










Improving Early Fire and Smoke Detection with YOLOv12N via Histogram-Based Contrast Enhancement

Christine Dewi^{1,2}, Hanna P. Chernovita³, Melati V. V. Santoso¹, Evangs Mailoa¹, Stephen A. Philemon¹,
Stephen A. Sutresno⁴, Abbott P. S. Chen^{5,6*}

¹ Department of Information Technology, Satya Wacana Christian University, Salatiga 50711, Indonesia

² School of Information Technology, Deakin University, Burwood Campus, Burwood Victoria 3125, Australia

³ Department of Information Systems, Satya Wacana Christian University, Salatiga 50711, Indonesia

⁴ Department of Information System, Faculty of Bioscience, Technology, and Innovation, Atma Jaya Catholic University of Indonesia, Tangerang 15339, Indonesia

⁵ Department of Marketing and Logistics Management, Chaoyang University of Technology, Taichung City 413310, Taiwan

⁶ Artificial Intelligence Department, Honchita Co. Ltd., Changhua County 503005, Taiwan

Corresponding Author Email: chprosen@gm.cyut.edu.tw

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijse.160217>

ABSTRACT

Received: 10 December 2025

Revised: 11 February 2026

Accepted: 21 February 2026

Available online: 28 February 2026

Keywords:

fire and smoke detection, wildfire monitoring, YOLOv12N, contrast-limited adaptive histogram equalization, image enhancement, low-contrast imagery, D-Fire dataset

Early visual detection of wildfire-related smoke and flame remains difficult under low-contrast conditions, especially when smoke is thin, spatially dispersed, or partially obscured by haze and complex backgrounds. This study examines whether histogram-based contrast enhancement can improve the detection performance of a lightweight YOLOv12N detector for early fire and smoke recognition. Three preprocessing strategies—global histogram equalization (HE), contrast-limited adaptive histogram equalization (CLAHE), and Dynamic Block Size Technique with Local Contrast Modification (DBST-LCM) CLAHE—were evaluated on the D-Fire dataset under a unified training configuration. The results show that CLAHE provides the most consistent improvement across detection metrics. Compared with the baseline YOLOv12N model trained on unenhanced images, CLAHE-enhanced YOLOv12N increases precision from 78.8% to 84.4%, recall from 73.0% to 78.5%, mAP50 from 80.1% to 86.7%, and mAP50-95 from 47.6% to 57.2%. DBST-LCM CLAHE ranks second, improving local contrast in difficult scenes but occasionally smoothing fine smoke structures, while global HE produces the weakest results due to noise amplification and unstable feature representation. Loss-curve analysis and qualitative comparisons further confirm that localized contrast enhancement improves both convergence behavior and visual sensitivity to faint smoke. Runtime analysis indicates that the proposed preprocessing introduces only negligible latency and remains compatible with real-time deployment. These findings demonstrate that localized histogram-based enhancement, particularly CLAHE, is an effective and computationally practical strategy for strengthening.

1. INTRODUCTION

Wildfires, also referred to as Forest and Land Fires (FLF), have become one of the most severe environmental hazards of recent decades. Prolonged droughts, land-use change, and climate variability have intensified fire activity worldwide, resulting in extensive ecological damage, biodiversity loss, and increased greenhouse gas emissions that further reinforce global warming trends. Recent global assessments report a sustained increase in both wildfire frequency and burned area, accompanied by sharp rises in fine particulate matter that degrade air quality and pose significant public health risks [1, 2]. During the 2023–2024 fire season alone, approximately 3.9 million km² of land burned globally, with associated carbon emissions exceeding the historical average by 16% [3].

Beyond direct environmental impacts, wildfire smoke

represents a growing threat to human health. Exposure has been linked to respiratory illness, elevated cancer risk, and long-term psychological effects. Projections for the United States alone estimate between 1,300 and 32,000 premature deaths annually by 2050 due to wildfire-related smoke exposure [4]. Recent large-scale fire events—including the 2024 Amazon fires (> 51,000 km² burned), the 2023 Northern Australia wildfires (≈ 840,000 km² affected), and tens of thousands of fire incidents across Brazil's Pantanal, Cerrado, and Amazon regions—highlight the urgent need for reliable early-stage wildfire detection systems [5, 6].

Despite the expansion of wildfire monitoring infrastructure, early detection remains challenging. Thin smoke plumes, low contrast, haze, and visually complex backgrounds often obscure the earliest visual cues of ignition. While satellite-based monitoring and public reporting systems provide broad

spatial coverage, they frequently suffer from delayed response times and limited sensitivity during the initial stages of fire development. These constraints have motivated increasing interest in automated vision-based detection using deep learning, particularly for real-time applications.

In this context, this study investigates the integration of histogram-based contrast enhancement techniques with a lightweight yet architecturally advanced YOLOv12 detector to improve early-stage wildfire detection under adverse visual conditions. The underlying premise is twofold: localized contrast enhancement can expose subtle smoke structures that are otherwise difficult to detect, while architectural refinements in You Only Look Once version 12 (YOLOv12) enable stable and efficient feature learning without increasing model complexity or computational cost.

Vision-based wildfire monitoring systems are already deployed at scale in operational settings. The ALERTCalifornia network, for example, operates more than 1,144 high-definition pan-tilt-zoom cameras equipped with near-infrared capabilities, providing continuous surveillance across fire-prone regions [7]. Parallel to this, YOLO-based detectors have demonstrated strong performance in wildfire-related tasks. Spiller et al. [8] leveraged hyperspectral imagery for early detection using convolutional networks, while Al-Smadi et al. [9] reported a Mean Average Precision (mAP)_{@0.5} of 96.8% for smoke detection using YOLOv5x. However, many existing approaches rely on raw or minimally processed imagery, which limits robustness under low illumination, haze, or smoke-dominated conditions.

YOLOv12 is selected in this study not only for inference speed, but for its architectural evolution relative to earlier YOLO variants. While YOLOv11 introduced enhancements such as Spatial Pyramid Pooling - Fast (SPPF) and Cross Stage Partial with Spatial Attention (C2PSA) to improve representational capacity, YOLOv12 consolidates context modeling and attention mechanisms within a unified and more efficient design. In particular, the YOLOv12-N configuration avoids residual scaling effects, allowing architectural behavior to be evaluated without confounding influences from model size. This makes YOLOv12-N well suited for examining how visual preprocessing interacts with detector performance in early wildfire detection scenarios.

Within this framework, three histogram-based image enhancement techniques—global Histogram Equalization (HE), Contrast-Limited Adaptive Histogram Equalization (CLAHE), and Dynamic Block Size Technique with Local Contrast Modification (DBST-LCM) CLAHE—are evaluated for their ability to improve visual representation prior to detection. While previous studies combining YOLO with global HE reported improved brightness, they often failed to preserve local contrast, leading to incomplete smoke representation or increased false positives under haze-dominated conditions. By integrating localized and adaptive enhancement strategies with YOLOv12, this study aims to improve detection reliability while maintaining operational feasibility.

This work pursues three objectives:

1. To evaluate the effectiveness of image enhancement techniques such as HE and its variants in improving the visual quality of wildfire-related imagery, especially in conditions where thin smoke or distant flames are difficult to detect.

2. To implement and assess the performance of the YOLOv12 model in detecting fire and smoke within enhanced images, focusing on detection accuracy, precision, recall, and inference speed.

3. To compare the YOLOv12-enhanced detection pipeline against baseline scenarios, including unenhanced inputs and earlier YOLO versions (YOLOv11 and prior architecture), to quantify the added value of visual preprocessing and architectural improvements.

The main contributions of this study include the integration of histogram-based image enhancement with YOLOv12 for early wildfire detection, a comprehensive evaluation of detection performance under varied enhancement conditions, and a comparative analysis with unenhanced imagery and previous YOLO models to assess improvements in both accuracy and operational feasibility.

The remainder of this paper is structured as follows: Section 2 reviews related work on YOLO-based wildfire detection; Section 3 describes the materials, methods, dataset, image enhancement techniques, and evaluation metrics; Section 4 presents experimental results and discussion; and Section 5 concludes the study and highlights potential directions for future research.

2. RELATED WORK

2.1 Forest and Land Fire detection based on smoke and fire features

FLF detection focuses on identifying visual cues of smoke and flame across diverse environments using platforms such as satellites, fixed CCTV systems, and unmanned aerial vehicles (UAVs). Early approaches relied on handcrafted features, including color segmentation and motion analysis, but these methods often failed in complex scenes due to small object sizes, low contrast, and frequent false alarms. The emergence of deep learning, particularly convolutional neural networks (CNNs), significantly improved detection accuracy; however, adapting pretrained models to relatively small and imbalanced FLF datasets remains challenging. To mitigate catastrophic forgetting during fine-tuning, techniques such as Learning without Forgetting (LwF) have been introduced, enabling models to retain generic visual representations while adapting to fire and smoke detection tasks [10].

Object detection frameworks, especially those from the YOLO family, have further advanced FLF detection by offering an effective balance between accuracy and real-time performance. YOLOFM extends YOLOv5N through multi-scale feature integration, hardware-aware pyramid design, and optimized loss functions, leading to improvements in both mAP₅₀ and mAP₅₀₋₉₅ while maintaining efficiency [11]. Lightweight architectures such as GNN-YOLO employ GhostConv, coordinate attention, and semantic-preserving pooling to reduce computational cost and suppress false positives [12]. Other approaches include Detection Transformer (DETR)-based detectors enhanced with normalization strategies and multi-scale deformable attention to improve small-object recognition [13], as well as integrated systems combining traditional image processing (e.g., color segmentation and contour extraction) with deep learning and real-time alert mechanisms [14].

