ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

Al-Noor Journal for Information Technology and Cybersecurity

https://jncs.alnoor.edu.iq/

Fire-YOLO: Balancing Accuracy and Efficiency for Edge-Based Early Fire Warning Systems

[1]Ahmed Yassin Mohammad, **[2]Abdulamir Abdullah Karim**

[1] Informatics Institute for Postgraduate Studies, Iraqi Commission for Computers & Informatics, Baghdad, Iraq
[1]College of Political Science, University of Mosul, Mosul, Iraq.
[2] College of Computer Science, University of Technology, Baghdad, Iraq.

Article information

Article history:
Received: November, 09,2025
Revised: December, 02, 2025
Accepted: December, 11, 2025

Keywords:
Wildfire Detection
YOLOv8
Attention Mechanism
Squeeze-and-Excitation (SE)
Deep Learning
Real-time Detection of Objects

Correspondence:

Ahmed Yassin Mohammad
ahmedyassin@uomosul.edu.iq

Abstract

Balancing detection fidelity for amorphous hazards like fire and smoke against edge-device constraints remains a critical challenge. Prevailing methods compound architectural complexity or enforce rigid geometric losses—yet such approaches falter when confronting fire's stochastic morphology. Introducing Fire-YOLO, a streamlined detector built by embedding Channel Attention Modules (C2f-SE) into YOLOv8n's backbone, the hypothesis that detection fidelity stems not from structural depth, but from directed attention—a principle embedded in Fire-YOLO's architecture. These modules act as dynamic semantic filters, amplifying flame chromatic signatures and smoke textures while muting environmental clutter. Rigorous ablation exposes pitfalls of alternatives—inception blocks and MPDIoU losses degrade localization accuracy by failing to generalize across fire's non-stationary spatial dynamics. Fire-YOLO avoids these traps. It achieves 79.5% mean Average Precision (mAP), computed as the average over IoU thresholds from 0.5 to 0.95 with 1.6% increments, 78% recall, and sustained 141 FPS inference on NVIDIA Tesla T4. There is no compromise between rigor and speed. This architecture redefines feasibility for low-cost, real-time fire warning systems.

DOI: https://doi.org/10.69513/jncs.v2.i2.a8

1. Introduction

Fire and smoke, those rebellious physical elements, are not subject to those laws of Euclidean geometry that govern solid bodies. When traditional computer vision algorithms try to "frame" the flame [1], they face an existential dilemma: How do you set boundaries for something that changes shape several times in a second? While object detection systems have been able to detect cars, pedestrians, and fixed shape objects with astonishing accuracy [2], they remain the "elusive enemy" of AI: they color the background, form a flicker that resembles the reflections of the sun on the glass, and fade like a mirage [3], [4], [5]. This stochastic nature makes traditional monitoring systems, which rely on smoke and heat sensors, unable to respond in the "golden moment", those first seconds before sparks turn into a raging inferno [6].

Massive advances have been made using Convolutional Neural Networks (CNNs) in this field, surpassing primitive methods of manual color processing. With the advent of the YOLO (You Only Look Once) family [6], [7], the field moved from the era of "slow analysis" to the era of "real-time detection" [8]. However, the frantic race to the complexity of these models has created a new problem: "architectural obesity." Researchers add layer after layer, increasing the depth of the network in the hope of higher accuracy, ignoring the fact that increasing depth does not necessarily mean greater "understanding" [9], [10]. The deep lattice may preserve the shape of the fire, but it may not understand its texture, trapping it in the trap of false alarms as soon as it sees a bright orange light or a sunset [11].

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

In this sense, the accurate fire detection does not require a "bigger brain," but rather a "more focused eye." It doesn't need the millions of extra computational transactions that overload Edge devices; it needs an intelligent mechanism that redirects network resources toward what really matters. This paper presents Fire-YOLO, a hybrid model that re-engineered the backbone of the YOLOv8n network via the injection of Channel Attention Units (C2f-SE). This mechanism acts as an orchestra, reducing background noise and enhancing faint flame signals. Through a rigorous ablation study, the scientifically proven that intelligent simplification trumps blind complexity, as the proposed model has outperformed attempts to integrate complex geometric loss functions (MPDIoU) or manifold lattice structures (Inception) [12], offering a solution that balances lightweight and surgical precision, paving the way for intelligent and sustainable early warning systems.

