5l	1.951
YOLOv8s	1.269
YOLOv8m	1.732
YOLOv8l	2.275

Table IV encapsulates the average performance metrics of each variant of the YOLOv5 model on validation. The standard deviation (SD) is also included to offer insights into the consistency of the performance of each variant.

TABLE IV. METRICS RESULTS FOR EACH YOLOV5 VARIANT ON VALIDATION

Model	Precision		Recall		F1-score		mAP	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
YOLOv5s	0.902	0.052	0.886	0.044	0.893	0.048	0.916	0.053
YOLOv5m	0.901	0.055	0.883	0.048	0.891	0.051	0.910	0.057
YOLOv5l	0.895	0.043	0.883	0.038	0.889	0.040	0.908	0.046

The YOLOv5s variant stands out for its exceptional performance across all metrics, especially in wildfire and smoke detection. It achieves a precision score of 0.902, demonstrating its ability to make highly accurate predictions while minimizing false positives, a crucial feature for our specific application.

The YOLOv5s model attains a high recall score of 0.886, highlighting its ability to identify true positive instances effectively. This capability is crucial for avoiding potential false negatives in wildfire detection scenarios. Additionally, YOLOv5s not only excels in precision and recall but also achieves top scores for F1-Score and mAP@50. These scores firmly establish the model's outstanding performance in object detection tasks.

The relatively small standard deviations for YOLOv5s across these metrics emphasize the model's consistent and stable performance. It's also worth mentioning that the performance variations among the YOLOv5 variants are relatively minor, indicating a consistent level of performance throughout the YOLOv5 architecture.

In addition to numerical results, we conducted graphical tests to assess the performance of the models in a context more closely resembling real-world conditions. Fig. 4 presents the results obtained using YOLOv5 models.