To improve robustness across diverse environmental conditions, Adversarial Fusion Networks (AFN) have been proposed to fuse shallow and deep features, enhancing sensitivity to small smoke plumes and improving generalization to unseen domains [15]. A comprehensive survey further introduced a taxonomy of wildfire detection

scenarios based on fire size and background complexity, analyzing 153 studies and 17 datasets to highlight data gaps and the need for stronger validation protocols [16].

Beyond detection, recent research has explored spatiotemporal wildfire forecasting using U-shaped convolutional neural network architecture (U-Net) variants, Convolutional Long Short-Term Memory (ConvLSTM), and Temporal Convolutional Networks (TCN) to model fire spread and environmental risk over time [17]. While these methods are valuable for long-term monitoring and risk assessment, they typically rely on aggregated temporal

patterns and are not designed for early-stage visual detection of faint smoke or small flames in real-time imagery. As a result, detection-oriented approaches that enhance visual sensitivity while preserving computational efficiency remain essential.

Despite ongoing progress, reliably detecting thin smoke under degraded visibility—such as haze, low illumination, or complex backgrounds—continues to be an open challenge, particularly when real-time deployment constraints are considered.

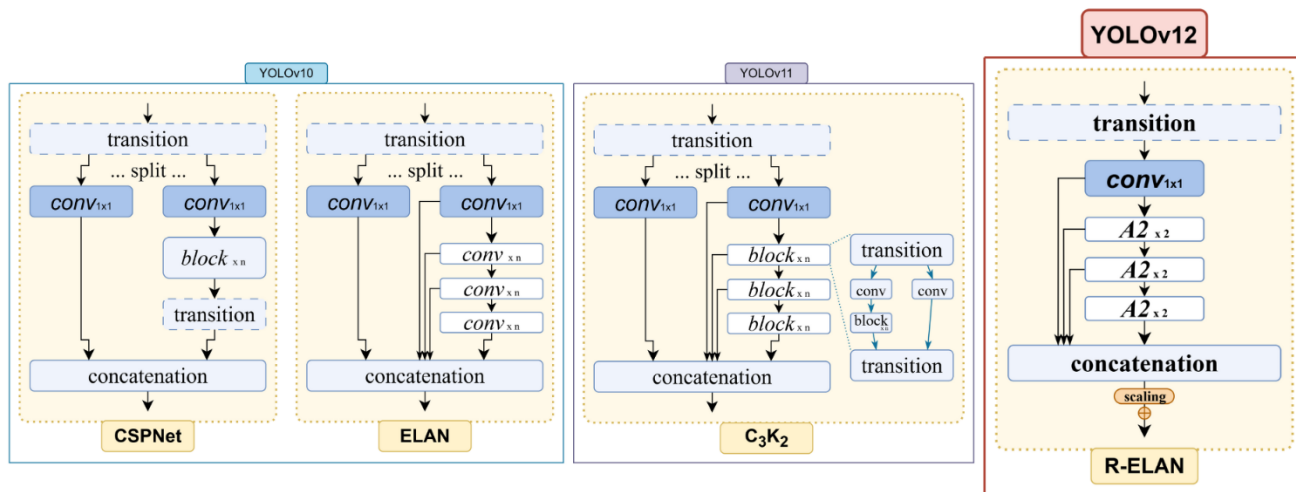


Figure 1. Residual Efficient Layer Aggregation Network (R-ELAN) backbone architecture in YOLOv12
 Note: CSPNet: Cross Stage Partial Network; ELAN: Efficient Layer Aggregation Network; C₃K₂: Cross Stage Partial with kernel size 2

2.2 YOLOv12

Over more than a decade of development, the YOLO family has evolved from the coarse object localization of YOLOv1 to architectures capable of high-accuracy, real-time detection of small and visually ambiguous objects. Successive versions have refined backbone networks, neck structures, detection heads, and training strategies, leading to improved accuracy–speed trade-offs across applications including surveillance, environmental monitoring, and wildfire detection [18].

Building on these developments, YOLOv12 introduces several architectural enhancements aimed at strengthening feature representation while maintaining inference efficiency. As illustrated in Figure 1, the Residual Efficient Layer Aggregation Network (R-ELAN) backbone improves feature aggregation through residual connections and efficient layer reuse, enabling deeper representations without excessive computational overhead. In addition, 7×7 separable convolutions expand the receptive field, allowing the model to capture broader spatial context while keeping parameter growth modest. FlashAttention further enhances spatial selectivity by emphasizing visually salient regions, which is particularly beneficial in cluttered or low-contrast scenes.

For wildfire detection, the architectural components of YOLOv12 address key visual challenges inherent in fire and smoke imagery [19, 20]. The R-ELAN backbone helps preserve subtle intensity gradients and thin smoke textures that are often lost in shallower networks, reducing missed detections under low-visibility conditions. In addition, the larger receptive field introduced by 7×7 separable convolutions supports the recognition of elongated and dispersed smoke plumes over wide areas, while the

FlashAttention mechanism improves sensitivity to transient flame cues and low-contrast regions. Together, these design choices enhance detection robustness across varying lighting conditions, haze levels, and partial occlusions.

Although previous studies have examined YOLO-based models up to YOLOv11 for fire and smoke detection [19, 20], the interaction between YOLOv12’s architectural design and visual preprocessing strategies has not yet been systematically investigated. The YOLOv12 implementation is publicly available on GitHub.

2.3 Visual preprocessing

Visual preprocessing plays a critical role in enhancing detection sensitivity for FLI imagery, particularly under low light, haze, or visually cluttered conditions. This study focuses on histogram-based enhancement methods—HE, Contrast Limited Adaptive HE (CLAHE), and DBST-LCM combined with CLAHE—because of their ability to improve global and local contrast, making subtle smoke and flame features more distinguishable.

HE enhances contrast by redistributing pixel intensities across the full dynamic range of the image, offering a simple and computationally efficient solution suitable for real-time pipelines [21]. CLAHE extends this concept by applying localized contrast enhancement within small tiles while limiting amplification, thereby preserving fine details and reducing over-enhancement artifacts [22, 23]. DBST-LCM CLAHE further refines this approach by dynamically adjusting block sizes based on image characteristics, improving noise suppression while retaining salient visual structures [24, 25]. Within DBST-LCM CLAHE, PSNR is

used as a relative stability indicator to guide block size selection, rather than as a direct measure of perceptual quality. This adaptive strategy has demonstrated consistent improvements in precision and mAP on the D-Fire dataset.

Despite its simplicity, global HE can be problematic in wildfire scenes. High-intensity flame regions often dominate the global histogram, leading to excessive contrast amplification in bright areas while compressing low-intensity smoke gradients. This imbalance may exaggerate flame cores, introduce artificial edges, and increase background noise, ultimately destabilizing the representation of diffuse smoke structures.

Overall, these histogram-based methods act as low-cost visibility enhancement tools that increase the signal-to-noise ratio and support more reliable detection in degraded imagery [26]. However, most prior studies evaluate preprocessing techniques and detection models independently. Their combined effects on detection accuracy, robustness, and inference behavior—particularly when integrated with modern architectures such as YOLOv12—remain insufficiently explored. This study addresses this gap by systematically integrating HE, CLAHE, and DBST-LCM CLAHE with YOLOv12 and evaluating their impact on detection performance and efficiency for early-stage wildfire recognition.

2.4 Performance metrics

To evaluate wildfire detection performance, this study adopts standard object detection metrics. Intersection over Union (IoU), defined in Eq. (1), measures the overlap between predicted and ground-truth bounding boxes as the ratio of their intersection to their union. A detection is considered correct when the IoU exceeds a predefined threshold.

$$IoU = \frac{Area_{pred} \cap Area_{gt}}{Area_{pred} \cup Area_{gt}} \quad (1)$$

Precision (P) and Recall (R), defined in Eqs. (2) and (3), quantify detection correctness and completeness, respectively. Precision reflects the proportion of true positive detections among all predicted positives, while recall measures the proportion of true positives identified among all ground-truth instances [27].

$$Precision (P) = \frac{TP}{TP + FP} \quad (2)$$

$$Recall (R) = \frac{TP}{TP + FN} \quad (3)$$

Because precision and recall often exhibit a trade-off, the F1 score is used as a balanced metric, representing the harmonic mean of precision and recall, as shown in Eq. (4) [28].

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Average Precision (AP) captures the area under the precision–recall curve for a single class, while Mean Average Precision (mAP) averages AP across all classes. This study focuses on mAP@0.5 (mAP50), which computes mean precision at a fixed IoU threshold of 0.5, as defined in Eq. (5).

This metric is widely adopted in object detection research because it balances localization accuracy and classification performance and allows direct comparison with prior wildfire detection studies [29, 30].

$$mAP@0.5 = \frac{1}{N} \sum_{c=1}^N AvgPrecision_c^{IoU=0.5} \quad (5)$$

During training, the model minimizes a composite loss function consisting of bounding box regression loss, classification loss, and Distribution Focal Loss (DFL). DFL refines bounding box localization by modeling distance distributions, improving accuracy for small and thin smoke features commonly observed in wildfire imagery [31].

Together, these metrics provide a comprehensive evaluation framework, ensuring that improvements in accuracy, localization quality, and detection robustness are consistently assessed.

3. METHODOLOGY

3.1 D-Fire dataset

The D-Fire dataset was designed to support accurate and real-time detection of fire and smoke under realistic environmental conditions. It consists of images grouped into four scenarios: fire-only, smoke-only, fire-and-smoke, and negative cases that contain visually similar patterns but no actual fire or smoke. The distribution of images across these scenarios is summarized in Table 1.