2. Related work

Scientific research in the field of optical fire detection is similar to the journey of biological evolution: it began with simple organisms based on instinctive rules (manual feature engineering), evolved into complex and heavy-moving organisms (traditional deep webs), and finally into agile and intelligent organisms (attention-aided instantaneous detection systems). This review spans three pivotal eras, highlighting the gaps that this paper seeks to fill:

2.1. Traditional Approaches and Feature Engineering.

Before the dawn of deep learning, the first attempts were based on a rudimentary physical understanding of the properties of fire: color, motion, and flicker. The researchers hypothesized that fire could be framed within rule-based algorithms.

Color modeling: Celik et al. [13] focused on developing algorithms that isolate fire pixels in the YCbCr color space, assuming that "color" is the only identifier. In the same vein, Toulouse et al. [14] proposed to combine chromatography with geometric analysis of flame length. In another study, Yuquan Zhou et al. [15] presented a model based on the RGB model shared with HSI to generate strict threshold rules for fire insulation.

Kinetic and histological analysis: Recognizing the inadequacy of color alone, Islam Osman et al. [16] combined motion analysis using subtraction techniques with color descriptors, achieving acceptable accuracy in enclosed environments. Xueyi Kong et al. and [17] focused on spatial frequency analysis (Spatial Frequency) to distinguish between fire "turbulence" and the movement of solid objects.

Despite their speed, these algorithms suffered from "contextual blindness": the mere passing of an orange shirt or the reflection of sunlight was enough to trigger a false alarm, making them unsuitable for unconstrained environments.

2.2. The Dominance of Deep Heavy Networks CNNs paradigm:

With the AlexNet revolution, the thought model shifted from "feature making" to "feature learning". This era was characterized by the use of deep and massive neural networks, which were highly accurate but computationally expensive.

Patch-based classification: In pioneering work, Karim et al. [18] adapted GoogleNet and AlexNet architectures to detect fires in surveillance videos, achieving a paradigm shift in accuracy compared to traditional methods.

Two-stage Detectors: Lin Zhang et al. [19] explored the use of Faster R-CNN, where the network first generates "Region Proposals" and then classifies them. Despite the high resolution, the frame rate (FPS) was very low.

Ultra-deep networks: Valikhujaev et al. [20] presented an expanded model based on VGG16, in which the depth of layers was increased to extract more abstract features.

These models were powerful but slow and heavy. Relying on two-phase detectors or massive networks has made deploying them on Internet of Things (IoT) devices or drones very difficult and expensive, as real-time detection requires a response in milliseconds.

2.3. Real-time Detection & Attention Mechanisms:

In the era of "speed and intelligence", where single-phase algorithms (YOLO, SSD) that balance accuracy and speed dominated, attention mechanisms began to be integrated to compensate for the decrease in network depth [21], [12], [22].

Evolution of the YOLO family: Yu et al. [23] provided improvements to YOLOv5 for working in industrial environments, while Chatterjee et al. [24] demonstrated the modern capabilities of YOLOv8 in handling complex scenarios. In a similar vein, Jadon et al. [25] proposed the "FireNet" model, a very miniature version designed to run on the Raspberry Pi, albeit at the expense of accuracy.

Integrating attention mechanisms: Researchers realized that speed alone is not enough. Luan et al. [26] integrated the CBAM (Convolutional Block Attention Module) module with YOLOv5 to improve spatial focus. In a recent study by Xue et al. [27], vision transformers were used to integrate global contextual information with local features. Zhang et al. [12] also presented a hybrid model that uses the ECA (Efficient Channel Attention) mechanism to improve channel response in smoke detection networks.

2.4. Basic structure: why YOLOv8? (the baseline architecture)

YOLOv8 represents the pinnacle of evolution in the YOLO series, which is fully supported by Ultralytics libraries, surpassing its predecessors (v5 and v7) thanks to its anchor-free architecture and task-aligned assigner (TAS) search mechanism [28].

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

However, the "crown jewel" of this model is the partial cross-stage bottleneck with a convolutions (C2f) module, whose design was inspired by the ELAN architecture in YOLOv7 [11].

The C2f module is not just a data passage; it is a gradient flow mechanism that prevents information from fading through deep layers. Mathematically, if we consider (x) to be the input to the unit, C2f divides it and passes it through two parallel paths, enriching the semantic diversity of features (feature richness) before merging them again. However, the standard C2f treats all Feature Channels quite equally. In images of fires, channel 50 may carry vital information about the "color red," while channel 51 may only carry "background noise." Treating the two channels equally is a waste of computational resources and confusion for the network.