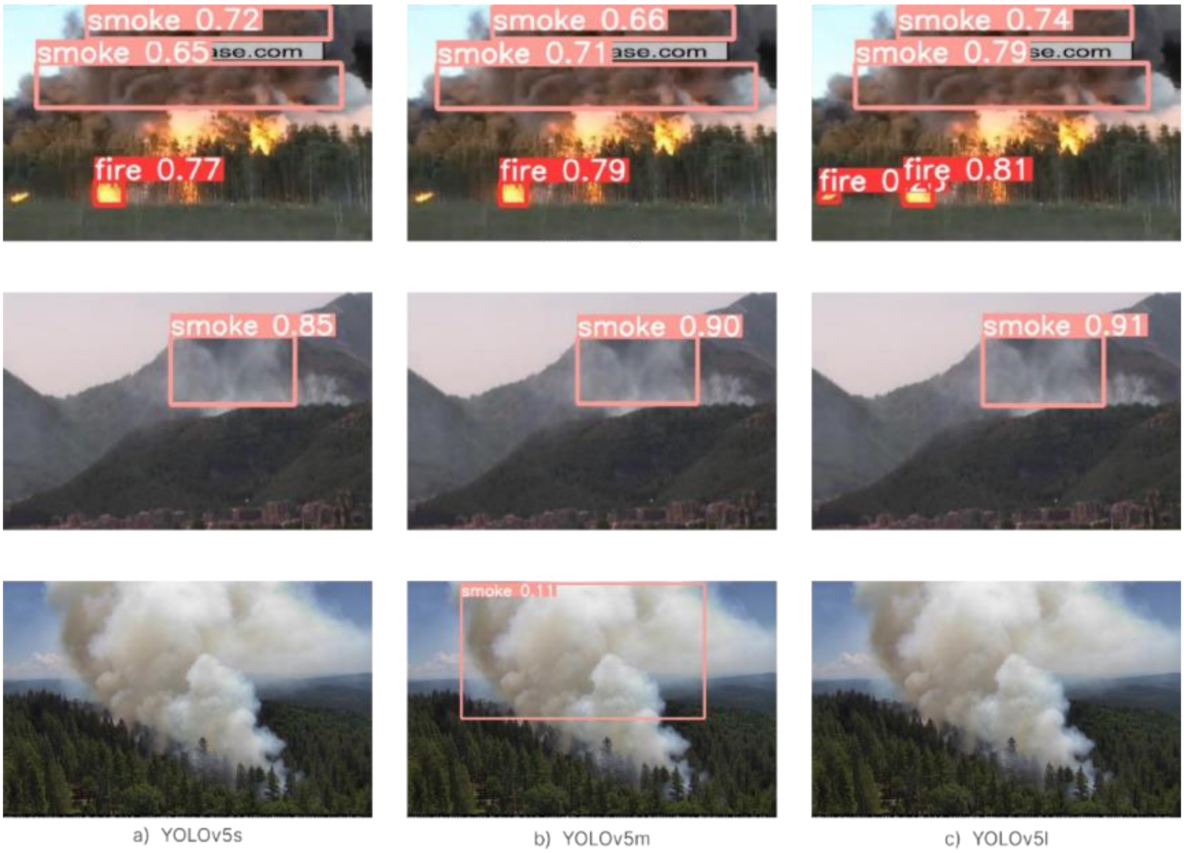


Fig. 4. Inference using YOLOv5 models used.

Generally speaking, all versions of the YOLOv5 model appear to be capable of detecting both smoke and fire in the provided images. However, subtle variations in their performance were observed. Moreover, this model yielded the highest confidence scores among all the evaluated models. The YOLOv5s and YOLOv5m versions showed a slight decrease in confidence scores, particularly when detecting instances of smoke. This could point to potential challenges these models face in differentiating between dense smoke and background elements, or in their sensitivity to certain patterns and lighting conditions.

In the middle row, a high degree of parity in performance across all models is observed. The bounding boxes are well-positioned and virtually in the same location. Additionally, the confidence scores for all models exceed 85%, indicating strong performance in detecting columns of less dense smoke under favorable lighting conditions.

In the bottom row, another scene featuring a large column of smoke is presented. Interestingly, unlike in previous scenarios, only the YOLOv5m model was able to detect the smoke in this scene, albeit with a very low confidence score. The other two models, YOLOv5s and YOLOv8l, failed to detect anything in the image despite the presence of a very clear and easily identifiable column of smoke. This is noteworthy because, in prior scenes, the models correctly identified the smoke, demonstrating that the models have limitations in detecting columns of smoke with specific characteristics, as displayed in this scene.

Examining the performance of YOLOv8 model variants on validation, as shown in Table V, YOLOv8s stands out in terms of precision, averaging a score of 0.886. The relatively low standard deviation of 0.048 indicates that this model not only delivers high prediction accuracy but also maintains consistency across various runs. This consistent precision is crucial in wildfire and smoke detection tasks, as it helps minimize false alarms, preserving resources and reducing the burden on emergency services.

TABLE V. METRICS RESULTS FOR EACH YOLOV8 VARIANT ON VALIDATION

Model	Precision		Recall		F1-score		mAP	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
YOLOv8s	0.886	0.048	0.860	0.059	0.873	0.054	0.906	0.059
YOLOv8m	0.882	0.040	0.866	0.050	0.874	0.044	0.905	0.051
YOLOv8l	0.878	0.057	0.863	0.073	0.867	0.065	0.898	0.075

Looking at recall, an essential metric in scenarios like wildfire detection where missed detections can lead to dire consequences, the YOLOv8m variant yields the highest average score of 0.866. However, the standard deviation for this metric is slightly higher, indicating a touch more variability in the model’s performance than with precision.

Turning to the F1-Score, the YOLOv8s and YOLOv8m models display similar performance, both slightly surpassing the YOLOv8l variant. Regarding the mAP@50 metric, which gauges the overall detection performance across various thresholds, the YOLOv8s variant again takes the lead.

Thus, YOLOv8s shines with the highest precision and mAP@50, while YOLOv8m stands out with the highest recall. The relatively minor differences in performance among the variants point to an inherent stability within the YOLOv8 architecture.

It is evident that the YOLOv5 models, specifically the small variant, are exhibiting superior performance across all metrics. Notably, YOLOv5s with its higher precision, recall, F1-Score, and mAP@50, implies an enhanced ability to accurately detect and locate wildfires and smoke in our dataset.

The difference in average values for these key metrics between YOLOv5s and YOLOv8s, although marginal, is noteworthy. This highlights YOLOv5s’ reliability in this specific detection task. Moreover, YOLOv5s displays relatively low standard deviations, suggesting a higher level of consistency across different runs.

Taking time efficiency into account alongside performance, YOLOv5 models hold an advantage due to their shorter training duration. Nevertheless, both YOLOv5 and YOLOv8 models demonstrate commendable performance in wildfire and smoke detection, with YOLOv5s having a slight edge. It’s essential to recognize that these findings are based on a single dataset, and further experiments with diverse datasets could yield additional insights.