Following the original compilation by de Venâncio et al. [32], the dataset contains a total of 21,527 labeled images. Although the fire-only category includes fewer images, it exhibits a higher object density, with an average of 2.52 fire annotations per image. In contrast, smoke objects in the smoke-only and combined fire-and-smoke categories appear more sparsely, averaging 1.13 annotations per image. Overall, the dataset includes 26,557 bounding boxes, comprising 14,692 fire instances and 11,865 smoke instances, as detailed in Table 2.

Table 1. Distribution of images per scenario in the D-Fire dataset

Scenario	Description of Scenario	Image
Fire	Images containing only fire	1164
Smoke	Images containing only smoke	5867
Fire and smoke	Images containing both fire and smoke	4658
None	Images containing neither fire nor smoke	9838
Total of image		21527

Table 2. Annotation counts and average per image in the D-Fire dataset

Object Category	Bounding-Box	Average per-Image
Fire	14692	2.52
Smoke	11865	1.13
Total	26557	

Images were collected from diverse sources, including online repositories, controlled fire experiments conducted at the Technological Park of Belo Horizonte (Brazil),

surveillance cameras at the Universidade Federal de Minas Gerais (UFMG), and natural scenes from Serra Verde State Park. To further increase visual diversity, a limited number of synthetic samples were generated by compositing artificial smoke onto natural landscape backgrounds, simulating realistic forest conditions. Representative examples from the dataset, along with their ground-truth annotations, are shown in Figure 2.

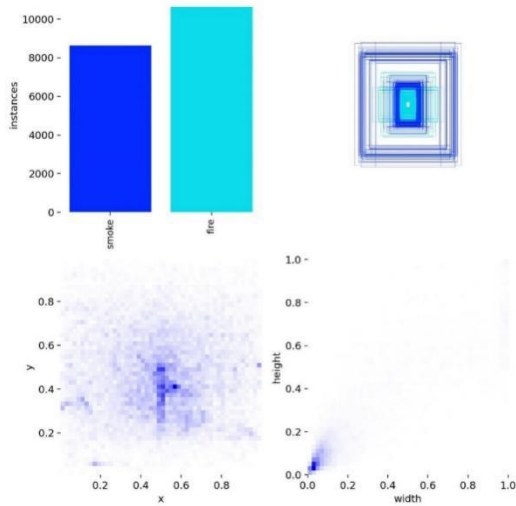


Figure 2. D-Fire dataset instances

The dataset covers a wide range of environments, including forests, parks, and semi-urban areas, and captures variations in

camera viewpoints, smoke density, and illumination. This diversity makes D-Fire particularly suitable for evaluating contrast enhancement techniques, especially in challenging scenarios involving thin or dispersed smoke, low visibility, or rapidly changing lighting conditions. To avoid bias toward artificial textures, synthetically augmented images were excluded from the test split. The inclusion of data from multiple acquisition sources and camera configurations provides strong intra-dataset diversity, supporting generalization beyond a single visual distribution.

3.2 Experimental configuration

In its original form, the D-Fire dataset is divided into training and testing subsets using an 80:20 split. To ensure consistency with prior studies—particularly the work of Alkhamash [19]—the training portion was further split into 72% for training and 8% for validation.

After the split, three image enhancement techniques were applied to generate three dataset variants: X Dataset, Y Dataset, and Z Dataset. The X Dataset applies global HE, which redistributes pixel intensity values to enhance overall contrast. The Y Dataset uses CLAHE, which performs localized contrast enhancement while constraining amplification to suppress noise. The Z Dataset incorporates DBST-LCM CLAHE, where Dynamic Block Size Technique-based noise reduction and Local Contrast Modification (LCM) are applied prior to CLAHE to further enhance subtle visual features. Examples of the enhanced images produced by these methods are shown in Figure 3.



Figure 3. Examples of enhanced images: (a) HE; (b) CLAHE; (c) DBST-LCM CLAHE

Note: HE: Histogram Equalization; CLAHE: Contrast-Limited Adaptive Histogram Equalization; DBST-LCM CLAHE: Dynamic Block Size Technique with Local Contrast Modification.

All enhancement procedures were conducted in the LAB color space to decouple luminance information from chromatic components. In wildfire imagery, smoke visibility is primarily governed by subtle luminance variations rather than color saturation. By applying enhancement only to the L* channel, contrast manipulation directly targets brightness gradients associated with semi-transparent smoke and flame glow, while preserving chromatic consistency in the a* and b* channels. This separation helps reduce color distortion under nonlinear illumination effects such as backlighting and haze, which are common in fire–smoke scenes.

Table 3. Preprocessing configuration for enhanced datasets

Method	Image Resize	Color Space	Core Technique	Key Parameters
HE			Global Histogram Equalization (HE)	-
CLAHE	Original Size	RGB → LAB	Contrast Limited Adaptive HE	ClipLimit = 1.0, TileGridSize = (20, 20)
DBST-LCM CLAHE			Dynamic Block Size Technique + Local Contrast Modification + CLAHE	ClipLimit = 1.0, BlockSizes = [8, 16, 32, 64]

For reproducibility, the parameters used in each enhancement technique are explicitly listed. HE was applied on the L-channel of the LAB color space without tunable parameters. CLAHE was configured with a clip limit of 1.0 and a tile grid size of 20×20 , while DBST-LCM CLAHE used a clip limit of 1.0 and dynamically selected the best-performing block size among 8, 16, 32, 64 based on Peak Signal-to-Noise Ratio (PSNR) values. These configurations are summarized in Table 3, and all corresponding preprocessing scripts are available in the public repository at GitHub.

For model training, YOLOv12N was selected as the base detector. This nano-scale variant uses a depth multiplier of 0.5, a width multiplier of 0.25, and a maximum channel limit of 1024 [33]. Compared with larger variants such as YOLOv12X, YOLOv12N significantly reduces model complexity, resulting in faster inference and lower memory usage. These characteristics make it well suited for real-time wildfire detection in resource-constrained deployment environments.

Figure 4 illustrates the YOLOv12N architecture, which consists of three main components: backbone, neck, and detection head. The backbone extracts hierarchical visual features through convolutional and C3 blocks, capturing both fine-grained smoke textures and broader fire regions. The neck aggregates multi-scale features via concatenation and upsampling, preserving contextual information across resolutions. The detection head produces predictions at three scales (80×80 , 40×40 , and 20×20), enabling accurate localization of both small smoke plumes and larger flame regions.

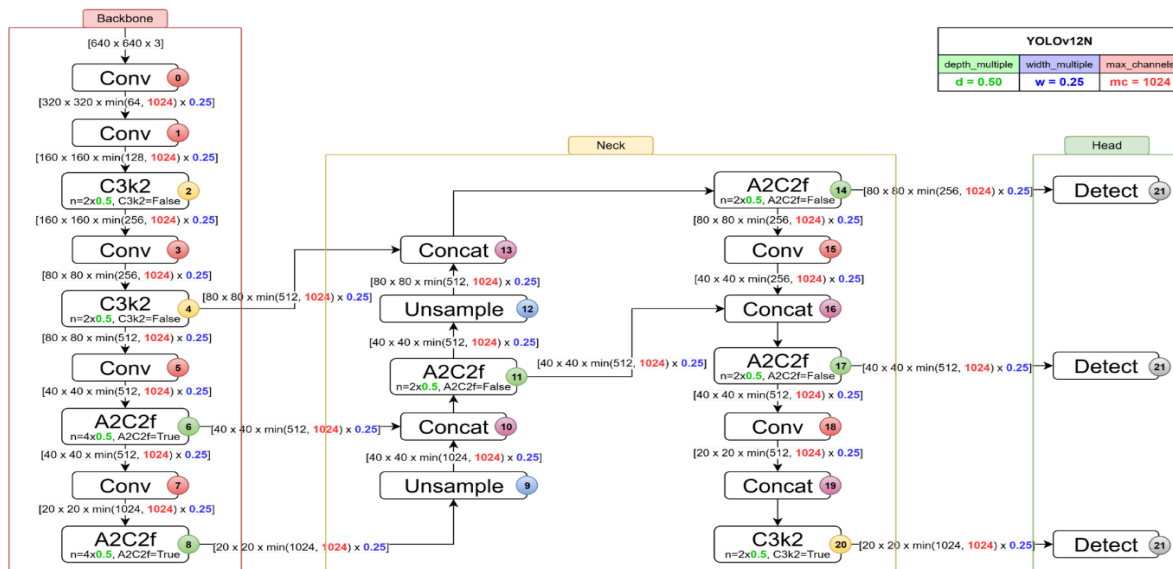


Figure 4. YOLOv12N model architecture

Note: A2C2f: Area Attention Enhanced Cross Stage Partial with Fusion (as detailed in Appendix)

The choice of YOLOv12N is motivated not only by computational efficiency, but also by its suitability for detecting weak visual signals characteristic of early-stage wildfires. Reduced depth and channel width limit excessive feature compression, which can otherwise attenuate faint smoke textures. At the same time, multi-scale prediction heads preserve sufficient spatial context, allowing YOLOv12N to maintain sensitivity to low-contrast patterns while remaining lightweight.

3.3 Training and validation

All experiments were conducted using the hardware configuration summarized in Table 4. Although this setup provides improved computational resources, training parameters were intentionally kept consistent with prior work to ensure fair comparison across YOLO variants. As shown in Table 5, all models were trained for 100 epochs with an input resolution of 640×640 , a batch size of 64, and the Stochastic Gradient Descent (SGD) optimizer.