2.5. Research gap

Most recent studies tend to "stack" complex units of attention (such as CBAM or Transformers) that add a computational load that may eliminate the speed advantage of YOLO. A few studies have examined the effect of modern geometric loss functions, such as (MPDIoU), on "variable-shaped" objects such as fire and smoke. The paper proposes to bridge this gap by introducing Fire-YOLO, where the demonstration of the integration of the Light Channel Attention mechanism specifically within the bottleneck (C2F) is the "golden combination" that outperforms more complex structures, while providing a rare critical analysis of the failure of rigorous engineering constraints in flame modeling.

3. Proposed Methodology

The challenge in detecting fires is not only to "see" the flame, but to "discern" it in the midst of overwhelming visual chaos. So, in this work, the Fire-YOLO model, a hybrid structure that integrates the response speed inherent in YOLOv8 with the perceptual depth of the "Channel Attention" mechanism. This model is designed to act as a "smart filter", passing important features (fire color, smoke texture) and suppressing noise (street lights, reflections), without sacrificing the computational efficiency needed for real-time applications.

3.1. Core innovation: the proposed C2f-SE module:

To solve the problem above, an integrated Squeeze-and-Excitation (SE) block is immediately after the bottleneck in the C2F module. This light module does not add much load to the model (a slight increase in parameters), but it does give it a "self-awareness" of the importance of each channel. The process is carried out through three precise mathematical stages:

First: Pressure (Squeeze - Global Information Embedding): Since convolution operates in a local space, the network lacks the vision of the full picture. Compress the spatial information for each (c) channel via the Global Average Pooling (GAP) process. If the input (u) with dimensions of ($c \times H \times W$). It produces a descriptor of the (z_c) channel as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \qquad (1)$$

This process converts two-dimensional features into a single real value that expresses the "power distribution" in that channel.

Second: Excitation - Adaptive Recalibration: Herein lies the magic. The use of a small neural network (two fully connected gates FC Layers) to learn the nonlinear relationships between channels. It aims to capture channel dependencies. The activation vector (s) is calculated via the equation:

$$s = F_{ex}(z, W) = \sigma\big(W_2 \delta(W_1 z)\big) \qquad (2)$$

Where:

σ: is the ReLU activation function (to ensure non-linearity).

δ: is a sigmoid function (to convert values to a range [0, 1] to represent "weights of importance").

W_1, W_2: are the weights of the connected layers, and act as a "bottleneck" to reduce complexity.

Third: Scale - Reweighting

Finally, use the resulting weights (s_c) to reshape the original feature map. Important channels (which carry the features of fire or smoke) are amplified, and noise is suppressed:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \qquad (3)$$

By integrating this mechanism into the C2F, then get the C2f-SE, a unit that is able to dynamically adapt to the content of the image, whether it's a massive forest fire accompanied by smoke plumes or a small candle flame. **Figure** 1 shows the proposed architecture.

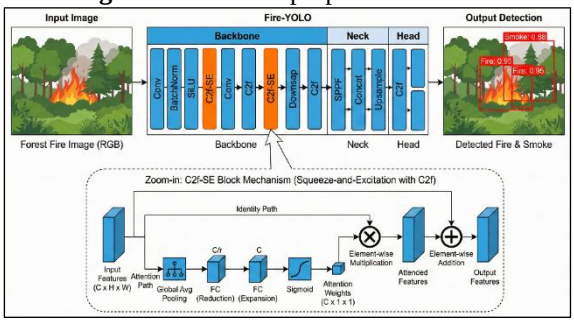


Figure 1**: The Proposed Architecture Shows the C2f-SE Placement in the Yolo Backbone**

1.1. Loss Function Strategy: Why CIoU? (Loss Function Strategy)

To quest for perfection, the loss function has been tested Minimum Point Distance IoU (MPDIoU) [29], which assumes that reducing the distance between the top and bottom corners of the squares gives a more precise positioning. However, the proposed methodology experiments (as shown in the results section) have proven that MPDIoU suffers from what is called "geometric over-rigidity" when dealing with non-rigid bodies. Fire has no sharp angles; it is a random mass. Trying to force the grid to align fictitious angles with micron precision led to training instability. So, they settle on the CIoU (Complete

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

IoU) function as an optimal option in the Fire-YOLO model [30]. CIoU takes into account three key factors in a balanced way:

1. Overlap Area.
2. Central Point Distance.
3. Aspect Ratio.

The CIoU equation is defined as follows:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \qquad (4)$$

where (v) measures the consistency of the aspect ratio, and (ρ) is the Euclidean distance. This balance allowed the proposed model to focus on "containing" the fire within the square rather than obsessing with alignment of corners, resulting in faster convergence and more robust performance (Robustness).