Table VI presents the best recall values and their corresponding epochs for each model variant of YOLOv5 and YOLOv8. The highest recall among YOLOv5 models was attained by the small variant (YOLOv5s) at 0.907 during the 107th epoch, showcasing its strong ability to identify true positives. In the case of YOLOv8, the medium variant (YOLOv8m) achieved the highest recall at 0.897 in the 187th epoch.

Interestingly, the epoch of achieving the best recall varies among the models, possibly reflecting differences in model complexity and learning rate evolution. YOLOv5s reached peak performance earlier than YOLOv5m, with YOLOv5l doing so even sooner, by the 89th epoch. In contrast, both medium and large variants of YOLOv8 required more epochs to attain their best recall compared to the small variant.

Table VII shows the inference time in milliseconds for each ‘best model’ in both YOLOv5 and YOLOv8. Clearly, YOLOv5 models consistently exhibit faster inference times compared to their YOLOv8 counterparts. Notably, YOLOv5s achieves the fastest inference time at 0.9 ms, highlighting its efficiency in delivering high performance with minimal computational load.

Conversely, YOLOv8 models exhibited slower inference times across all variants. However, it’s worth noting that these models still maintained a competitive time range, with the fastest being YOLOv8s at 1.3 ms. Despite the increased complexity of YOLOv8, its inference time performance remains commendable, although not as efficient as YOLOv5.

TABLE VI. BEST RECALL AND CORRESPONDING EPOCH FOR EACH MODEL ON VALIDATION

Model	Best recall	Epoch
YOLOv5s	0.907	107
YOLOv5m	0.906	153
YOLOv5l	0.904	89
YOLOv8s	0.895	93
YOLOv8m	0.897	187
YOLOv8l	0.895	182

TABLE VII. INFERENCE TIME FOR EACH "BEST MODEL"

Model	Time (ms)
YOLOv5s	0.9
YOLOv5m	1.4
YOLOv5l	1.8
YOLOv8s	1.3
YOLOv8m	2.1
YOLOv8l	3.0

Table VIII depicts the testing results of the 'best model' for the YOLOv5 variants on the two distinct classes (fire and smoke), as well as the consolidated class (all).

TABLE VIII. RESULTS OF EACH YOLOV5 'BEST MODEL' ON TESTING

Model	Class	Precision	Recall	F1	mAP
YOLOv5s	All	0.899	0.888	0.894	0.907
	Fire	0.837	0.844	0.841	0.854
	Smoke	0.962	0.932	0.947	0.960
YOLOv5m	All	0.903	0.888	0.895	0.905
	Fire	0.852	0.847	0.850	0.864
	Smoke	0.953	0.929	0.941	0.945
YOLOv5l	All	0.888	0.880	0.884	0.905
	Fire	0.828	0.835	0.831	0.861
	Smoke	0.949	0.924	0.936	0.949

In the "All" category, representing combined performance for "Fire" and "Smoke" classes, all three models show relatively high performance. Among them, YOLOv5m achieves the highest precision and F1-score at 0.903 and 0.895, respectively, closely followed by the YOLOv5s variant. Interestingly, the mAP@50 score, which combines both recall and precision, is the same (0.905) for both YOLOv5m and YOLOv5l models, despite YOLOv5l having slightly lower precision. This suggests that the YOLOv5l model's higher recall compensates for its slightly lower precision when considering overall performance.

Focusing on the 'Smoke' class, the YOLOv5s model is evidently superior with a precision of 0.962, which is significantly higher than the other two models. Its recall rate is also the highest at 0.932. This superiority in both precision and recall is reflected in the higher F1-score (0.947) and mAP@50 (0.960). The YOLOv5s model, therefore, offers an optimal choice for high precision and recall in "Smoke" detection.

In the "Fire" class evaluation, YOLOv5m stands out with a recall of 0.847, slightly surpassing YOLOv5s (0.844) and YOLOv5l (0.835). Its higher precision score of 0.852 indicates YOLOv5m's tendency to return more relevant results. This balanced performance is confirmed by its leading F1-score of 0.850.

The mAP@50 score highlights YOLOv5m's superior performance in object detection and bounding box accuracy. With a score of 0.864, YOLOv5m is more accurate and reliable at identifying and locating instances of fire within an image compared to YOLOv5s (0.854) and YOLOv5l (0.861). In real-world applications like wildfire detection, this advantage can potentially lead to fewer false alarms and missed detections, demonstrating the reliability and practical utility of YOLOv5m.