Maintaining identical training settings across YOLOv10N, YOLOv11N, and YOLOv12N ensures that observed performance differences can be attributed to architectural advancements—such as the R-ELAN backbone, 7×7 separable convolutions, and FlashAttention—rather than differences in optimization or data handling. While it is acknowledged that further hyperparameter tuning could yield additional improvements, the uniform configuration adopted here prioritizes comparability and experimental fairness.

Table 4. System configuration for model training

Name	Type
CPU	8vCPUs
GPU	Nvidia A100-SXM4-40GB
RAM	83 GB
Framework	PyTorch 2.6.0 + CUDA 12.4
Accelerator	GPU A100

Table 5. Training parameters vs. prior work

Parameter	Values		
	YOLOv10	YOLOv11	YOLOv12
epochs	100	100	100
imgsz	640	640	640
batch	64	64	64
optimizer	SGD	SGD	SGD
model weight	YOLOv10N	YOLOv11N	YOLOv12N
number of parameters (million)	2.3	2.6	2.6

3.4 Training and inference settings

During training, standard YOLO data augmentation strategies were employed to improve generalization under diverse wildfire conditions. These included random horizontal flipping, random scaling, Hue, Saturation, and Value (HSV) color jittering, and Mosaic augmentation, which combines multiple images to enhance multi-scale feature learning. MixUp augmentation was intentionally disabled, as excessive blending between fire and smoke regions can distort thin smoke boundaries and degrade detection accuracy.

All models were trained using the SGD optimizer with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 5×10^{-4} . A cosine learning rate scheduler was employed over 100 epochs to ensure smooth and stable convergence.

During inference, non-maximum suppression (NMS) was applied with a confidence threshold of 0.25 and an IoU threshold of 0.45, following default YOLO evaluation settings. Class-agnostic NMS was disabled to preserve class-specific suppression between fire and smoke detections.

Runtime profiling was performed during validation using YOLOv12’s built-in speed reporting on the same hardware configuration listed in Table 4. As summarized in Table 6, the baseline YOLOv12N achieved an average inference time of 1.8 ms per image, with preprocessing and postprocessing latencies of 0.1 ms and 0.9 ms, respectively. When histogram-based enhancement methods were applied, preprocessing time increased marginally to approximately 0.2 ms per image. However, inference time decreased to 1.4 ms for HE and 1.3 ms for both CLAHE and DBST-LCM CLAHE, while postprocessing latency remained stable at 0.9 ms and 0.8 ms, respectively.

These results indicate that the proposed enhancement methods introduce negligible computational overhead, and the observed reduction in inference latency suggests that improved contrast facilitates more efficient feature extraction without affecting NMS behavior or real-time deployment feasibility.

Table 6. Per-image runtime analysis of YOLOv12N under different preprocessing methods

Model	Preprocess (ms)	Inference (ms)	Postprocess (ms)
YOLOv12N (baseline)	0.1	1.8	0.9
YOLOv12N + HE	0.2	1.4	0.9
YOLOv12N + CLAHE	0.2	1.3	0.8
YOLOv12N + DBST-LCM CLAHE	0.2	1.3	0.8

4. RESULT AND DISCUSSION

Table 7 summarizes the baseline training performance of YOLOv10N, YOLOv11N, and YOLOv12N. Among the three models, YOLOv12N achieves the strongest overall results, with a precision of 78.8%, a recall of 73.0%, and an mAP50 of 80.1%. Notably, the precision of YOLOv10N (78.3%) is closer to that of YOLOv12N than to YOLOv11N, suggesting that earlier YOLO variants were already effective at limiting false positives. In contrast, recall shows a clearer upward trend across successive versions, increasing by roughly two percentage points per generation, which reflects gradual improvements in capturing fire and smoke instances. Although the gain in mAP50 from YOLOv11N to YOLOv12N is relatively modest, YOLOv12N exhibits a more pronounced advantage under stricter localization criteria, reaching an mAP@50-95 of 47.6%. This indicates more consistent bounding-box localization and class prediction, supporting the use of YOLOv12N as a stable baseline for subsequent enhancement experiments.

Table 7. YOLOv10-YOLOv12 training performance

Models	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLOv10N	78.3	69.5	77.1	-
YOLOv11N	77.2	71.8	78.0	45.8
YOLOv12N	78.8	73.0	80.1	47.6

Before examining localized contrast enhancement methods, it is important to clarify the performance degradation observed when applying global HE. In wildfire imagery, HE tends to over-amplify high-intensity flame regions, increasing local intensity variance and disrupting smooth gradients around flame boundaries. These distortions negatively affect bounding-box regression and reduce localization reliability. At the same time, the altered intensity distribution interferes with Distribution Focal Loss (DFL) optimization for small and thin smoke objects, leading to lower recall and mAP. This behavior highlights the limitations of global enhancement and motivates the use of localized and adaptive contrast enhancement techniques.

As discussed in Section 2.2, the architectural design of

YOLOv12 plays a central role in addressing the visual challenges associated with wildfire detection. The R-ELAN backbone improves the preservation of subtle smoke gradients, while the expanded receptive field of the 7×7 separable convolutions supports the detection of elongated and dispersed smoke patterns. In addition, FlashAttention enhances sensitivity to low-contrast and partially occluded flame regions. Together, these components contribute to the observed improvements in precision, recall, and localization

accuracy compared with earlier YOLO variants.

Table 8 shows the training results of YOLOv12N on the D-Fire dataset using three image enhancement methods: X Dataset (HE), Y Dataset (CLAHE), and Z Dataset (DBST-LCM CLAHE). The results reveal distinct performance characteristics for each enhancement strategy, underscoring the different trade-offs introduced by global, local, and adaptive contrast enhancement.

Table 8. Performance metrics of three enhanced datasets trained using YOLOv12N

Class	X Dataset				Y Dataset				Z Dataset			
	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)
Smokes	80.5	78.1	83.4	51.7	88.1	85.1	91.3	65.5	82.2	81.3	86.0	55.1
Fire	71.0	59.9	68.5	35.6	80.7	72.0	82.2	48.9	71.9	63.7	72.6	38.6
All	75.8	69.0	76.0	43.7	84.4	78.5	86.7	57.2	77.0	72.5	79.3	46.9

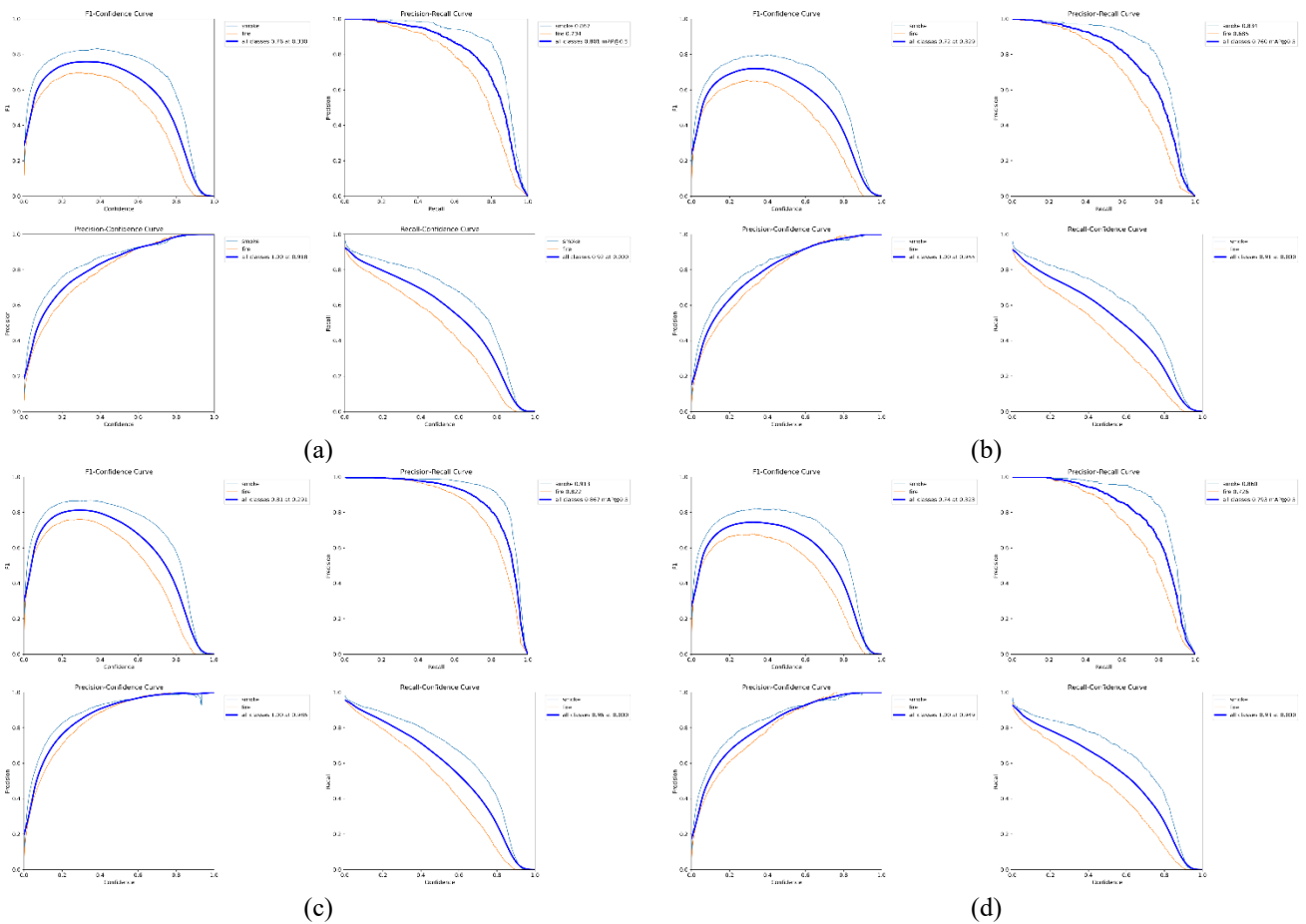


Figure 5. Performance evaluation curves of YOLOv12N and enhanced models on the D-Fire dataset: (a) Original YOLOv12N; (b) HE-enhanced; (c) CLAHE-enhanced; (d) DBST-LCM CLAHE-enhanced

The Y Dataset, which applies CLAHE, delivered the strongest performance across all evaluation metrics. It achieved a precision of 84.4%, a recall of 78.5%, an mAP50 of 86.7%, and an mAP50-95 of 57.2%. These results clearly exceed those of the baseline YOLOv12N model without preprocessing, which obtained 78.8% precision, 73.0% recall, 80.1% mAP50, and 47.6% mAP50-95. Relative to the baseline, CLAHE yields absolute gains of 5.6 percentage points in precision, 5.5 in recall, 6.6 in mAP50, and 9.6 in mAP50-95. These improvements indicate that localized contrast enhancement substantially benefits both detection accuracy and bounding-box localization. By enhancing local

intensity variations, CLAHE improves the visibility of low-contrast structures such as faint smoke, leading to more reliable detection of both fire and smoke regions.