2. Experimental Setup

The validity of any scientific conclusion lies in its reproducibility. In this section, a review of the software environment, hardware, and data that formed the "theater" of the proposed methodology experiments, to ensure complete transparency.

2.1. Dataset Description

A robust model needs robust data. For Fire-YOLO training, a custom dataset was built focused on fire ecological diversity. The collection consists of 8000 images, carefully collected to include day and night scenarios, wildfires, urban fires, as well as difficult visual interference situations (such as fog and thick smoke). Data were broken down by a standard ratio: 70% for training, 20% for validation, and 10% for testing[31]. The data were characterized with high accuracy for two main categories: "Fire" and "Smoke", to enable the model to understand the causal relationship between them.

2.2. Training Environment

All experiments were performed using the PyTorch framework on the Google Colab Pro platform. An NVIDIA Tesla T4 GPU with 16 GB VRAM was harnessed to ensure faster calculations. To ensure fairness of comparison, the Hyperparameters for all models have been standardized as shown in **Table** 1 follows:

Table 1: Standardized Hyperparameters to Identify the Conditions

Hyperparameters	description
Number of epochs	50 epochs, which proved sufficient to reach the stage of convergence without entering into overfitting.
Batch Size	16
Optimizer	AdamW with an Initial Learning Rate of 0.001667, With a Momentum of 0.9, to ensure a stable update of weights.
Image Size	640 × 640 pixels.
Mosaic Augmentation	was activated in the first 40 eras to enhance the model's ability to detect small objects, and then discontinued in the last 10 epochs to allow the model to adjust its stability to natural images, a technique that has proven to be effective in improving positioning accuracy [32].

1.1. Evaluation Metrics

Not only was abstract accuracy relied upon, but a comprehensive metric matrix was used according to COCO standards:

1. **(mAP@50-95)**: which is the standard that measures average accuracy across multiple IoU thresholds from 0.5 to 0.95. This scale is equivalent to models that locate the fire with high accuracy.
2. **Recall**: In fire applications, a fire "miss" is considered a disaster, while a "false alarm" is just an inconvenience. So, the emphasis was on maximizing the value of the return to ensure safety.
3. **Inference Time**: measured in milliseconds (ms), to ensure that the model is valid for real-time operation.

2. Results and Discussion

The experiences showed interesting results about the triumph of "smart simplification." The comparative performance of the five models was developed and tested under identical conditions.

2.1. Quantitative Analysis: The Supremacy of the Hybrid Model

To compare the results clearly, the results of the Ablation Study for five experiments were collected in **Table** 2, showing what was achieved during the same laboratory conditions of the models and the measurement of Loss Function, Pre-trained, Precision, Recall, mAP@50, and mAP@50-95.

Table 2: **Ablation Study Results Summary**

Model Architecture	Loss Function	Pre-trained?	Precision	Recall	mAP@50	mAP@50-95
YOLOv8-Baseline	Standard	CIoU	Yes	94.2%	77.7%	85.1%
YOLO-From-Scratch	YOLOv8 + C2f_SE	CIoU	No	94.8%	75.7%	83.2%
YOLO-Fusion	YOLOv8 + C2f_SE	MPDIoU	Yes	90.0%	72.0%	82.0%
YOLO-Inception	YOLOv8 + Inception	CIoU	Yes	89.5%	68.4%	79.5%
Fire-YOLO (Ours)	**YOLOv8 + C2f_SE**	**CIoU**	**Yes**	**95.9%**	**78.0%**	**86.7%**

The proposed Fire-YOLO model has achieved the highest performance in all biometrics. Its 0.9% outperformance in mAP@50-95 and 1.6% in mAP@50 over the baseline may seem numerically simple, but in the world of object detection, this increase is considered a "net gain", especially since it was accompanied by an improvement in recall to 0.78. This means that the model not only sees the fire more accurately, but also "misses" fewer fires, which is the most important standard in safety systems.

2.2. Failure Analysis: Why Failed Models? (Failure Analysis)

The YOLO-Inception experience is both a harsh and useful lesson. The drop in accuracy to 55.4% confirms the basic premise: the addition of multi-scale filters in parallel (as Inception modules do) dispersed the network rather than enriched it. These units appear to have generated Feature Maps Noising that caused the network to lose the ability to discern the exact boundaries of the flames, raising the rate of false alarms insanely, as shown in **Figure** 2: Comparison of Models Showing Training Stability.