Regarding YOLOv8, the performance of its variants on different classes presents an interesting perspective, as shown in Table IX.

TABLE IX. RESULTS OF EACH YOLOV8 "BEST MODEL" ON TESTING

Model	Class	Precision	Recall	F1	mAP
YOLOv8s	All	0.896	0.866	0.881	0.912
	Fire	0.842	0.792	0.816	0.862
	Smoke	0.950	0.939	0.944	0.962
YOLOv8m	All	0.880	0.864	0.872	0.897
	Fire	0.830	0.807	0.818	0.844
	Smoke	0.930	0.921	0.926	0.950
YOLOv8l	All	0.886	0.876	0.881	0.899
	Fire	0.835	0.824	0.830	0.850
	Smoke	0.936	0.928	0.932	0.949

When considering the "All" class, YOLOv8s leads in terms of precision (0.896) and mAP@50 (0.912), highlighting its superior object detection accuracy across classes and its stronger overall performance in accurate bounding box localization. Its superior mAP@50 score signifies that YOLOv8s is likely to be more reliable in real-world applications, producing fewer false alarms and better identifying the correct object locations. However, its recall score (0.866) lags slightly behind that of YOLOv8l (0.876), suggesting YOLOv8l may be slightly more effective in identifying all relevant instances across classes.

Analyzing the "Smoke" class, we see that YOLOv8s again takes the lead with the highest precision (0.950) and mAP@50 (0.962) scores. These metrics underline its ability to accurately detect and localize smoke instances. Its recall score of 0.939, though marginally better than the others, indicates a nearly equal proficiency in identifying all relevant smoke instances.

Regarding the "Fire" class, the "models" performance is more closely matched. YOLOv8l demonstrates a slight edge with the highest recall score of 0.824, indicating its superior capability to identify all relevant fire instances. However, the highest precision and mAP@50 scores are held by YOLOv8s (0.842 and 0.862, respectively). This indicates that YOLOv8s is more accurate in detecting and localizing fire instances, hence more reliable in practical applications. Despite these differences, the models exhibit closely matched F1-Scores, demonstrating a balanced performance in precision and recall.

Overall, the YOLOv5 variants demonstrate better results in detecting both instances (smoke and fire). These variants show improvements in precision, recall, and F1-score compared to YOLOv8. However, they exhibit slightly lower performance in terms of mAP@50.

Regarding the "Smoke" class, the results vary significantly. Some YOLOv5 variants outperform YOLOv8 in metrics such as precision and F1-score, but have lower recall. Additionally, the mAP@50 results are similar across models. For the "Fire" class, the YOLOv5 variants generally show better performance in terms of precision, recall, and F1-Score. In other words, they are better at detecting fire instances. The mAP@50 results are more variable, but with very small differences between models.

Moving on to the visual results displayed in Fig. 5 in the top row, good performance is observed across all models. The bounding boxes are well-positioned, and the confidence scores are high (>70%) for all main instances. However, similar to what was observed with the YOLOv5 models, the YOLOv8l model detected an additional small instance of fire compared to the other models. Furthermore, its confidence score is moderately high, demonstrating a slight superiority of this variant.

In the middle row, mirroring what was seen with the YOLOv5 models, a high degree of parity in performance among all models is observed. The bounding boxes are well-placed, although the YOLOv8m model shows slight variations and compaction. The confidence scores are high, indicating strong detection capabilities.

Transitioning to the bottom row, we once again observe difficulties in the models' ability to detect the smoke column in this particular scene, similar to what was observed with the YOLOv5 models. However, this time, two models, YOLOv8s and YOLOv8l, were successful in detecting the smoke column. Notably, the YOLOv8s model has a smaller and tighter bounding box, while YOLOv8l features a much broader bounding box, capturing a larger extent of the smoke column. Nevertheless, the confidence scores in both cases are relatively low, which does not provide adequate assurance in the detection capabilities and renders the models inefficient for potential real-world applications.

In general terms, the YOLOv8 models display consistent and even superior performance compared to YOLOv5 models. Although neither model met the expected performance in certain scenarios, YOLOv8 appears to have greater potential than YOLOv5. With improved training data or optimized training configurations, YOLOv8 is likely to further distinguish itself in terms of performance, which is anticipated given the enhancements introduced in this version.