The Z Dataset, processed with DBST-LCM CLAHE, ranked second, with a precision of 77.0%, recall of 72.5%, mAP50 of 79.3%, and mAP50-95 of 46.9%. Although this method effectively suppresses noise and enhances local contrast, its adaptive smoothing can attenuate fine visual details, which likely contributes to its slightly lower performance compared with CLAHE. In contrast, the X Dataset using global HE exhibited the weakest results, recording 75.8% precision, 69.0% recall, 76.0% mAP50, and

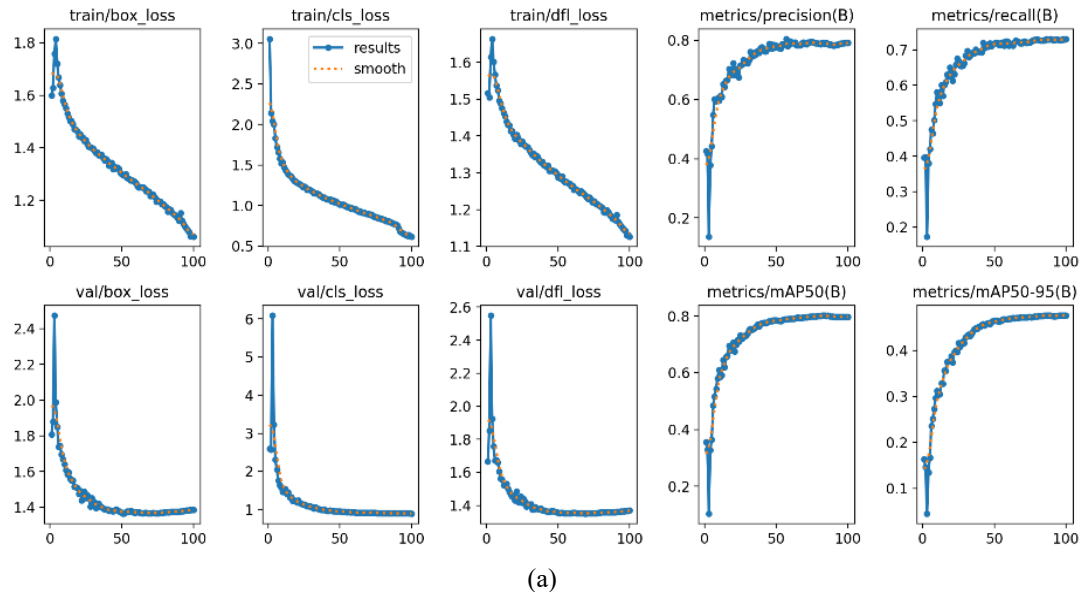
43.7% mAP50-95. This outcome reflects the limitations of uniform contrast redistribution, which tends to amplify noise and fails to adapt to local intensity variations commonly present in wildfire imagery.

Validation results closely followed the trends observed during training, indicating good generalization to unseen data and suggesting that the models neither overfit nor underfit. This consistency across splits supports further analysis using F1 scores, loss convergence, and qualitative comparisons. Figure 5 illustrates the confidence-based performance curves for YOLOv12N and the three enhanced datasets, highlighting differences in stability and emphasizing the robustness of localized enhancement strategies.

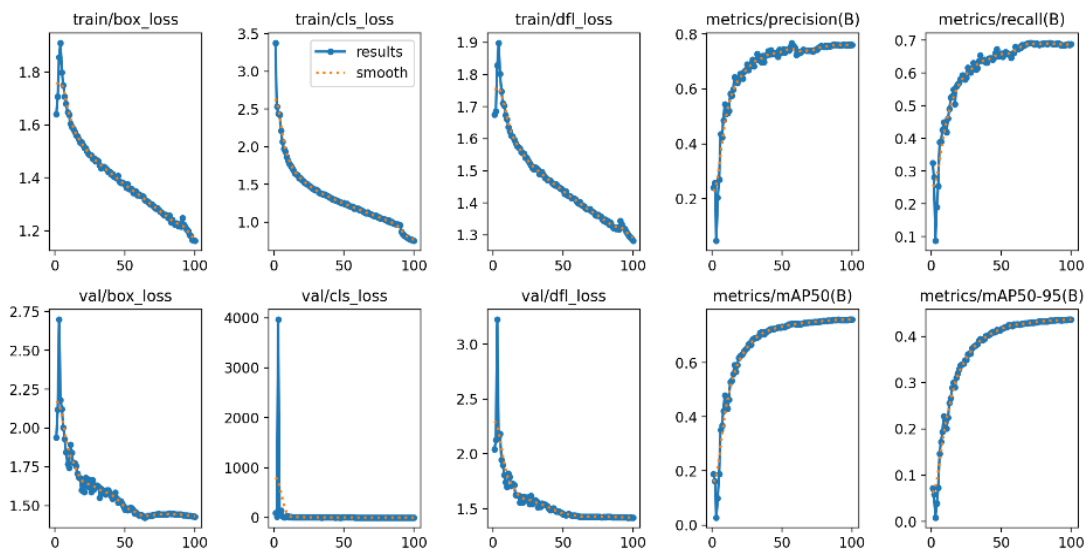
Figure 5 summarizes confidence-based performance across all models. Both the baseline YOLOv12N and the HE-enhanced variant show a faster decline in F1 score as the confidence threshold increases, indicating reduced robustness under stricter decision criteria. In contrast, the CLAHE-enhanced model maintains higher F1 scores over a wider confidence range, while DBST-LCM CLAHE demonstrates moderate but more stable behavior than global HE. These patterns are consistent with the quantitative results reported in Table 8 and further support the benefit of localized contrast enhancement for improving detection stability.

Beyond metric- and confidence-based evaluation, Figure 6 illustrates the training and validation loss curves for the baseline and enhanced models. All configurations exhibit clear convergence behavior, characterized by rapid loss reduction within the first 20–30 epochs followed by gradual stabilization. Alongside box loss, both classification (cls) and distribution focal loss (dfl) decrease steadily, with training cls converging around 1.0–1.2 and dfl around 1.1–1.3, closely tracked by the corresponding validation curves.

Among the evaluated methods (Figure 6(a-d)), the CLAHE-enhanced model (Figure 6(c)) reaches the lowest final cls and dfl values, indicating more effective feature discrimination against fire and smoke, particularly in low-contrast scenes. The DBST-LCM CLAHE model (Figure 6(d)) also converges smoothly but stabilizes at slightly higher loss values, which is consistent with the smoothing effects introduced during preprocessing. In comparison, the HE-enhanced model (Figure 6(b)) shows greater early fluctuation in validation loss, likely due to noise amplification, while the baseline YOLOv12N (Figure 6(a)) remains stable but converges more gradually. Overall, these observations reinforce earlier conclusions that CLAHE offers the most reliable enhancement, leading to improved learning efficiency and reduced classification uncertainty.