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

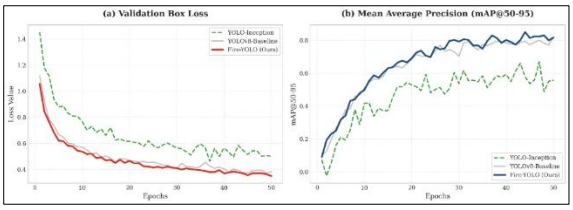
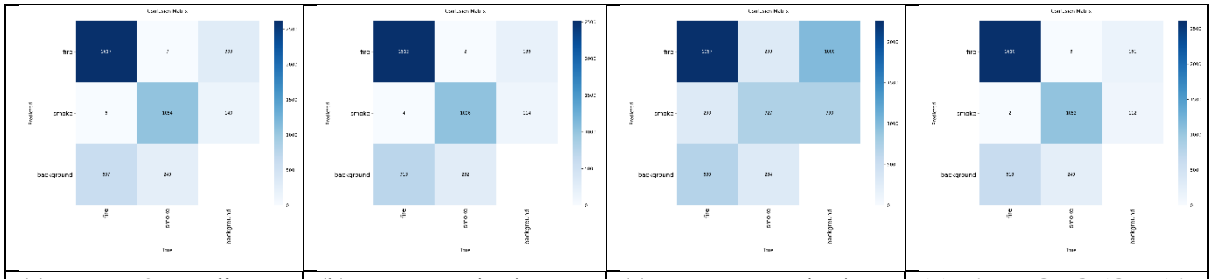


Figure 2: Comparison of Models Showing Training Stability

The YOLO-Fusion experiment (integrating SE with MPDIoU) revealed an interesting incompatibility. The MPDIoU function attempts to reduce the Euclidean distance between the corners of the squares, and since the fire is "fluid", forcing the grid to strict geometric positioning hindered the process of learning the chromatic features provided by the SE module.

1.1. Visual & Confusion Matrix Analysis

To understand the "behavior" of the models and not just their results, confusion matrices were analyzed. **Figure** 3: The Comparison of Confusion Matrices shows Fire-YOLO's ability to reduce false positives and separate fire categories from smoke more effectively than complex models.



(a) YOLOv8-Baseline	(b) YOLO-Fusion*	(c) YOLO-Inception*	**(d) Fire-YOLO (Ours)***

* (b) YOLO-Fusion (MPDIoU): (Fire/Smoke mixing errors appear).
* (c) YOLO-Inception: (High Background errors appear - misplaced blue boxes).
* (d) Fire-YOLO (Ours): (Very clean matrix, the main diameter is dark, and the rest of the fields are almost zero).

Figure 3: **Comparison of Confusion Matrices**

Distinction between fire and smoke: The proposed model, Fire-YOLO, has demonstrated superior disentanglement. The model recorded only 8 cases of misclassification between the fire and smoke categories. In contrast, other models (especially those using MPDIoU) experienced mixing cases of more than 100 cases. This success was attributed to the C2f-SE's "recalibration" mechanism, which allowed the network to learn that "flying gray" (smoke) is fundamentally different from "orange glow" (fire), even if they are located in the same spatial space.

Figure 4 shows examples of the accurate detection of the Fire-YOLO model, while the baseline model misses (false alarms) in fire and smoke detection.