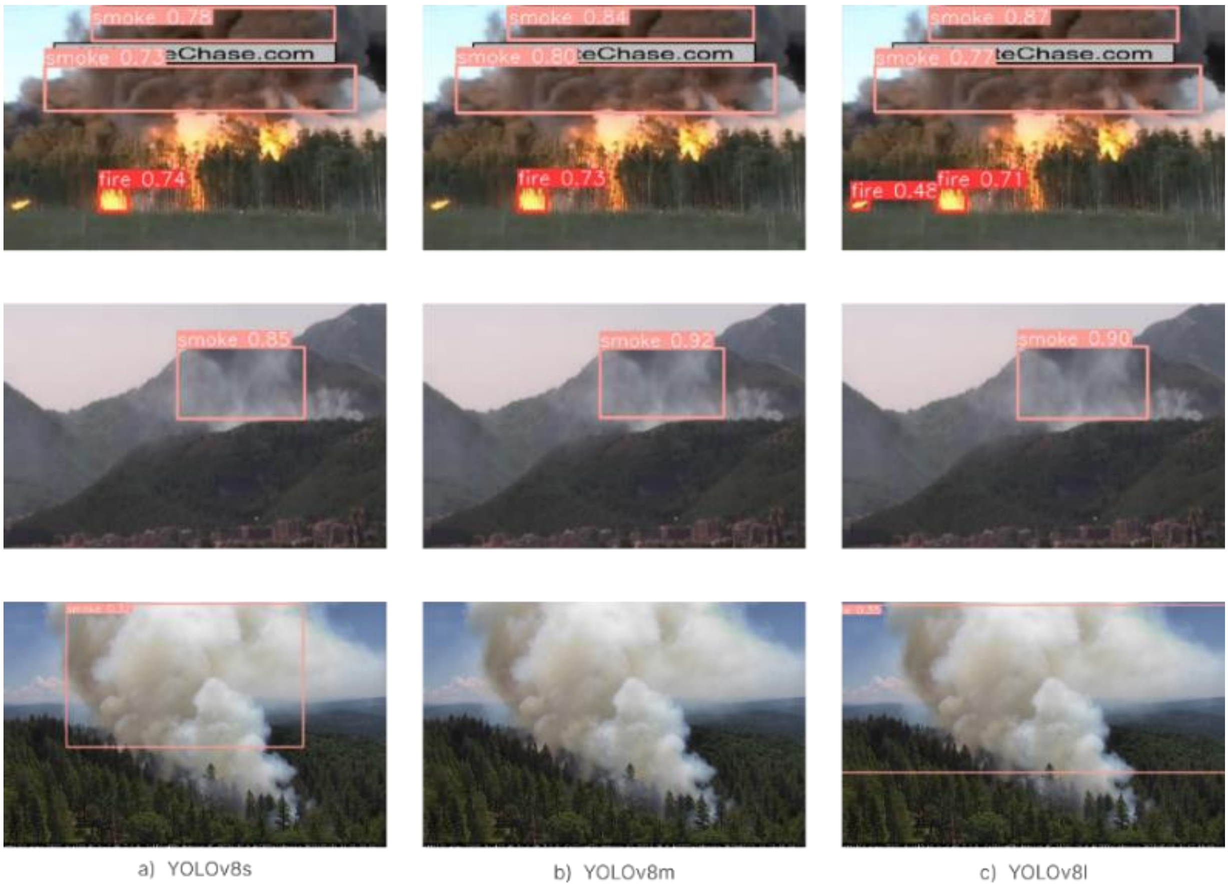


Fig. 5. Inference using YOLOv8 models studied.

IV. Conclusions and Future Works

In this study, we meticulously examined and compared the performance of YOLOv5 and YOLOv8, two notable iterations of the "You Only Look Once" (YOLO) object detection architecture. Our comprehensive analysis involved various critical factors, including metrics such as precision, recall, F1-Score, and mAP@50, as well as training times, inference times, and the number of epochs required to achieve optimal recall. To ensure robustness in our evaluation, we employed the Foggia dataset for our experiments.

The comparative study between YOLOv5 and YOLOv8 models has yielded insightful distinctions in performance. Across standard metrics, YOLOv5 consistently excels, achieving higher precision, recall, and F1-Scores, coupled with shorter training times and fewer epochs to optimal recall. These factors underscore YOLOv5's efficiency, making it particularly suitable for scenarios with limited resources or time constraints.