(a)



(b)

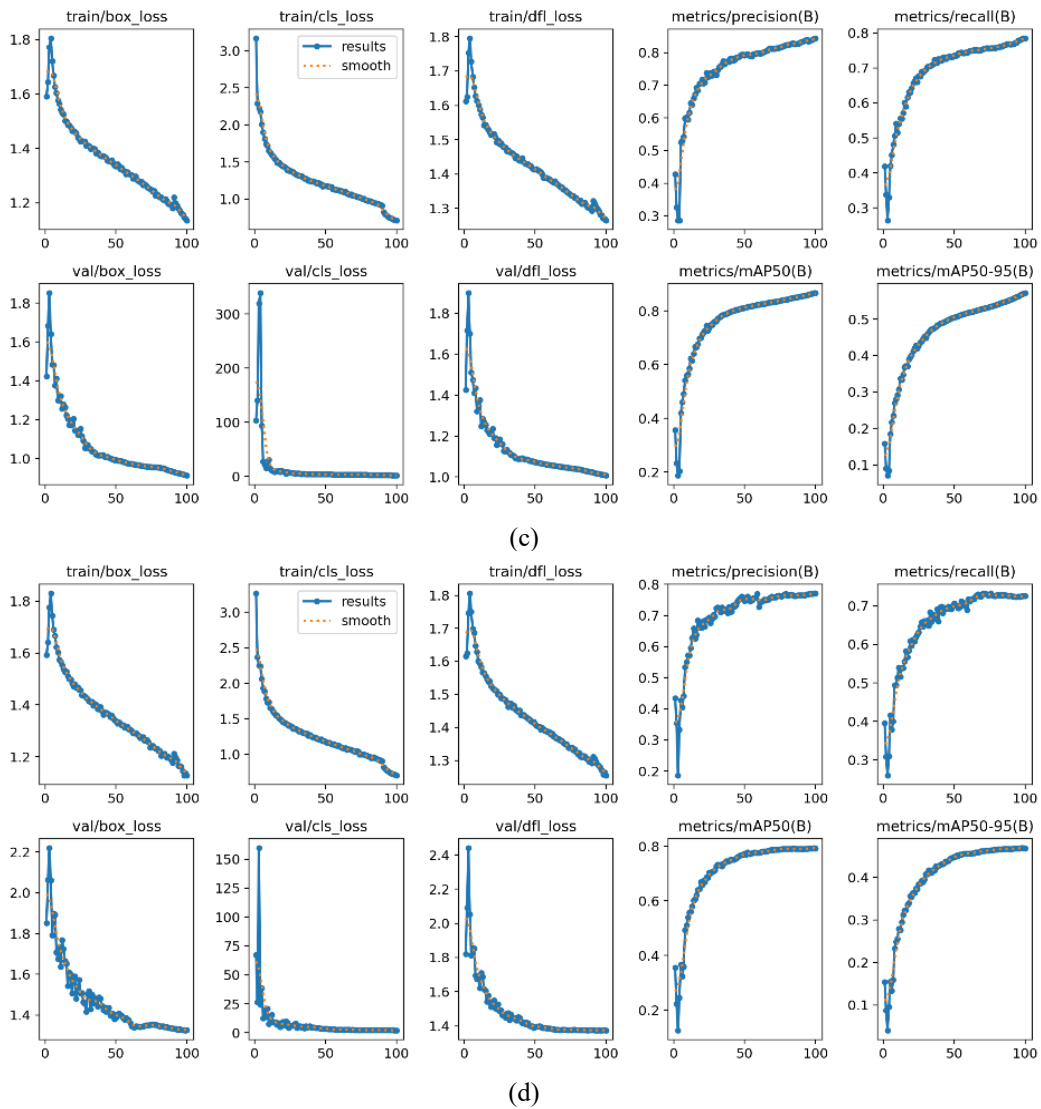


Figure 6. Training and validation loss curves for (a) Original YOLOv12N; (b) HE-enhanced; (c) CLAHE-enhanced; (d) DBST-LCM CLAHE-enhanced

Note: cls: Class Loss; df_l: Distribution Focal Loss; box: Box Loss (Regression); val: Validation Data; B: Best/Box Metrics.

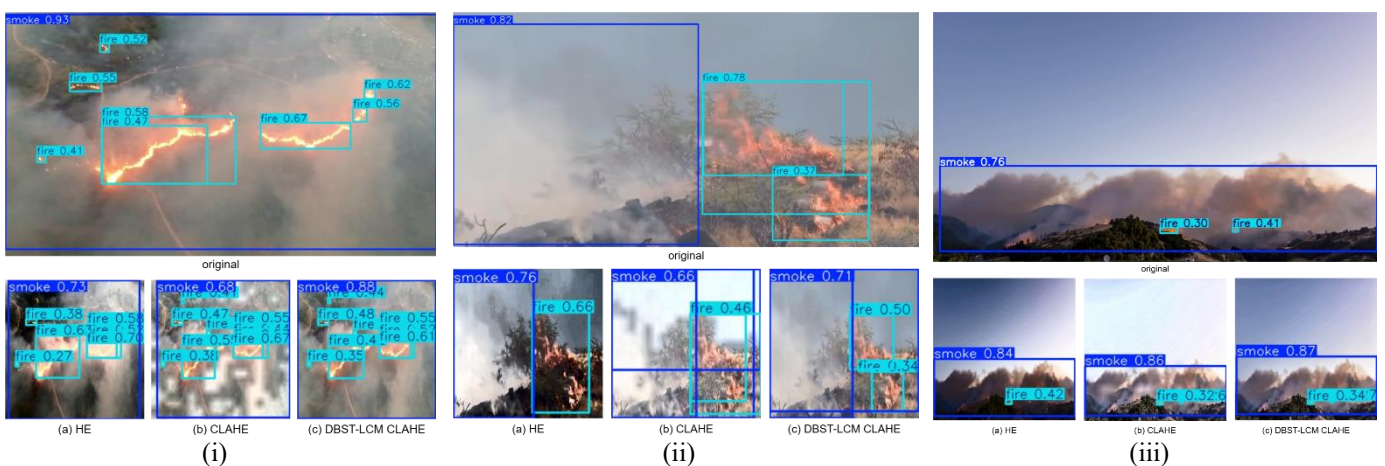


Figure 7. Qualitative comparison of YOLOv12 detection results on wildfire imagery under different scenarios from the same dataset: (i) scenario 1; (ii) scenario 2; (iii) scenario 3, using (a) Histogram Equalization (HE), (b) CLAHE, and (c) DBST-LCM Contrast-Limited Adaptive Histogram Equalization (CLAHE)

Although the quantitative metrics and curve-based analyses demonstrate the clear advantage of CLAHE, they do not fully reflect model behavior on real wildfire imagery. To address

this, Figure 7 provides a qualitative comparison of detection results across three representative scenarios from the D-Fire dataset. These examples illustrate how different enhancement

strategies influence the localization of fire and smoke under visually challenging conditions, such as low contrast, haze, and dispersed smoke.

In addition, Table 9 compares the proposed YOLOv12N + CLAHE approach with several recent YOLO-based models, including YOLOv4, Tiny-YOLOv4, and YOLOv11X + HE, all evaluated on the same D-Fire dataset. To ensure a fair comparison, all models were trained and tested using the unified configuration summarized in Table 5, with identical dataset splits, input resolution (640 × 640), and hyperparameter settings. Under these consistent conditions, YOLOv12N + CLAHE achieves the highest performance, reaching an mAP50 of 86.7% and an mAP50-95 of 57.2%. This improvement is obtained while preserving lightweight architecture, making the proposed method well suited for real-time wildfire monitoring. Overall, these results confirm that combining localized contrast enhancement with YOLOv12N leads to more robust and accurate detection of both fire and smoke regions.

Table 9. Previous research comparison

Models	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLOv4 [34]	-	-	76.56	-
Tiny-YOLOv4 [34]	-	-	63.34	-
YOLOv11X+HE [35]	78.4	70.3	77.1	46.2
YOLOv12N+CLAHE	84.4	78.5	86.7	57.2

In Scenario 1 (Figure 7(i)), the baseline YOLOv12N produced the strongest smoke detection, reaching a confidence of 93%, with fire detections remaining relatively stable between 41% and 67%. When global HE was applied, smoke confidence dropped to 73%, and fire predictions became noticeably unstable (27%–70%), likely due to noise amplification. CLAHE yielded more consistent outputs, with smoke detected at 68% and fire confidence ranging from 38% to 67%, although overall confidence was lower than the baseline. DBST-LCM CLAHE increased smoke confidence to 88%, approaching the baseline performance, but fire detection remained comparatively weaker (35%–61%).

In Scenario 2 (Figure 7(ii)), the baseline model achieved smoke confidence of 82% and fire detections at 37% and 78%. HE slightly reduced smoke confidence to 76% but produced a relatively stable fire prediction at 66%. CLAHE resulted in lower confidence for both smoke (66%) and fire (46%), suggesting that localized contrast enhancement may reduce feature visibility in already well-lit scenes. DBST-LCM CLAHE produced moderate results, with smoke at 71% and fire between 34% and 50%, offering improved stability over HE but still underperforming relative to the baseline. These observations indicate that enhancement methods do not uniformly improve detection in high-contrast conditions, underscoring the importance of dataset-level evaluation rather than reliance on individual cases.

In Scenario 3 (Figure 7(iii)), where visual contrast is limited, all enhancement methods improved detection performance. The baseline YOLOv12N achieved smoke confidence of 76% and fire confidence between 30% and 41%. HE increased smoke confidence to 84% and fire to 42%. CLAHE further improved smoke detection to 86%, with fire confidence ranging from 32% to 56%. DBST-LCM CLAHE produced the strongest qualitative result in this scenario, reaching 87% smoke confidence and fire confidence between 34% and 67%. This scenario highlights the effectiveness of

contrast enhancement in low-visibility conditions, where faint smoke and small flame regions are otherwise difficult to detect.

Overall, the qualitative results show that CLAHE and DBST-LCM CLAHE enhance smoke visibility more effectively than global HE, with DBST-LCM CLAHE providing the strongest improvements in certain low-contrast cases. However, this behavior differs from the quantitative evaluation, where CLAHE achieved the highest overall performance. This discrepancy reflects the distinction between case-specific behavior and dataset-level robustness. While DBST-LCM CLAHE can yield visually stronger results in selected scenarios, its adaptive smoothing may suppress fine structural details when applied broadly. In contrast, CLAHE offers a more balanced enhancement, preserving critical features while avoiding excessive noise amplification or smoothing.

Accordingly, CLAHE-enhanced YOLOv12N emerges as the most reliable configuration for real-time wildfire detection. Although DBST-LCM CLAHE may outperform CLAHE in specific low-contrast scenes, CLAHE demonstrates greater consistency across diverse visual conditions. These findings emphasize the value of combining quantitative metrics with qualitative analysis when evaluating detection systems for highly variable wildfire imagery.