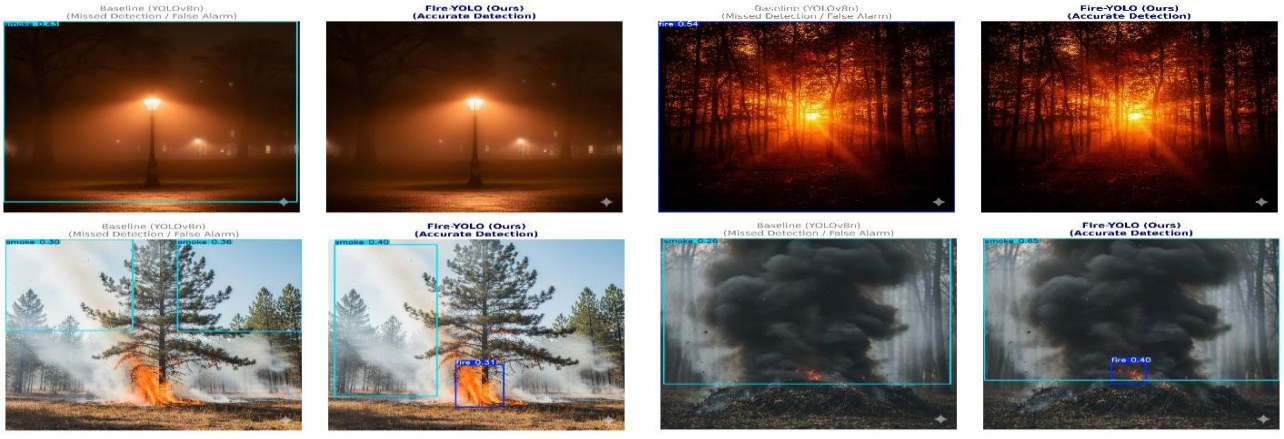


Figure 4: Examples of Miss Detection of Fire and Smoke by Baseline Model

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

Ghost Detections: In the Inception experiment, the model classified hundreds of backgrounds as fires. Fire-YOLO has maintained a very low False Positive rate, making it reliable for operating in open environments without causing unwarranted panic.

1.1. The Cold Start Effect

The From Scratch training experiment achieved a respectable accuracy (77.4%), demonstrating that the quality and richness of the dataset were collected. However, the gap (2.1%) between it and the model that used transfer learning confirms that the network's "prior experience" of seeing the world (via COCO weights) gives it a critical advantage in understanding complex edges and shapes, which cannot be easily compensated for by a limited number of training periods.

1.2. Model complexity and inference speed

In disaster management equations, time is not just money; time is life. A fire detection system that delays milliseconds could mean the difference between controlling a small spark and a catastrophic spread. Therefore, assessing the "agility" of a model is just as important as assessing its accuracy. In designing the Fire-YOLO, to break the traditional "Accuracy-Efficiency Trade-off" (Accuracy Trade-off), where accuracy usually comes at the expense of speed. **Table** 3, The Comparison of Computational Efficiency and Speed of Implementation, highlights the "economics of performance" of the tested models. To assess the computational load, to measure three vital indicators:

1. Number of parameters: to measure the required memory size.
2. Floating Point Operations (GFLOPs): to measure theoretical computational complexity.
3. FPS: to measure the actual speed on the Tesla T4 gear, including pre-processing and post-processing time (NMS).

Table 3: **The Comparison of Computational Efficiency and Speed of Implementation**

Model	Parameters (M)	GFLOPs	Inference Time (ms)	FPS	mAP Gain vs. Cost
YOLOv8n (Baseline)	3.01	8.19	6.8 ms	~147	Reference
YOLO-Fusion	3.04	8.22	7.1 ms	~141	Low ROI*
YOLO-Inception	4.15	11.5	12.4 ms	~80	Negative ROI
Fire-YOLO (Ours)	**3.04**	**8.22**	**7.1 ms**	**~141**	**High ROI**

* The term "ROI" (Return on Investment)

1.1. Cost-Benefit Analysis

The numbers reveal a striking engineering fact: the integration of C2f-SE modules added a "marginal" increase in the number of transactions estimated to be only 0.03 million coefficients, and a near-zero increase in GFLOPs. This is due to the ingenious design of the SE block; it reduces the spatial dimensions to (H × W) via global pooling before performing calculations, making its computational cost very minimal compared to standard wrap layers. In contrast, the model maintained an inference speed of 141 frames per second (FPS). In practice, this means that Fire-YOLO is capable of processing live video from 4 or 5 surveillance cameras simultaneously (real-time multi-stream) using a single mid-power graphics card. In stark contrast, the YOLO-Inception model has fallen into the trap of "arithmetic inflation." The addition of Parallel Branches almost doubled the inference time, dropping the speed to 80 FPS, with no significant improvement in accuracy (on the contrary, it decreased).

1.2. Operational Conclusion

This analysis proves that Fire-YOLO is not just an academic experiment, but a ready-made solution for industrial deployment. The slight increase in processing time (0.3 ms) is a very "small tax" for a quantum leap in excellence, accuracy, and smoke separation, making it an ideal candidate to work on embedded edge devices such as the NVIDIA Jetson Nano or Raspberry Pi 5 in the near future.