While YOLOv5 has the upper hand in quantitative metrics, YOLOv8 is commendably close, particularly in mAP@50 scores, suggesting that it remains a competitive alternative in the object detection landscape. In terms of real-time application, YOLOv5's faster inference times are advantageous; however, YOLOv8's respectable inference times cannot be disregarded, as they reflect its viability in scenarios where slightly longer processing can be accommodated.

Visual test results reveal a nuanced dynamic: YOLOv8 excels in detecting dark and dense smoke, a critical attribute for real-time detection in variable conditions. On the other hand, YOLOv5 shows a slight advantage in identifying fire more confidently. Despite this, both models encounter challenges with light-colored smoke, indicating an area for potential improvement.

In summary, while YOLOv5 outstrips YOLOv8 in efficiency and core metric performance, YOLOv8's robust visual detection capabilities, especially for complex smoke scenarios, illustrate its practical utility. The decision between the two should be informed by the specific requirements of the intended application whether it is the precision and speed of YOLOv5 or the visual acuity of YOLOv8 that is more critical. With distinct strengths, both YOLO models hold promise for varied deployment in object detection tasks, particularly for enhancing wildfire and smoke detection solutions.

Future studies could yield substantial benefits by expanding the scope of evaluation to datasets that encompass a wider variety of wildfire scenarios, which may further delineate the conditions under which each model excels. Such research is essential to refine detection accuracy, particularly in diverse and unpredictable environmental conditions. This will directly support efforts in environmental protection by enhancing early detection capabilities, ultimately contributing to more effective disaster mitigation strategies.

Additionally, a systematic exploration of hyperparameters and model architectures could uncover optimizations that enhance model performance. These potential future directions aim to further improve the performance, flexibility, and applicability of the YOLOv5 and YOLOv8 models, allowing for their effective deployment in various real-world scenarios and expanding their potential impact in the field of object detection. Emphasis on these areas of research could significantly advance the field of computer vision in environmental monitoring applications.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

EC: Formal analysis and investigation, Writing-review and editing. LR: Conceptualization, Formal analysis and investigation, Writing-original draft preparation, Writing-review and editing. EB: Formal analysis and investigation, Writing-review and editing. FRE: Formal analysis and investigation, Writing-review and editing. All authors reviewed the manuscript.

References

[1] D. Kinaneva, G. Hristov, J. Raychev, and P. Zahariev, "Early forest fire detection using drones and artificial intelligence," in *Proc. 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2019, pp. 1060–1065.

[2] S. Dutta and S. Ghosh, "Forest fire detection using combined architecture of separable convolution and image processing," in *Proc. 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, 2021, pp. 36–41.

[3] D. Kinaneva, G. Hristov, G. Georgiev, P. Kyuchukov, and P. Zahariev, "An artificial intelligence approach to real-time automatic smoke detection by unmanned aerial vehicles and forest observation systems," in *Proc. 2020 International Conference on Biomedical Innovations and Applications (BIA)*, 2020, pp. 133–138.

[4] E. Buza and A. Akagic, "Unsupervised method for wildfire flame segmentation and detection," *IEEE Access*, vol. 10, pp. 55213–55225, 2022.

[5] J. Johnston, K. Zeng, and N. Wu, "An evaluation and embedded hardware implementation of yolo for real-time wildfire detection," in *Proc. 2022 IEEE World AI IoT Congress (AIIoT)*, 2022, pp. 138–144.

[6] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[7] Y. Nuwara, W. K. Wong, and F. H. Juwono, "Modern computer vision for oil palm tree health surveillance using yolov5," in *Proc. 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, 2022, pp. 404–409.

[8] G. Jocher, A. Chaurasia, and J. Qiu. (Jan. 2023). YOLO by ultralytics. [Online]. Available: https://github.com/ultralytics/ultralytics

[9] A. Dumitriu, F. Tatui, F. Miron, R. T. Ionescu, and R. Timofte, "Rip current segmentation: A novel benchmark and yolov8 baseline results," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2023, pp. 1261–1271.