Each enhancement method introduces distinct trade-offs. HE improves global brightness but tends to amplify noise; DBST-LCM CLAHE enhances local contrast while potentially smoothing subtle smoke textures; and CLAHE achieves a balanced improvement with minimal artifact introduction. The choice of enhancement strategies should therefore be guided by operational priorities and environmental conditions.

Regarding deployment feasibility, YOLOv12N is expected to maintain competitive inference speed due to its lightweight architecture. Turmaganbet et al. [36] report approximately 7 FPS for YOLOv12N on moderate hardware, compared with 5 FPS for YOLOv8s and 4 FPS for YOLOv11s, while larger variants such as YOLOv12s operate near 2 FPS. Although direct runtime and preprocessing latency measurements were not conducted in this study, architectural complexity and prior benchmarking suggest that integrating CLAHE with YOLOv12N remains suitable for real-time or near-real-time wildfire monitoring applications.

Finally, it is noted that all evaluations were conducted under controlled conditions using fixed dataset splits and identical training configurations. Rather than statistical population inference, reliability is assessed through consistent performance gains observed across multiple metrics, confidence-based curves, convergence behavior, and qualitative examples. The alignment of these evaluation perspectives supports the robustness of the reported improvements and indicates that the observed gains are not attributable to random variation.

5. CONCLUSIONS

This study demonstrates that integrating YOLOv12 with histogram-based visual preprocessing improves wildfire detection performance on the D-Fire dataset. Among the evaluated methods, CLAHE provides the most consistent gains. The CLAHE-enhanced YOLOv12N achieves the highest precision (84.4%), recall (78.5%), and mAP@0.5

(86.7%), outperforming both the baseline model and alternative enhancement strategies. DBST-LCM CLAHE ranks second, effectively reducing noise and enhancing local contrast but occasionally smoothing fine details, while global HE shows the weakest performance due to noise amplification.

Loss curve analysis supports these findings, with CLAHE yielding the lowest classification and distribution focal losses, indicating improved separation between fire and smoke features. Qualitative evaluations further show that CLAHE enhances smoke visibility under faint and hazy conditions, whereas DBST-LCM CLAHE can produce stronger visual improvements in certain low-contrast scenarios. At the same time, the comparison reveals differences between averaged quantitative metrics and case-specific behavior, highlighting the influence of contextual visual conditions on real-world performance.

From a deployment perspective, the results point to several opportunities for further refinement rather than fundamental limitations. The reported improvements reflect end-to-end pipeline performance and do not yet isolate the individual effects of preprocessing and YOLOv12's internal attention mechanisms, which is a valuable direction for future analysis, particularly in resource-constrained settings. In addition, robustness to practical imaging degradations—such as motion blur, compression artifacts, and abrupt illumination changes—remains to be systematically evaluated, despite their prevalence in operational wildfire monitoring systems. Runtime efficiency and preprocessing latency were also inferred from architectural characteristics and prior benchmarks rather than directly measured, suggesting that deployment-oriented profiling on edge devices, UAV platforms, or large-scale camera networks would further strengthen validation.

Overall, CLAHE-enhanced YOLOv12N offers the most reliable balance between accuracy and stability, making it a strong candidate for real-time wildfire detection. While DBST-LCM CLAHE shows promise in extreme low-contrast conditions, further refinement is needed to mitigate its smoothing effects. Future work will explore adaptive hybrid enhancement strategies, multimodal data integration such as thermal or satellite imagery, and deployment-focused optimization on edge platforms, extending the applicability of this approach to broader environmental monitoring and operational fire management contexts.

ACKNOWLEDGMENT

This research is supported by the Vice-Rector of Research, Innovation, and Entrepreneurship at Satya Wacana Christian University 007/SPK-PF/RIK/03/2025 and Honchita Co., Ltd. (Uniform Number: 60304896), Changhua, Taiwan.

AUTHOR CONTRIBUTION

Conceptualization and study design were carried out by Christine Dewi; Melati Viaeritas Vitrieco Santoso; Abbott Po Shun Chen; and Hanna Prillysca Chernovita. Data collection was performed by Melati Viaeritas Vitrieco Santoso; Stephen Abednego Philemon; and Stephen Aprius Sutresno. Analysis and interpretation of the results were conducted by Evangs Mailoa; Christine Dewi; Hanna Prillysca Chernovita; Stephen Aprius Sutresno; and Abbott Po Shun Chen. The manuscript

was drafted by Abbott Po Shun Chen; Christine Dewi; Hanna Prillysca Chernovita; Evangs Mailoa; Stephen Abednego Philemon; and Stephen Aprius Sutresno. All authors reviewed the results and approved the final version of the manuscript.

AVAILABILITY OF DATA AND MATERIALS

D-Fire: An image dataset for fire and smoke detection, GitHub repository, accessed 1 May 2025.

CONFLICT OF INTEREST AND GENERATIVE AI USE STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

Generative artificial intelligence (AI) tools were used solely to assist with language editing, grammar correction, and improving the clarity of the manuscript. The research design, experiments, data analysis, and interpretation of results were conducted entirely by the authors, who take full responsibility for the final content of this manuscript.

REFERENCES

- [1] Jones, M.W., Kelley, D.I., Burton, C.A., Di Giuseppe, F., et al. (2024). State of wildfires 2023–2024. *Earth System Science Data*, 16: 3601–3685. <https://doi.org/10.5194/essd-16-3601-2024>
- [2] Samborska, V., Ritchie, H. (2024). Wildfires: Explore global and country-level data on the extent of wildfires and how they've changed over time. *Our World in Data*. <https://ourworldindata.org/wildfires>.
- [3] Singh, S. (2022). Forest fire emissions: A contribution to global climate change. *Frontiers in Forests and Global Change*, 5: 925480. <https://doi.org/10.3389/ffgc.2022.925480>
- [4] Grant, E., Runkle, J.D. (2022). Long-term health effects of wildfire exposure: A scoping review. *Journal of Climate and Health*, 6: 100110. <https://doi.org/10.1016/j.joclim.2021.100110>
- [5] The Economist. (2025). Wildfires devastated the Amazon basin in 2024. *The Economist*. <https://www.economist.com/the-americas/2025/05/21/wildfires-devastated-the-amazon-basin-in-2024>.
- [6] Fisher, R. (2024). Vastly bigger than the Black Summer: 84 million hectares of northern Australia burned in 2023. *The Conversation*. <https://doi.org/10.64628/AA.4fga4p5vs>
- [7] University of California San Diego. (2023). ALERTCalifornia. <https://alertcalifornia.org/>.
- [8] Spiller, D., Carbone, A., Amici, S., Thangavel, K., Sabatini, R., Laneve, G. (2023). Wildfire detection using convolutional neural networks and PRISMA hyperspectral imagery: A spatial–spectral analysis. *Remote Sensing*, 15(19): 4855. <https://doi.org/10.3390/rs15194855>
- [9] Al-Smadi, Y., Alauthman, M., Al-Qerem, A., Aldweesh, A., Quaddoura, R., Aburub, F., Mansour, K., Alhmiedat, T. (2023). Early wildfire smoke detection using different YOLO models. *Machines*, 11(2): 246.