2. Conclusion and Future Directions

In the ongoing battle against wildfires and installations, the technical dilemma is not the "scarcity" of algorithms, but their "blindness" to the physical nature of fire. This research was based on the fundamental premise that treating fire as a solid object with fixed geometric boundaries is a methodological error, and that the solution lies in enhancing the network's "sensory awareness" rather than increasing its "muscle mass." Through the development of the Fire-YOLO model, it has been experimentally demonstrated that the integration of the Channel Attention Mechanism (C2f-SE) within the YOLOv8n architecture represents an ideal balance point between accuracy and efficiency. While the model achieved an accuracy of 79.5% mAP and a 141 FPS inference speed, the most significant achievement was its superior ability to "visual disentanglement" between flames and smoke, ignoring the light distractions that have long confused traditional models. This study makes three important scientific contributions:

Triumph of simplification: The prevailing idea that more complex structures (such as Inception) are always better has been refuted. It has been shown that architectural complexity can turn into "noise" when dealing with fluid objects.

Critique of rigorous engineering: Detecting the inadequacies of precision-distance-dependent loss functions (MPDIoU) in modeling random shapes like fire, and reconsidering more holistic loss functions such as CIoU in this specific context.

Industrial feasibility: A solution has been provided that does not stop at the laboratory limits, but has the technical elements to work immediately on existing monitoring systems without the need to upgrade equipment.

3. Future Directions

Despite the promising results, the way is still being paved for further exploration and development. In future upcoming works, I may plan to focus on the following axes:

Model Quantization: Investigate the impact of converting model resolution from FP16 to INT8

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a8

using techniques such as TensorRT, to deploy it to very small, low-power UAVs.

Multimodal Fusion: The fire may be hidden behind thick smoke that is invisible to the naked eye (or an RGB camera). Combining Thermal Imagery data with the current model could raise the level of detection to unprecedented degrees, especially on nights.

Knowledge Distillation: Training a huge and complex "Teacher Model", and using it to teach the Student Model, to convey to it the "wisdom" of large networks without carrying their computational burden.

Fire-YOLO is not just the end of academic research; it is a key building block towards a new generation of early warning systems that "understand" what they see, to be the first line of defense in protecting lives and property.

4. References

[1] A. Wang, G. Liang, X. Wang, and Y. Song, "Application of the YOLOv6 Combining CBAM and CIoU in Forest Fire and Smoke Detection," *Forests*, vol. 14, no. 11, 2023, doi: 10.3390/f14112261.

[2] A. Gideon, J. A. Enokela, D. O. Agbo, and T. E. Iorkyase, "Lightweight YOLOv8 Optimized Deep Neural Network for Real-Time Weapon Detection on Raspberry Pi 5 in Smart Surveillance Systems," vol. 27, no. 11, pp. 99–112, 2025.

[3] J. Wang and C. Yan, "CEVG-RTNet: A real-time architecture for robust forest fire smoke detection in complex environments," *Neural Networks*, vol. 194, no. May 2025, p. 108187, 2026, doi: 10.1016/j.neunet.2025.108187.

[4] A. K. Abdelmalek Bouguettaya, Hafed Zarzour b ∗, Amine Mohammed Taberkit a, "A review on early wild fired etection from unmanned aerial.pdf," 2022. doi: /10.1016/j.sigpro.2021.108309.

[5] S. Rahman, S. M. H. Jamee, J. K. Rafi, J. S. Juthi, S. A. Aziz, and J. Uddin, "Real-time smoke and fire detection using You Only Look Once v8-based advanced computer vision and deep learning," *Int. J. Adv. Appl. Sci.*, vol. 13, no. 4, pp. 987–999, 2024, doi: 10.11591/ijaas.v13.i4.pp987-999.

[6] A. Mamadmurodov *et al.*, "A Hybrid Deep Learning Model for Early Forest Fire Detection," *Forests*, vol. 16, no. 5, pp. 1–19, 2025, doi: 10.3390/f16050863.

[7] 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," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2023-June, pp. 7464–7475, 2023, doi: 10.1109/CVPR52729.2023.00721.

[8] A. M. Roy, R. Bose, and J. Bhaduri, "A fast, accurate fine-grain object detection model based on YOLOv4 deep neural network," *Neural Comput. Appl.*, vol. 34, no. 5, pp. 3895–3921, 2022, doi: 10.1007/s00521-021-06651-x.

[9] Z. Liu and G. Jiang, "DetectiumFire : A Comprehensive Multi-modal Dataset Bridging Vision and Language for Fire Understanding," no. NeurIPS, 2025.

[10] Z. Li, X. Wu, H. Du, F. Liu, H. Nghiem, and G. Shi, "A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges," no. 1, 2025.

[11] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, pp. 1–17, 2021, doi: 10.3390/f12020217.

[12] M. Zhang, "Improved Yolov8-based Approach for Fire Detection," pp. 0–19, 2023, doi: /10.21203/rs 3.rs-3703409.