[10] A. Yazan, M. Alauthman, A. Al-Qerem, A. Aldweesh, R. Quaddoura, F. Aburub, K. Mansour, and T. Alhmiedat, "Early wildfire smoke detection using different YOLO models" *Machines*, vol. 11, no. 2, 2023. https://doi.org/10.3390/machines11020246

[11] F. Detection. (Feb. 2023). Foggia-all dataset. [Online]. Available: https://universe.roboflow.com/fire-detection-uoeha/foggia-all

[12] M. Shahid and K. Hua, "Fire detection using transformer network." in *Proc. the 2021 International Conference on Multimedia Retrieval (ICMR)*, New York, NY, USA, 2021, pp. 627–630. https://doi.org/10.1145/3460426.3463665

[13] K. Muhammad, J. Ahmad, and S. W. Baik, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," *Neurocomputing*, vol. 2, pp. 30–42, 2017. doi: 10.1016/j.neucom.

[14] F. Saeed, A. Paul, and P. Karthigaikumar, "Convolutional neural network based early fire detection." *Multimed Tools Appl*, vol. 79, pp. 9083–9099, 2020. https://doi.org/10.1007/s11042-019-07785-w

[15] G. Jocher. (May 2020). YOLOv5 by ultralytics. [Online]. Available: https://github.com/ultralytics/yolov5

[16] E. Casas, L. Ramos, E. Bendek, and F. Rivas-Echeverria, "Assessing the effectiveness of yolo architectures for smoke and wildfire detection," *IEEE Access*, vol. 11, pp. 96554–96583, 2023.

[17] M. L. Mekhalfi, C. Nicolo, Y. Bazi, M. M. A. Rahhal, N. A. Alsharif, and E. A. Maghayreh, "Contrasting yolov5, transformer, and efficientdet detectors for crop circle detection in desert," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[18] J. K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in

Proc. the IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.
[19] L. H. Den, J. Ito, and A. Kokaze, "Diagnostic accuracy of a deep learning model using yolov5 for detecting developmental dysplasia of the hip on radiography images," *Scientific Reports*, vol. 13, no. 1, 2023.
[20] P. Wu, H. Weng, W. Luo, Y. Zhan, L. Xiong, and H. Zhang, "An improved Yolov5s based on transformer backbone network for detection and classification of bronchoalveolar lavage cells," *Computational and Structural Biotechnology Journal*, vol. 6, no. 21, pp. 2985–3001, 2023. doi:10.1016/j.csbj.2023.05.008
[21] P. K. Yadav, J. A. Thomasson, S. W. Searcy, R. G. Hardin, U. Braga-Neto, S. C. Popescu, D. E. Martin, R. Rodriguez, K. Meza, J. Enciso, J. S. Diaz, and T. Wang, "Assessing the performance of yolov5 algorithm for detecting volunteer cotton plants in corn fields at three different growth stages," *Artificial Intelligence in Agriculture*, vol. 6, pp. 292–303, 2022.
[22] H. Lou, X. Duan, J. Guo, H. Liu, J. Gu, L. Bi, and H. Chen, "Dcyolov8: Small-size object detection algorithm based on camera sensor," *Electronics*, vol. 12, no. 10, 2023.
[23] A. Vats and D. C. Anastasiu, "Enhancing retail checkout through video inpainting, yolov8 detection, and deepsort tracking," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2023, pp. 5529–5536.
[24] D. D. Van, "Application of advanced deep convolutional neural networks for the recognition of road surface anomalies," *Engineering, Technology &; Applied Science Research*, vol. 13, no. 3, pp. 10765–10768, Jun. 2023.
[25] M. Chetoui and M. A. Akhloufi "Object detection model-based quality inspection using a deep CNN," in *Proc. Sixteenth International Conference on Quality Control by Artificial Vision*, July 2023. https://doi.org/10.1117/12.2689921
[26] N. Aishwarya and R V. Kumar, "Banana ripeness classification with deep CNN on NVIDIA jetson Xavier AGX," in *Proc. 2023 7th International Conference on I-SMAC*, Kirtipur, Nepal, 2023, pp. 663–668. doi: 10.1109/I-SMAC58438.2023.10290326
[27] M. Vergara, L. Ramos, N. D. Rivera-Campoverde and F. Rivas-Echeverría, "EngineFaultDB: A novel dataset for automotive engine fault classification and baseline results," *IEEE Access*, vol. 11, pp. 126155-126171, 2023. doi: 10.1109/ACCESS.2023.3331316
[28] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.
[29] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics*, vol. 10, no. 3, 2021.
[30] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics and methods," *Applied Sciences*, vol. 10, no. 21, 2020.