- <https://doi.org/10.3390/machines11020246>
- [10] Sathishkumar, V.E., Cho, J., Subramanian, M., Naren, O.S. (2023). Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecology*, 19: 9. <https://doi.org/10.1186/s42408-022-00165-0>
- [11] Geng, X., Su, Y.X., Cao, X.H., Li, H.Z., Liu, L.G. (2024). YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n. *Scientific Reports*, 14: 4543. <https://doi.org/10.1038/s41598-024-55232-0>
- [12] Dou, X.K., Wang, T., Shao, S.L. (2023). A lightweight YOLOv5 model integrating GhostNet and attention mechanism. In 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL), Zhuhai, China, pp. 348-352. <https://doi.org/10.1109/CVIDL58838.2023.10166155>
- [13] Li, Y.M., Zhang, W., Liu, Y.Y., Jing, R.D., Liu, C.S. (2022). An efficient fire and smoke detection algorithm based on an end-to-end structured network. *Engineering Applications of Artificial Intelligence*, 116: 105492. <https://doi.org/10.1016/j.engappai.2022.105492>
- [14] Prasad, H., Singh, A., Thakur, J., Choudhary, C., Vyas, N. (2023). Artificial intelligence-based fire and smoke detection and security control system. In 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, pp. 1-6. <https://doi.org/10.1109/NMITCON58196.2023.10276198>
- [15] Li, T.T., Zhang, C.C., Zhu, H.W., Zhang, J.G. (2022). Adversarial fusion network for forest fire smoke detection. *Forests*, 13(3): 366. <https://doi.org/10.3390/f13030366>
- [16] Gragnaniello, D., Greco, A., Sansone, C., Vento, B. (2024). Fire and smoke detection from videos: A literature review under a novel taxonomy. *Expert Systems with Applications*, 255: 124783. <https://doi.org/10.1016/j.eswa.2024.124783>
- [17] Zakari, R.Y., Malik, O.A., Wee-Hong, O. (2025). Spatio-temporal wildfire forecasting in Australia using deep learning and explainable AI. *Modeling Earth Systems and Environment*, 11: 425. <https://doi.org/10.1007/s40808-025-02621-7>
- [18] Sapkota, R., Flores-Calero, M., Qureshi, R., Badgujar, C., et al. (2025). YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series. *Artificial Intelligence Review*, 58: 274. <https://doi.org/10.1007/s10462-025-11253-3>
- [19] Alkhamash, E.H. (2025). A comparative analysis of YOLOv9, YOLOv10, YOLOv11 for smoke and fire detection. *Fire*, 8(1): 26. <https://doi.org/10.3390/fire8010026>
- [20] Ngoc, P.V.B., Hoang, L.H., Hieu, L.M., Nguyen, N.H., Thien, N.L., Doan, V.T. (2023). Real-time fire and smoke detection for trajectory planning and navigation of a mobile robot. *Engineering, Technology & Applied Science Research*, 13(5): 11843-11849. <https://doi.org/10.48084/etasr.6252>
- [21] Buriboev, A.S., Abduvaitov, A., Jeon, H.S. (2025). Binary classification of pneumonia in chest X-ray images using modified contrast-limited adaptive histogram equalization algorithm. *Sensors*, 25(13): 3976. <https://doi.org/10.3390/s25133976>
- [22] Zhang, Z.Q., Su, Y.F. (2025). Optimizing visual communication in online classrooms using image processing technology. *Traitement du Signal*, 42(1): 373-381. <https://doi.org/10.18280/ts.420132>
- [23] Werdingsih, I., Puspitasari, I., Hendradi, R. (2025). Impact of image enhancement on deep learning-based recognition of activity prompts in children with autism using Motion History Images. *Mathematical Modelling of Engineering Problems*, 12(10): 3718-3728. <https://doi.org/10.18280/mmep.121035>
- [24] Chakraverti, S., Agarwal, P., Pattanayak, H.S., Chauhan, S.P.S., Chakraverti, A.K., Kumar, M. (2024). De-noising the image using DBST-LCM-CLAHE: A deep learning approach. *Multimedia Tools and Applications*, 83: 11017-11042. <https://doi.org/10.1007/s11042-023-16016-2>
- [25] De Venâncio, P.V.A.B., Rezende, T.M., Lisboa, A.C., Barbosa, A.V. (2021). Fire detection based on a two-dimensional convolutional neural network and temporal analysis. In 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Temuco, Chile, pp. 1-6. <https://doi.org/10.1109/LA-CCI48322.2021.9769824>
- [26] Patel, O., Maravi, Y.P.S., Sharma, S. (2013). A comparative study of histogram equalization-based image enhancement techniques for brightness preservation and contrast enhancement. *Signal & Image Processing: An International Journal*, 4(5): 11-26. <https://doi.org/10.5121/sipij.2013.4502>
- [27] Conciatori, M., Valletta, A., Segalini, A. (2024). Improving the quality evaluation process of machine learning algorithms applied to landslide time series analysis. *Computers & Geosciences*, 184: 105531. <https://doi.org/10.1016/j.cageo.2024.105531>
- [28] Goutte, C., Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science*, pp. 345-359. https://doi.org/10.1007/978-3-540-31865-1_25
- [29] Padilla, R., Netto, S.L., da Silva, E.A.B. (2020). A survey on performance metrics for object-detection algorithms. In 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, pp. 237-242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
- [30] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, C., Zisserman, A. (2010). The PASCAL Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88: 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- [31] Tong, C.H., Yang, X.H., Huang, Q., Qian, F.Y. (2022). NGIoU loss: Generalized intersection over union loss based on a new bounding box regression. *Applied Sciences*, 12(24): 12785. <https://doi.org/10.3390/app122412785>
- [32] de Venâncio, P.V.A.B., Lisboa, A.C., Barbosa, A.V. (2022). An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Computing and Applications*, 34: 15349-15368. <https://doi.org/10.1007/s00521-022-07467-z>
- [33] Tian, Y.J., Ye, Q.X., Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *arXiv*

preprint arXiv:2502.12524.
<https://doi.org/10.48550/arXiv.2502.12524>

- [34] de Venâncio, P.V.A.B., Campos, R.J., Rezende, T.M., Lisboa, A.C., Barbosa, A.V. (2023). A hybrid method for fire detection based on spatial and temporal patterns. *Neural Computing and Applications*, 35: 9349-9361. <https://doi.org/10.1007/s00521-023-08260-2>
- [35] Dewi, C., Santoso, M.V.V., Chernovita, H.P., Mailoa, E., Philemon, S.A., Chen, A.P.S. (2025). Integration of YOLOv11 and histogram equalization for fire and smoke-based detection of Forest and Land Fires. *Computers, Materials & Continua*, 84(3): 5361-5379. <https://doi.org/10.32604/cmc.2025.067381>
- [36] Turmaganbet, U.K., Zhexebay, D.M., Turlykozhayeva, D.A., Skabylov, A.A., Akhtanov, S.N., Temesheva, S.A., Tao, M., Masalim, P.C. (2025). Thermal infrared object detection with YOLO models. *Eurasian Physical Technical Journal*, 22(2): 121-132. <https://doi.org/10.31489/2025N2/121-132>

APPENDIX

A Glossary is provided to enable readers to quickly refer to and understand the key concepts mentioned in the paper.

This appendix provides definitions for the key terms and abbreviations used in the text.

A2 – An attention-based module used in deep neural networks to capture long-range dependencies and enhance feature representation by modeling relationships between spatial regions in feature maps.

A2C2f – An architectural module that integrates the A2 attention mechanism with the C2f structure to improve feature extraction efficiency and contextual representation in object detection networks.

AFN (Adaptive Feature Network) – A network component designed to enhance feature representation by adaptively adjusting feature maps to improve object detection performance.

box (Bounding Box Loss) – A loss component used in object detection models to measure the accuracy of predicted bounding box coordinates relative to the ground truth object locations.

C2PSA (Cross Stage Partial with Spatial Attention) – A neural network module that combines the Cross-Stage Partial architecture with spatial attention mechanisms to improve feature representation and enhance the model's representational capacity while maintaining computational efficiency.

C3 – A feature extraction module derived from the Cross-Stage Partial (CSP) architecture that improves gradient flow and reduces computational complexity in convolutional neural networks.

C3K2 – An improved variant of the C3 module that incorporates optimized kernel configurations to enhance feature representation and detection performance in YOLO architectures.

CLAHE (Contrast Limited Adaptive Histogram Equalization) – An image enhancement technique that improves local contrast by applying histogram equalization within small regions of an image while limiting noise amplification through contrast clipping.

cls (Classification Loss) – A loss component used to measure the accuracy of predicted object class probabilities

compared to the ground truth class labels.

CNN (Convolutional Neural Network) – A deep learning architecture designed to automatically learn hierarchical feature representations from images using convolutional operations.

CSPNet (Cross Stage Partial Network) – A neural network architecture designed to reduce computational cost while maintaining accuracy by partitioning feature maps into separate paths for gradient flow and feature reuse.

DBST-LCM CLAHE (Dynamic Block Size Technique – Local Contrast Modification Contrast Limited Adaptive Histogram Equalization) – An enhanced image processing method that combines dynamic block size adjustment and local contrast modification with CLAHE to improve image visibility and detail while minimizing noise amplification.

dfl (Distribution Focal Loss) – A loss function used in object detection to improve bounding box regression by modeling the probability distribution of bounding box coordinates.

ELAN (Efficient Layer Aggregation Network) – A neural network design that aggregates features from multiple convolutional layers to improve feature propagation and representation in deep architectures.

FLF (Forest and Land Fire) – A type of fire occurring in forested or land areas, often associated with environmental and ecological damage and requiring early detection systems for mitigation.

HE (Histogram Equalization) – A global image enhancement technique that redistributes the intensity values of an image to improve overall contrast.

IoU (Intersection over Union) – A metric used in object detection to evaluate the overlap between the predicted bounding box and the ground truth bounding box.

mAP50 (Mean Average Precision at IoU 0.50) – An evaluation metric that measures the average precision of object detection results using an Intersection over Union threshold of 0.50.

mAP50-95 (Mean Average Precision at IoU thresholds from 0.50 to 0.95) – A comprehensive evaluation metric that calculates the average precision across multiple IoU thresholds ranging from 0.50 to 0.95.

P (Precision) – A performance metric that measures the proportion of correctly predicted positive detections among all predicted positive detections.

PR Curve (Precision-Recall Curve) – A graphical representation that illustrates the trade-off between precision and recall at different confidence thresholds.

PSNR (Peak Signal-to-Noise Ratio) – A metric used to evaluate image quality by measuring the ratio between the maximum possible signal power and the noise affecting the image reconstruction.

R (Recall) – A performance metric that measures the proportion of correctly detected positive samples among all actual positive samples.

R-ELAN (Residual Efficient Layer Aggregation Network) – An extension of ELAN that incorporates residual connections to enhance feature aggregation, improve gradient flow, and enable more stable training in deep neural networks.

SGD (Stochastic Gradient Descent) – An optimization algorithm used to train machine learning models by iteratively updating model parameters using randomly selected subsets of training data.

SPPF (Spatial Pyramid Pooling Fast) – A pooling module used in YOLO architectures that applies multiple pooling

operations to capture multi-scale spatial features while maintaining computational efficiency.

YOLO (You Only Look Once) – A real-time object detection framework that performs object localization and classification in a single forward pass through a neural network.

YOLOv10 – A version of the YOLO object detection architecture designed to improve detection accuracy and efficiency through architectural and training optimizations.

YOLOv11 – An improved YOLO architecture that introduces enhanced feature extraction modules and attention

mechanisms to improve object detection performance.

YOLOv12 – A recent development of the YOLO architecture that integrates advanced backbone structures and attention mechanisms to enhance feature representation and detection capability.

YOLOv12N – The nano variant of YOLOv12 designed for lightweight deployment with reduced parameters and computational requirements.

YOLOv12X – The extra-large variant of YOLOv12 designed to achieve higher detection accuracy through larger model capacity and deeper architecture.