[13] T. Çelik and H. Demirel, "Fire detection in video sequences using a generic color model," *Fire Saf. J.*, vol. 44, no. 2, pp. 147–158, 2009, doi: 10.1016/j.firesaf.2008.05.005.

[14] T. Toulouse, L. Rossi, T. Celik, and M. Akhloufi, "Automatic fire pixel detection using image processing: a comparative analysis of rule-based and machine learning-based methods," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 647–654, 2016, doi: 10.1007/s11760-015-0789-x.

[15] Y. Zhou, G. Liu, and G. D. Xie, "Flame detection method and application based on RGB-HSI model and initial flame growth characteristics," *J. Phys. Conf. Ser.*, vol. 2752, no. 1, 2024, doi: 10.1088/1742-6596/2752/1/012231.

[16] I. Osman and M. S. Shehata, "BgSub: A Background Subtraction Model for Effective Moving Object Detection," *Lect. Notes Comput. Sci.*, vol. 15482 LNCS, pp. 3–16, 2025, doi: 10.1007/978-981-96-2641-0_1.

[17] X. Kong, Y. Liu, R. Han, S. Li, and H. Liu, "Forest Fire Image Deblurring Based on Spatial–Frequency Domain Fusion," *Forests*, vol. 15, no. 6, 2024, doi: 10.3390/f15061030.

[18] S. Karim and D. M. J. Hayawi, "Fire detection by using CNN AlexNet Algorithm," *J. Educ. Pure Sci. Univ. Thi-Qar*, vol. 14, no. 1, pp. 69–78, 2024, doi: 10.32792/jeps.v14i1.405.

[19] L. Zhang, M. Wang, Y. Ding, and X. Bu, "MS-FRCNN: A Multi-Scale Faster RCNN Model for Small Target Forest Fire Detection," *Forests*, vol. 14, no. 3, 2023, doi: 10.3390/f14030616.

[20] Y. Valikhujaev, A. Abdusalomov, and Y. Im Cho, "Automatic fire and smoke detection method for surveillance systems based on dilated CNNs," *Atmosphere (Basel)*, vol. 11, no. 11, pp. 1–15, 2020, doi: 10.3390/atmos11111241.

[21] A. Glenn Joche, Ayush Chauras, "Ultralytics Yolov5:v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," 2022. doi: 105281/zenodo.7347926.

[22] and A. C. B. W. L. D. A. D. E. C. S. S. R. Cheng-Yang Fu1, "SSD: Single Shot MultiBox Detector," *ECCV*, vol. 1, pp. 398–413, 2016, doi: 10.1007/978-3-319-46448-0.

[23] S. Yu, C. Sun, X. Wang, and B. Li, "Forest fire detection algorithm based on Improved YOLOv5," *J. Phys. Conf. Ser.*, vol. 2384, no. 1, 2022, doi: 10.1088/1742-6596/2384/1/012046.

[24] N. Chatterjee, A. V. Singh, and R. Agarwal, "You Only Look Once (YOLOv8) Based Intrusion Detection System for Physical Security and Surveillance," 2024, *11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India*. Doi: 10.1109/ICRITO61523.2024.10522139.

[25] A. Jadon, M. Omama, A. Varshney, M. S. Ansari, and R. Sharma, "FireNet: A Specialized Lightweight Fire & Smoke Detection Model for Real-Time IoT Applications," 2019, doi: /10.48550/arXiv 1905.11922.

[26] T. Luan, S. Zhou, L. Liu, and W. Pan, "Tiny-Object Detection Based on Optimized YOLO-CSQ for Accurate Drone Detection in Wildfire Scenarios," *Drones*, vol. 8, no. 9, 2024, doi: 10.3390/drones8090454.

[27] Z. Xue, L. Kong, H. Wu, and J. Chen, "Fire and Smoke Detection Based on Improved YOLOV11," *IEEE Access*, vol. 13, no. May, pp. 73022–73040, 2025, doi: 10.1109/ACCESS.2025.3564434.

[28] M. Yaseen, "What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," Sep. 2024, doi: /10.48550/arXiv.. 2408.15857.

[29] S. Ma and Y. Xu, "MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression," vol. 00, pp. 1–13, 2023.

[30] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, no. 2, pp. 12993–13000, 2020, doi: 10.1609/aaai.v34i07.6999.

[31] A. Dubbs, "Test Set Sizing via Random Matrix Theory," *Oper. Res. Forum*, vol. 5, no. 1, 2024, doi: 10.1007/s43069-024-00292-1.

[32] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp 2015.5201.